



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Jan Vágner

Ekonometrické SUR modely závislých run-off trojúhelníků v neživotním pojištění

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: prof. RNDr. Tomáš Cipra, DrSc.

Studijní program: Matematika

Studijní obor: MFMAT

Praha 2021

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

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 má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 14. 5. 2021

Jan Vágner

Na úvod bych rád poděkoval vedoucímu práce panu prof. RNDr. Tomáši Ciprovi, DrSc. za ochotu konzultovat jakýkoliv problém, který se při vypracovávání práce vyskytl a také za promptnost, se kterou na všechny mé dotazy s ochotou odpovídal.

Název práce: Ekonometrické SUR modely závislých run-off trojúhelníků v neživotním pojištění

Autor: Jan Vágner

Katedra pravděpodobnosti a matematické statistiky: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: prof. RNDr. Tomáš Cipra, DrSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Bakalářská práce se zabývá odhadem technických rezerv v neživotním pojištění. Pro odhady se využívá SUR modelu, který umožňuje do výpočtu zahrnout korelace mezi vývojovými trojúhelníky, vstupními daty. Zvolený model je tedy schopný při výpočtu uvažovat závislosti, které nastávají mezi jednotlivými vývojovými trojúhelníky. Díky tomuto rozšíření klasického modelu sloužícímu k odhadování rezerv dostáváme výrazně přesnější odhady, které lépe vystihují reálná pozorování. V práci kromě představení tohoto modelu uvádíme také aplikaci daného modelu na reálných datech za použití různých variací tohoto modelu a jeho porovnání s klasickou metodou pro odhady rezerv.

Klíčová slova: Závislé run-off trojúhelníky; SUR model; neživotní pojištění

Title: Econometric SUR models of dependent run-off triangles in non-life insurance

Author: Jan Vágner

Department of Probability and Mathematical Statistics: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Tomáš Cipra, DrSc., Department of Probability and Mathematical Statistics

Abstract: This bachelor thesis deals with the estimation of loss reserves in non-life insurance. To construct these estimates we use SUR model, which allows to include correlations between run-off triangles, input data. This SUR based model due to inclusion of correlations embraces dependencies between each of run-off triangles. Thanks to this extension of standard loss reserving technique it makes our estimates more precise and more corresponding to real observations. In this thesis apart from introduction of the model we also elaborate numerical example using different variations of this specific model and lastly we compare SUR based model with standard model for loss reserve estimation.

Keywords: Dependent run-off triangles; SUR model; non-life insurance

Obsah

Úvod	2
1 Vývojové trojúhelníky a odhad rezerv v neživotním pojištění	3
1.1 Klasická chain ladder metoda	4
1.2 Závislé vývojové trojúhelníky	6
2 Metodika SUR pro soustavy ekonometrických rovnic	8
3 Použití metodiky SUR pro závislé trojúhelníky	11
3.1 Specifikace modelu	11
3.2 Odhady v modelu GMCL	12
3.3 Porovnání odhadů v GMCL modelu	16
4 Aplikace na reálných datech	17
4.1 Vstupní data	17
4.2 Aplikace modelů	18
4.3 Porovnání modelů	21
Závěr	25
Seznam použité literatury	26
Seznam tabulek	27
A Příloha	28
A.1 Zdrojový kód spolu s výstupy	28

Úvod

Tato práce se věnuje tématu odhadu budoucích škod a výpočtu rezerv v neživotním pojištění. Obecně se rozlišují dva druhy škod, na které pojišťovna vytváří technické rezervy. Jedná se o rezervy typu RBNS („reported but not settled“) a IBNR („incurred but not reported“). Právě na druhý typ těchto rezerv se soustředíme v této práci. V případě IBNR rezerv se zabýváme, jak již z názvu vyplývá, škodami, o kterých pojišťovna zatím neví. Není znám skutečný počet škod, které již v minulosti vznikly, ale nebyly nahlášený. Z toho důvodu se z pohledu pojišťovny jedná o nějaké náhodné veličiny. Pro předstihu o budoucím vývoji se zavádí statistické modely, které slouží k odhadování takových škod a následnému výpočtu rezerv vytvářených k jejich pokrytí. Právě touto problematikou se zabýváme i my dále v práci.

V první části práce se věnujeme vymezení pojmů týkající se vstupních dat ve formě vývojových trojúhelníků a druhů rezerv, které je pojišťovna povinna udržovat. Dále uvádíme základní (klasický chain ladder) model pro výpočet rezerv typu IBNR, na který později navazujeme prostřednictvím zobecněného modelu. V navazující části je rozveden model SUR („seemingly unrelated regression“), který následně hraje nezastupitelnou roli v hlavní části práce jakožto klíčový prvek zobecněného modelu. Stěžejní kapitola práce se věnuje zavedení obecného modelu GMCL („generalized multivariate chain ladder“) pro odhady budoucích škod, který pro odhady využívá právě zmíněnou metodiku SUR. Narozdíl od klasického chain ladder modelu je tento nově odvozený model schopný díky rozšíření o SUR model zahrnovat do výpočtu také závislosti mezi vývojovými trojúhelníky. Díky tomu jsme schopni obdržet výsledky, které lépe odpovídají skutečným škodním průběhům.

V závěrečné části práce řešíme numerický příklad s reálnými vstupními daty, který využívá pro výpočet rezerv GMCL modelu. Naším cílem v této části je obdržet odhady škod a rezerv pomocí tohoto zobecněného modelu. Nakonec se zaměříme na porovnání různých variací GMCL modelu a klasického modelu využívaného pro odhady IBNR rezerv v neživotním pojištění.

1. Vývojové trojúhelníky a odhad rezerv v neživotním pojištění

Tato úvodní kapitola se věnuje popisu základních metod využívaných k odhadování rezerv v neživotním pojištění. Jde především o rezervy na pojistná plnění, které jsou stanovovány takovým způsobem, aby pojišťovna byla schopna v budoucnu vyplácet pojistná plnění z pojistných událostí, které proběhly v minulých obdobích, ale v době výpočtu nebyly tyto škody zatím nahlášený nebo zcela zlikvidovány.

Dvěma základními typy rezerv na pojistná plnění, které v neživotním pojištění rozeznáváme, jsou IBNR („incurred but not reported“) a RBNS („reported but not settled“). Prvně jmenovaný představuje v budoucnu vyplacená pojistná plnění za škody, které vznikly v minulosti, ale dosud nebyly pojišťovně nahlášený. Na druhou stranu RBNS se vztahuje na pojistná plnění za škody, které již byly nahlášený, nicméně ještě nebyly zlikvidovány. Zatímco rezerva RBNS se ve většině případů stanovuje na základě zkušeností likvidátora a jeho posouzení nastalé škodní události, rezervy IBNR se odhadují pomocí statistických metod založených na historických datech vztahujících se k obdobným pojistným událostem. Do výpočtů takových odhadů se v praxi také často zahrnují vnější vlivy ovlivňující samotnou výši rezervy, např. inflace.

Dále se text práce věnuje zejména samotnému odhadu rezerv s využitím statistických metod. Pro popis vstupních dat, která budeme dále analyzovat, zavádíme tzv. *vývojové trojúhelníky* (angl. *run-off triangles*), které v závislosti na předmětu zájmu vyjadřují proplacené nebo nahlášené pojistné nároky. Obecně se tedy z pohledu pojišťovny jedná o náklady, které očekává v budoucích obdobích.

Vývojové trojúhelníky definujeme jako horní trojúhelníkovou matici A , kde její řádky označují *rok vzniku* (angl. *accident year*) nebo-li rok, kdy došlo ke škodní události, a sloupce této matice symbolizují zpoždění od nahlášení dané události nebo-li *vývojové období* (angl. *development year*). Matici pak lze pro přehlednost schématicky zapsat do tabulky 1.1.

Období vzniku	Vývojové období						
	1	2	...	j	...	$n-1$	n
1	$a_{1,1}$	$a_{1,2}$...	$a_{1,j}$...	$a_{1,n-1}$	$a_{1,n}$
2	$a_{2,1}$	$a_{2,2}$...	$a_{2,j}$...	$a_{2,n-1}$	
⋮	⋮		⋱		⋱		
i	$a_{i,1}$	$a_{i,2}$...	$a_{i,j}$			
⋮	⋮		⋱				
$n-1$	$a_{n-1,1}$	$a_{n-1,2}$					
n	$a_{n,1}$						

Tabulka 1.1: Vývojový trojúhelník

Prvek $a_{i,j}$ tedy představuje výši škod vzniklých v období i , které byly zaplacený, resp. nahlášený, do konce vývojového období j . Obvykle se využívají dva druhy vývojových trojúhelníků. První z nich se nazývá *nekumulativní*, též *inkrementální*, kde jeho jednotlivé prvky značí hodnoty škod pouze za jedno konkrétní období. V tomto textu budeme využívat výhradně druhého typu, tzv. *kumulativního* trojúhelníku, ve kterém se hodnoty škod z jednotlivých vývojových let pro jeden rok vzniku postupně nasčítávají od prvního do požadovaného období. Prvek $a_{i,j}$ kumulativního trojúhelníku lze tedy získat jako součet prvků $b_{i,k}$ nekumulativního trojúhelníku předcházejících vývojových období.

$$a_{i,j} = \sum_{k=1}^j b_{i,k}$$

Pro odhady rezerv v neživotním pojištění se využívají dva základní odlišné přístupy. Prvním z nich je parametrický a druhým neparametrický model. Pro obě tyto skupiny jsou vhodná odlišná vstupní data. Zatímco v parametrických modelech je obvykle výhodnější pracovat s nekumulativními trojúhelníky, v neparametrických modelech, kterými se zabývá tato práce, se nejčastěji setkáváme se vstupními daty ve formě kumulativních vývojových trojúhelníků. V základních parametrických modelech se využívají např. gamma nebo logaritnicko-normální rozdělení, ale v praxi používanější je poněkud obecněji aplikovatelné GEV rozdělení (generalized extreme value distribution). Neparametrické modely jsou narozdíl od parametrických ve své podstatě méně náročné na výpočty, ale zobecnění těchto metod častokrát také využívají regrese, která může poměrně zkomplikovat proces odhadování technických rezerv z hlediska složitosti výpočtu. Jeden ze základních neparametrických modelů, sloužící k odhadování rezerv v neživotním pojištění, je představen v následující podkapitole, kde již uvažujeme pouze kumulativní vývojové trojúhelníky.

1.1 Klasická chain ladder metoda

Metoda chain ladder (někdy se lze setkat s překladem této metody jako tzv. stupňová metoda) je jeden z nejčastěji využívaných statistických přístupů sloužící k odhadování IBNR rezerv v pojišťovnictví. Jedná se o neparametrický model, který má dva předpoklady. Zaprvé musí platit, že jednotlivé inkrementální škody vzniklé v různých letech jsou na sobě vzájemně nezávislé. Druhým předpokladem je poté fakt, že očekáváme určité opakování obdobných schémat pozorovaných v minulosti i v budoucnosti. Samotný model je poměrně nenáročný z hlediska výpočetní techniky ve srovnání s jinými například parametrickými metodami. Tato výhoda je ale vykompenzována nevýhodou, která spočívá v citlivosti odhadů na změnu ve vstupních datech. I malá změna v pozorovaných datech může zásadně ovlivnit odhady rezerv zejména v nejnedávnějších škodních obdobích.

K samotnému odhadování s využitím chain ladder metody je potřebné hned ze začátku aplikovat uvažovaný předpoklad o přímé návaznosti minulých vývojových let na nadcházející roky a odvodit tzv. *vývojový faktor* (angl. *development*

factor), značíme f_j . Tento faktor představuje konstantu, pomocí níž je možné přecházet od jednoho vývojového období ke druhému. Jeho význam tedy prakticky spočívá v určení daného vývojového schématu, pomocí kterého následně tvoříme odhady budoucích škod. Odhad vývojového faktoru \hat{f}_j pak má tvar poměru škod dvou po sobě jdoucích vývojových let v daném roku vzniku, kde stříškou označujeme zde i dále v textu nějaký odhad dané neznámé. Explicitní vyjádření \hat{f}_j lze zapsat ve tvaru

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j} a_{i,j}}{\sum_{i=1}^{n-j} a_{i,j-1}}, \quad j \in \{2, \dots, n\}.$$

Samotné vývojové faktory jsou pak dále využívány k odvozování neznámých hodnot představující budoucí kumulativní výše škod. Vývojový trojúhelník z tabulky 1.1 tedy postupně doplňujeme těmito očekávanými úhrny škod na obdélník. Výše škod vzniklých v roce i s uplynulými roky od vzniku škod j se odhaduje jako

$$\hat{a}_{i,j} = a_{i,n-i+1} \cdot \prod_{l=n-i+2}^j \hat{f}_l, \quad i \in \{2, \dots, n\}, j \in \{n-j+2, \dots, n\},$$

kde $\hat{a}_{i,j}$ tedy vyjadřuje odhad budoucího kumulativního úhrnu škod, neboli nároků na pojistná plnění. V následující tabulce 1.2 je znázorněno, jak vypadá kumulativní trojúhelník po doplnění odhadovanými hodnotami.

	Vývojové období						
	1	2	...	j	...	$n-1$	n
Období vzniku							
1	$a_{1,1}$	$a_{1,2}$...	$a_{1,j}$...	$a_{1,n-1}$	$a_{1,n}$
2	$a_{2,1}$	$a_{2,2}$...	$a_{2,j}$...	$a_{2,n-1}$	
⋮	⋮		⋱		⋱		⋮
i	$a_{i,1}$	$a_{i,2}$...	$a_{i,j}$...	$\hat{a}_{i,n}$
⋮	⋮		⋱		⋱		⋮
$n-1$	$a_{n-1,1}$	$a_{n-1,2}$					$\hat{a}_{n-1,n}$
n	$a_{n,1}$...		$\hat{a}_{n,n-1}$	$\hat{a}_{n,n}$

Tabulka 1.2: Vývojový trojúhelník doplněný na obdélník

Navíc pro zjednodušení takového postupu je dále vhodné uvažovat ukončení vývoje škod po n letech.

S doplněnou maticí (vývojovým trojúhelníkem) jsme schopni zpětně vypočítat potřebné rezervy na škody k danému roku vzniku i . Výsledná odhadovaná výše rezervy je pak dána vztahem

$$\hat{R}_i = \hat{a}_{i,n} - a_{i,n-i+1}, \quad i \in \{2, \dots, n\}.$$

Odhad rezervy pro celé portfolio pak můžeme stanovit jako součet všech jednotlivých odhadů rezerv vztahených k danému období vzniku. Tedy dostáváme

celkovou hodnotu rezervy pro vývojový trojúhelník

$$\widehat{R} = \sum_{i=2}^n \widehat{R}_i.$$

Základním problémem tohoto přístupu je, že využívá pro všechny roky vzniku totožný vývojový faktor pro daný vývojový rok. To však často neodpovídá realitě a v důsledku toho dochází k někdy poměrně zásadním odklonům od skutečnosti. Z toho důvodu se vytvářejí různé modifikace a zobecnění chain ladder metody, jako je například obecná vícerozměrná chain ladder metoda, které umožňují větší různorodost při samotném odhadování. Tyto komplexnější modely kromě jiného často zohledňují i některé vztahy mezi daty, konkrétněji jejich korelace.

1.2 Závislé vývojové trojúhelníky

V této podkapitole obecně uvažujeme více kumulativních vývojových trojúhelníků a zabýváme se závislostmi, nebo-li korelacemi, mezi nimi.

V chain ladder modelu, který byl představen v kapitole 1.1, jsme uvažovali, že se využívané vývojové trojúhelníky vztahují k jednomu *pojistnému odvětví* (*angl. line of business*). V takovém trojúhelníku se pak zároveň předpokládá, že jednotlivé roky vzniku jsou na sobě nezávislé. Při uvažování více trojúhelníků, pro které vytváříme společný odhad rezerv je však situace jiná. V praxi není běžně pozorovaný jev, aby jednotlivé trojúhelníky mezi sebou neměli žádný vztah. To je důsledkem vnějších jevů jako je například inflace, které ovlivňují všechny škody napříč pojistnými odvětvími. Jestliže uvažujeme portfolio složené z několika vývojových trojúhelníků pro více pojistných odvětví, je zásadní rozdíl mezi odhadem vzniklým metodou chain ladder pro celé portfolio a mezi součtem odhadů pro jednotlivá pojistná odvětví. Zmíněné externí vlivy popisujeme jako závislosti mezi vývojovými trojúhelníky a k jejich matematické interpretaci se využívá právě již předeslaná korelace.

Pro ilustraci si můžeme představit dva různé vývojové trojúhelníky, přičemž první z nich označme P a druhý I . P představuje trojúhelník zaplacených a I nahlášených škod. Dále uvažujme P/I poměr, nebo-li poměr zaplacených a evidovaných škod. Jak již bylo uvedeno v předešlém odstavci, jestliže nyní aplikujeme chain ladder metodu na tyto trojúhelníky odděleně a následně odhady rezerv sečteme, dostaneme rozdílný výsledek od dat, které bychom na základě zkušeností očekávali. Důvodem je, že odhadnuté hodnoty P/I zůstávají konstantní v oddělených výpočtech metody chain ladder a díky tomu se samotný odhad odklání od námi očekávaného výsledku. Z tohoto příkladu tedy vyplývá, že je nutné uvažovat určitý druh závislosti mezi vývojovými trojúhelníky.

Pro zohlednění výše zmíněných vztahů mezi vývojovými trojúhelníky se zavádí modely rozšiřující klasický chain ladder model. Mezi ně patří model GMCL, kterému je věnována převážná část této práce, nebo mnichovský chain ladder model, jejichž součástí je i regresní model odhadující korelace mezi jednotlivými

trojúhelníky. Posledně zmiňovaný model využívá reziduí, jež navíc umožňují pracovat se všemi pozorovanými daty. Bez nich by pro poslední vývojová období nemusel být dostatek použitelných dat pro vytvářené odhady.

V problematice závislosti vývojových trojúhelníků rozeznáváme kromě korelací mezi jednotlivými trojúhelníky také korelace v rámci samotného trojúhelníku. Ty se projevují například ve vztazích mezi jednotlivými roky vzniku i . Příčinou znatelnějších rozdílů a silnějších závislostí tohoto typu může být např. zásadnější změna v ekonomickém prostředí, což také může ovlivnit i výši rezerv pro jednotlivá období.

Samotná závislost mezi trojúhelníky (různými pojistnými odvětvími), která souvisí s rozložením rizika, je velmi důležitá vzhledem ke stanovení solventnostního kapitálového požadavku. Ten je pojišťovna povinna za zákona udržovat, aby neztratila schopnost dostávat svým závazkům. Na druhou stranu cílem pojišťovny je nedržet nadbytek kapitálu, protože kapitálem jsou myšleny zejména peníze a peněžní ekvivalenty, které obecně nenesou vysoký výnos. V praxi nejspíše nejběžnějším příkladem příčiny korelací mezi trojúhelníky je inflace.

2. Metodika SUR pro soustavy ekonometrických rovnic

Tato kapitola se zabývá popisem metody, která umožňuje do odhadování škod zahrnout korelace mezi závislými vývojovými trojúhelníky. Obecně uvažujeme model s několika lineárními rovnicemi, které mohou představovat například soustavu ekonometrických rovnic, tedy rovnic udávající jistý vztah mezi ekonomickými veličinami. V takovém modelu nelze za reálných podmínek předpokládat, že chybové členy budou nekorelované. Systém uvažovaných lineárních rovnic na první pohled může sice vypadat nepropejeně, tedy není prvotně zřejmý žádný vztah mezi jednotlivými rovnicemi. Ve skutečnosti ale existuje propojení ve formě korelací mezi jednotlivými chybovými členy. Z toho důvodu se metodika pro výpočet výše popsanych rovnic, dále je chápeme jako lineární regresní rovnice, nazývá *SUR* (*seemingly unrelated regression*) nebo-li „zdánlivě nesouvisející regrese“.

Uvažujme model skládající se z $m \geq 2$ lineárních regresních rovnic, kde i -tou rovnicí lze vyjádřit ve tvaru

$$\mathbf{y}_i = X_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, m. \quad (2.1)$$

V takovém zápise \mathbf{y}_i představuje n -rozměrný vektor *závislých proměnných* neboli *odezvu* i -tého závislého vektoru, X_i je matice *regresorů* typu $n \times k_i$ s plnou hodnotou složená z pozorování k_i nezávislých náhodných veličin, $\boldsymbol{\beta}_i$ je k_i -rozměrný vektor *regresních koeficientů* a nakonec $\boldsymbol{\varepsilon}_i$ je n -rozměrný vektor, který představuje tzv. *chybový člen*. Pro chybový člen dále předpokládáme

$$E\boldsymbol{\varepsilon}_i = \mathbf{0}$$

$$E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T) = \sigma_{i,i} I_n, \quad i = 1, 2, \dots, m,$$

kde $\sigma_{i,i}$ označuje předpokládaný rozptyl poruchy (chybového členu) v i -té regresní rovnici a I_n jednotkovou matici typu $n \times n$.

Z důvodu uvažovaného vztahu mezi jednotlivými rovnicemi předpokládáme také, že

$$E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^T) = \sigma_{i,j} I_n, \quad i, j = 1, 2, \dots, m,$$

kde $\sigma_{i,j}$ představuje kovarianci mezi chybovými členy i -té a j -té rovnice pro každou z pozorovaných rovnic.

Nakonec můžeme vhodněji přepsat model (2.1) pomocí následujícího maticového vyjádření, do tvaru

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_m \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_m \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix}, \quad (2.2)$$

což lze zkráceně zapsat jako

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.3)$$

kde \mathbf{y} je mn -rozměrný vektor, X je typu $mn \times K$, $\boldsymbol{\varepsilon}$ je opět mn -rozměrný vektor a $K = \sum_i k_i$.

Předpoklady takového modelu (2.3) pro všech m regresních rovnic lze zapsat ve tvaru

$$E\boldsymbol{\varepsilon} = \mathbf{0}$$

$$\text{cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \begin{pmatrix} \sigma_{1,1}I_n & \dots & \sigma_{1,m}I_n \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}I_n & \dots & \sigma_{m,m}I_n \end{pmatrix} = \Sigma \otimes I_n = \Psi,$$

kde \otimes značí *Kroneckerův součin* matic, který je pro dvě matice, A typu $p \times q$ a B typu $r \times s$, definován jako

$$A \otimes B = (a_{i,j}B) = \begin{pmatrix} a_{1,1}B & \dots & a_{1,q}B \\ \vdots & \ddots & \vdots \\ a_{p,1}B & \dots & a_{p,q}B \end{pmatrix}.$$

Výsledkem je pak matice typu $pr \times qs$. V našem předchozím vyjádření $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)$ dále máme matici Σ typu $m \times m$, která je symetrická pozitivně definitní a jejíž prvek na místě (i,j) je $\sigma_{i,j}$. Z definice Kroneckerova součinu matic pak dostáváme matici Ψ , jež je typu $mn \times mn$.

Závěr této kapitoly je věnován samotnému výpočtu regresních rovnic. Běžná metoda pro odhadování koeficientů v regresním modelu spočívala ve vyčíslování jednotlivých rovnic zvlášť, to ale, jak již bylo zmíněno dříve v textu, vede k zásadním odlišnostem od výpočtu, ve kterém odhadujeme více rovnic najednou. Tento jev nastává díky korelaci mezi regresními rovnicemi, které jsou korelované prostřednictvím chybových členů, což bylo ukázáno v [6]. Ve stejné práci bylo také popsáno, že pro přesnější odhady, které lépe popisují hodnoty nastávající v budoucnu, je smysluplné využívat sdružené odhady nebo-li odhadují se všechny regresní rovnice společně, niko-li jednotlivě.

Metody nejčastěji využívané k odhadování koeficientů v SUR modelu jsou založené na *metodě nejmenších čtverců* (angl. *least squares method*). Existuje několik alternativ pro volbu postupu výpočtu. Uvedeme několik z nich, přičemž bereme v úvahu regresní model (2.1).

Prvně uvedeme GLS (*generalized least squares*) metodu, ve které regresní koeficienty odhadneme pomocí vztahu, ve kterém zohledňujeme kovarianční matici chybových členů Ψ . Odhad je v takovém případě následujícího tvaru

$$\hat{\boldsymbol{\beta}}_G = (X^T(\Sigma^{-1} \otimes I_n)X^{-1})^{-1}X^T(\Sigma^T \otimes I_n)\mathbf{y} = (X^T\Psi^{-1}X)^{-1}X^T\Psi^{-1}\mathbf{y}. \quad (2.4)$$

Kovarianční matici V odhadu $\hat{\boldsymbol{\beta}}_G$ pak dostaneme jako

$$V(\hat{\boldsymbol{\beta}}_G) = (X^T\Psi^{-1}X)^{-1}.$$

Problémem přístupu GLS je, že počítá se známou maticí Σ . To může být velmi problematické ve výpočtech s reálnými daty, jelikož kovarianční matice není většinou dopředu dána. Z tohoto důvodu se zavádí FGLS (*feasible generalized least squares*) metoda, která narozdíl od GLS využívá ve výpočtu odhad kovarianční matice Σ a je tedy lépe použitelná v praxi. Odhad regresního koeficientu $\hat{\beta}_{FG}$ lze psát následně jako

$$\hat{\beta}_{FG} = (X^T(\hat{\Sigma} \otimes I_n)^{-1}X)^{-1}X^T(\hat{\Sigma} \otimes I_n)^{-1}\mathbf{y}. \quad (2.5)$$

$\hat{\Sigma}$ ve výše uvedeném zápisu představuje matici odhadů $\hat{\sigma}_{i,j}$ kovariancí chybových členů $\sigma_{i,j}$. Taková matice je regulární a za $\hat{\sigma}_{i,j}$ můžeme volit nějaký námi vybraný odhad, nejčastěji konzistentní. Jednou z možností je například jako $\hat{\Sigma}$ označit odhad, který má na pozici (i, j) prvek

$$\hat{\sigma}_{i,j} = \frac{1}{n} \tilde{\mathbf{e}}_i^T \tilde{\mathbf{e}}_j,$$

kde $\tilde{\mathbf{e}}_i$ získáme jako residua v i -té regresní rovnici aplikací klasické OLS (*ordinary least squares*) metody, která je dalším typem odhadu pomocí metody nejmenších čtverců. Narozdíl od přechozích metod avšak u metody OLS neuvažujeme žádné korelace mezi jednotlivými rovnicemi.

Výhodou této metodiky je, že odhady získané výše popsanými způsoby jsou konzistentní. Jestliže navíc předpokládáme normalitu dat obdržíme zároveň odhady maximálně věrohodné.

3. Použití metodiky SUR pro závislé trojúhelníky

V této kapitola je představen model *General multivariate chain ladder*, dále zkráceně GMCL. Jedná se o zobecnění klasického multivariate chain ladder modelu, které umožňuje do výpočtu zahrnout závislosti mezi vývojovými trojúhelníky, tedy jejich korelace, a to prostřednictvím metodiky SUR. Konkrétní námi popisovaný model je převzatý z textu [8].

3.1 Specifikace modelu

Pro potřeby této kapitoly se uvažuje N vývojových trojúhelníků o stejných rozměrech. Lze poté označit N -rozměrný vektor $\mathbf{Y}_{i,k} = (Y_{i,k}^{(1)}, \dots, Y_{i,k}^{(N)})$, který představuje kumulativní hodnoty škod vztahované k roku vzniku i a vývojovému roku k ve všech trojúhelnících $1, 2, \dots, N$. Model dále předpokládá, že se pracuje se stejným počtem let vzniku a vývojových let, tedy $i, k \in \{1, \dots, L\}$ a že jednotlivé kumulativní škody $Y_{i,k}^{(j)}$ jsou kladné pro každé $j \in \{1, \dots, N\}$.

Pro potřebu dalších výpočtů označíme množinu všech pozorovaných škod $D = \{\mathbf{Y}_{i,k}; i + k \leq L + 1, 1 \leq i \leq L, 1 \leq k \leq L\}$ a množinu škod příslušnou roku vzniku i přes vývojové roky $\{1, \dots, k\}$ jako $D_{i,k} = \{\mathbf{Y}_{i,j}; j \leq k\}$. V neposlední řadě zavedeme $(L - k + 1)$ -rozměrný vektor $\mathbf{Y}_{\leq, k}^{(n)} = (Y_{1,k}^{(n)}, \dots, Y_{L-k+1,k}^{(n)})$ prvních $(L - k + 1)$ škod ve vývojovém roce k v n -tém vývojovém trojúhelníku. Podobně označíme $(L - k)$ -rozměrný vektor $\mathbf{Y}_{<, k}^{(n)} = (Y_{1,k}^{(n)}, \dots, Y_{L-1,k}^{(n)})$ prvních $(L - k)$ škod. Nakonec pro model, podobně jako v modelu (2.1), předpokládáme, že chybové členy $\boldsymbol{\varepsilon}_{i,k}$ za daných napozorovaných hodnot $D_{i,k}$ mají nulovou střední hodnotu a kovarianci dānu předpisem

$$\text{cov}(\boldsymbol{\varepsilon}_{i,k} | D_{i,k}) = \text{diag}(\mathbf{Y}_{i,k})^{1/2} \cdot \Sigma_k \cdot \text{diag}(\mathbf{Y}_{i,k})^{1/2},$$

kde uvažujeme, že škody vzniklé v jednotlivých letech jsou nezávislé, $\boldsymbol{\varepsilon}_{i,k}$ mají symetrické rozdělení, přičemž

$$\Sigma_k = \begin{pmatrix} \sigma_{1,1} & \dots & \sigma_{1,N} \\ \vdots & \ddots & \vdots \\ \sigma_{N,1} & \dots & \sigma_{N,N} \end{pmatrix}$$

je symetrická pozitivně definitní matice typu $N \times N$. Pro n -rozměrný vektor \mathbf{v} představuje $\text{diag}(\mathbf{v})$ diagonální matici typu $n \times n$, kde na její diagonále jsou právě prvky vektoru \mathbf{v} .

Na základě výše uvedených předpokladů a označení zavádíme následující model, který udává vztah mezi vývojovými obdobími k a $(k + 1)$

$$\mathbf{Y}_{i,k+1} = B_k \mathbf{Y}_{i,k} + \boldsymbol{\varepsilon}_{i,k}. \quad (3.1)$$

V tomto modelu B_k představuje čtvercovou matici řádu N vztahující se k vývojovému roku k , která má na pozici (n, j) vývojový faktor $\beta_{n,j}$, tj. vývojový faktor v j -tém roku vzniku a n -tém vývojovém trojúhelníku. Prvek $\varepsilon_{i,k}$ je chybovým členem pro daný rok vzniku a vývojový rok. Jestliže uvažujeme libovolnou matici B_k , která není diagonální, pak jednotlivé škody v rámci jednoho vývojového trojúhelníku ve vývojovém roce k přímo závisí na škodách ostatních $(N - 1)$ trojúhelníku ze stejného období.

3.2 Odhady v modelu GMCL

Ze struktury modelu (3.1) lze usoudit, že žádná odezva $Y_{i,k+1}^{(n)}$ nevystupuje v dalších rovnicích jako nezávislá proměnná. Jednotlivé rovnice pro dané vývojové trojúhelníky jsou pak tedy provázány pouze prostřednictvím nenulové korelace chybových členů. Z toho důvodu je smysluplné pro odhady parametrů modelu (3.1) využívat metodiku SUR, která tento předpoklad také zahrnuje. Pro použití metodiky SUR je nicméně nutné model upravit, což je ukázáno v následující sekci. Dále se v této podkapitole zabýváme odhadem parametrů pro vývojový rok $(k + 1)$, podobně lze odvodit i odhad pro libovolné vývojové období.

Model (3.1) lze nejprve přepsat do tvaru

$$\begin{pmatrix} \mathbf{Y}_{\leq, k+1}^{(1)} \\ \mathbf{Y}_{\leq, k+1}^{(2)} \\ \vdots \\ \mathbf{Y}_{\leq, k+1}^{(N)} \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & X_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}. \quad (3.2)$$

Matice $X_n = (\mathbf{Y}_{<, k}^{(1)}, \dots, \mathbf{Y}_{<, k}^{(N)})$ typu $(L - k) \times N$ označuje prvních $(L - k)$ pozorování ve vývojovém roce k z každého vývojového trojúhelníku a platí $X_1 = X_2 = \dots = X_N$. Dále $\beta_n = (\beta_{n,1}, \dots, \beta_{n,N})$ je N -rozměrný vektor vývojových faktorů pro n -tý trojúhelník. ε_n je pak vektor délky $(L - k)$, který představuje chybový člen v n -té rovnici.

Model (3.2) lze dále zkráceně zapsat následovně

$$\mathbf{Y} = X\beta + \varepsilon, \quad (3.3)$$

kde $\mathbf{Y} = \text{vec}(\mathbf{Y}_{\leq, k+1}^{(1)}, \dots, \mathbf{Y}_{\leq, k+1}^{(N)})$ je $N(L - k)$ -rozměrný vektor závislých proměnných, přičemž $\text{vec}()$ je operátor vektorizace matice, tedy pro $A = (a_{i,j})$ typu $m \times n$ je $\text{vec}(A) = (a_{1,1}, \dots, a_{m,1}, a_{1,2}, \dots, a_{m,n})$ vektor délky mn . V modelu (3.3) dále vystupuje matice X , která má rozměr $N(L - k) \times N^2$. Tato matice je blokově diagonální a lze ji zapsat pomocí Kroneckerova součinu jako $X = I_N \otimes X_n$. Na diagonále takto získané matice pak vystupují právě matice X_1, \dots, X_N . Podobně jako v případě odezvy \mathbf{Y} vektor $\beta = \text{vec}(\beta_1, \dots, \beta_N)$ je vektorizace vývojových faktorů, jde o vektor délky NN , a $\varepsilon = \text{vec}(\varepsilon_1, \dots, \varepsilon_N)$ je opět vektorizace, tentokrát chybových členů, vytvářející vektor délky $N(L - k)$.

Model (3.3) má, jak již víme z podkapitoly (3.1), nulovou střední hodnotu chybových členů a navíc kovarianci

$$\text{cov}(\boldsymbol{\varepsilon}) = \text{diag}(\mathbf{W})^{1/2}(\Sigma_k \otimes I_{L-k})\text{diag}(\mathbf{W})^{1/2},$$

kde \mathbf{W} označuje $N(L-k)$ -rozměrný vektor $\text{vec}(\mathbf{Y}_{<,k}^{(1)}, \dots, \mathbf{Y}_{<,k}^{(N)})$ prvních $(L-k)$ pozorovaných škod ve vývojovém období k ve všech N trojúhelnících. Abychom obdrželi potřebnou strukturu kovarianční matice, podobnou kovarianci uvažované v modelu (2.1), využijeme ekvivalentní úpravy a vynásobíme obě strany rovnice faktorem $\text{diag}(\mathbf{W})^{-1/2}$. Dostáváme vztah

$$\text{diag}(\mathbf{W})^{-1/2}\mathbf{Y} = \text{diag}\mathbf{W}^{-1/2}X\boldsymbol{\beta} + \text{diag}(\mathbf{W})^{-1/2}\boldsymbol{\varepsilon}.$$

Jednotlivé členy označíme následujícím způsobem

$$\mathbf{Y}^* = \text{diag}(\mathbf{W})^{-1/2}\mathbf{Y},$$

$$X^* = \text{diag}(\mathbf{W})^{-1/2}X,$$

$$\boldsymbol{\varepsilon}^* = \text{diag}(\mathbf{W})^{-1/2}\boldsymbol{\varepsilon}.$$

Získáváme tak model s přeznačenými členy \mathbf{Y}^* , X^* , $\boldsymbol{\varepsilon}^*$

$$\mathbf{Y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*. \quad (3.4)$$

Model (3.4) má pak kovarianční strukturu chybových členů

$$\text{cov}(\boldsymbol{\varepsilon}^*) = \text{diag}(\mathbf{W})^{-1/2}\text{cov}(\boldsymbol{\varepsilon})\text{diag}(\mathbf{W})^{-1/2} = \Sigma \otimes I_{L-k},$$

což je požadovaná struktura kovarianční matice, které jsme chtěli dosáhnout. Takto získaný model pak totiž odpovídá modelu (2.1), jelikož splňuje potřebné předpoklady SUR modelu. Jinými slovy, struktura kovarianční matice chybových členů je shodná se strukturou v modelu (2.1). Na model (3.4) již tedy lze využít metody odhadu pro běžný model SUR, a proto pro výpočty nevyužíváme původní model (3.1), ale právě nyní odvozený model (3.4).

Nyní uvedeme několik způsobů, jak získat odhad pro SUR model (3.4) odvozený v předcházející úvaze. Základem všech těchto metod výpočtu je opět metoda nejmenších čtverců.

Nejjednodušším odhadem vývojových faktorů je odhad metodou OLS, který zanedbává korelace mezi chybovými členy. Metoda proto nedává nejpřesnější výsledky, jestliže jsou ve skutečnosti korelace chybových členů nezanedbatelné. Výsledný odhad lze zapsat jako

$$\hat{\boldsymbol{\beta}}_O = ((X^*)^T X^*)^{-1}(X^*)^T \mathbf{Y}^*. \quad (3.5)$$

Jde tedy prakticky o přímou aplikaci metody nejmenších čtverců na model (3.4). Takto obdrženy odhad $\hat{\boldsymbol{\beta}}_O$ je pak nestranný.

Další možností jak odhadovat vývojový faktor je odhad metodou GLS popsany v kapitole 2. Důležitým předpokladem pro výpočet tohoto odhadu je znalost

kovarianční matice chybových členů Σ_k . Vzorec pro tento odhad (2.4) lze přepsat pro model (3.4) do tvaru

$$\widehat{\beta}_G = ((X^*)^T(\Sigma_k \otimes I_{L-k})^{-1}X^*)^{-1}(X^*)^T(\Sigma_k \otimes I_{L-k})^{-1}Y^*. \quad (3.6)$$

Kovarianční matice odhadu je dána vztahem $V(\widehat{\beta}_G) = ((X^*)^T(\Sigma_k \otimes I_{L-k})^{-1}X^*)^{-1}$. Podobně jako v případě $\widehat{\beta}_O$, je odhad $\widehat{\beta}_G$ nestranným odhadem parametru β .

Nyní předpokládejme, že matice Σ_k není známá. V takovém případě, jak bylo zmíněno v kapitole 2, lze využít metody FGLS, která pracuje s odhadnutou maticí kovariancí chybových členů. Vzorec (2.5) pro odhad získaný FGLS metodou aplikovaný na model (3.4) je pak

$$\widehat{\beta}_{FG} = ((X^*)^T(\widehat{\Sigma}_k \otimes I_{L-k})^{-1}X^*)^{-1}(X^*)^T(\widehat{\Sigma}_k \otimes I_{L-k})^{-1}Y^*. \quad (3.7)$$

Existuje řada potenciálně použitelných odhadů matice Σ_k . Kromě odhadu zmíněného v kapitole 2 lze využít například odhad uvedený v práci [8], který je konzistentním, ale nikoli nestranným odhadem Σ_k . Odhad matice Σ_k pak nabývá tvaru

$$\widehat{\Sigma}_O = \frac{1}{L-k}(\widehat{\epsilon}_1^*, \dots, \widehat{\epsilon}_N^*)^T(\widehat{\epsilon}_1^*, \dots, \widehat{\epsilon}_N^*),$$

kde $\widehat{\epsilon}_n^*$ je $(L-k)$ -rozměrný vektor reziduí, získaný aplikací OLS metody na regresi model tvořený pouze z n -té rovnice původního modelu (3.4).

Jak již bylo několikrát zmíněno, metoda FGLS vyžaduje výpočet odhadu matice Σ_k , jelikož původní matice Σ_k není známá. Kvůli tomuto přidanému kroku se zvyšuje variabilita metody odhadu vektoru vývojových faktorů β . Ve snaze vylepšit odhad $\widehat{\beta}_{FG}$ se zavádí IFGLS (*iterative feasible generalized least square*) metoda odhadu, která má za cíl zvýšit eficienci odhadu. Nestranný odhad $\hat{\theta}$ parametru θ nazveme *eficientní*, jestliže splňuje $\frac{1}{F(\theta)} = \text{var}(\hat{\theta})$. $F(\theta)$ označuje Fisherovu informaci definovanou jako

$$\int \left(\frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 f(x; \theta) dx,$$

kde f je nějaká věrohodnostní funkce. V prvním kroku iterace nahradíme Σ_k za I_{L-k} , tak dostaneme odhad $\widehat{\beta}_O$ metodou OLS vektoru β . Následně odvodíme matici $\widehat{\Sigma}_k$ z reziduí založených na metodě OLS. Takto získaný odhad $\widehat{\Sigma}_k$ lze pak využít pro výpočet metodou FGLS. Podobně lze pokračovat i dalším krokem, ve kterém vezmeme rezidua z FGLS metody a pomocí nich vytvoříme opět odhad matice Σ_k . Tento proces lze aplikovat opakovaně a výslednému odhadu pak říkáme IFGLS odhad. Formálně pak definuje odhad v i -tém kroku předpisem

$$\widehat{\beta}_I^i = ((X^*)^T(\widehat{\Sigma}_k^i \otimes I_{L-k})^{-1}X^*)^{-1}(X^*)^T(\widehat{\Sigma}_k^i \otimes I_{L-k})^{-1}Y^*, \quad (3.8)$$

kde $\widehat{\Sigma}_k^i$ je konzistentní odhad Σ_k vytvořený z reziduí založených na odhadu $\widehat{\beta}_I^{i-1}$. Speciálně označíme odhad metodou OLS jako $\widehat{\beta}_I^0$, který využívá matice $\widehat{\Sigma}_k^0 = I_{L-k}$. Zbylé iterace $\widehat{\beta}_I^i$ jsou pak založené výhradně na FGLS metodě.

Pomocí odhadů vývojových faktorů lze vytvořit odhad celkových očekávaných škod užitím rekurzivního výpočtu aplikovaného přímo na původní GMCL model (3.1). Pomocí tohoto postupu dostaneme celkové hodnoty škod pocházejících z roku vzniku i za užití pozorovaných úhrnů škod $Y_{i,L-i+1}$ jako

$$\mathbf{Y}_{i,L} = \left(\prod_{k=L-1}^{L-i+1} B_k \right) \mathbf{Y}_{i,L-i+1} + \sum_{k=L-i+1}^{L-i} \left(\prod_{i=L-1}^{k+1} B_i \right) \boldsymbol{\varepsilon}_{i,k}, \quad (3.9)$$

kde součin $\prod_{k=L-1}^L B_k$ je definován jako identická matice. Za platnosti předpokladu nulové střední hodnoty chybových členů za daných napozorovaných dat a podmíněné kovariance odpovídající modelu (3.1) lze ukázat, že střední hodnota škod podmíněná napozorovanými hodnotami realizovaných škod je rovna $\left(\prod_{k=L-1}^{L-i+1} B_k \right) \mathbf{Y}_{i,L-i+1}$. Pro odhad očekávané celkové hodnoty škod využijeme odhad matice vývojových faktorů \widehat{B}_k a dostáváme

$$\widehat{\mathbf{Y}}_{i,L} = \left(\prod_{k=L-1}^{L-i+1} \widehat{B}_k \right) \mathbf{Y}_{i,L-i+1} \quad (3.10)$$

Práce [8] v lemmatu 1 shrnuje některé základní vlastnosti tohoto odhadu. Podle tohoto tvrzení platí, že $E(\widehat{B}_k) = B_k$, tj. odhady matic vývojových faktorů jsou navzájem nekorelované a odhad $\widehat{\mathbf{Y}}_{i,L}$ je nestranným odhadem $E(\mathbf{Y}_{i,L})$.

Je zřejmé, že GMCL model umožňuje poměrně velké množství různých postupů pro odhadování škod. Největší výhodou GMCL modelu oproti klasickému chain ladder modelu však není pouze velká variabilita použitelných odhadů, ale zejména možnost model upravovat takovým způsobem, aby se předešlo přeceňování velkých hodnot kumulativních škod a na druhou stranu umožnilo dát větší důležitost škodám relativně malým. Příčinou těchto problémů v klasickém modelu je, že se neuvažují body, ve kterých je protnuta osa y . Takový bod budeme dále nazývat intercept. V různých variacích chain ladder modelu může být poměrně obtížné přidat intercepty, ale za využití SUR modelu je toto rozšíření poměrně nativní.

Označíme $\mathbf{A}_k = (\beta_{1,0}, \dots, \beta_{N,0})$ N -rozměrný vektor interceptů, kde $\beta_{n,0}$ je intercept n -tého trojúhelníku v k -tém vývojovém roce. Nyní můžeme přidáním tohoto vektoru \mathbf{A}_k doplnit existující model (3.1) na model s intercepty

$$\mathbf{Y}_{i,k+1} = \mathbf{A}_k + B_k \mathbf{Y}_{i,k} + \boldsymbol{\varepsilon}_{i,k}, \quad (3.11)$$

ve kterém nadále platí stejné předpoklady jako v modelu (3.1). Samotný výpočet pomocí modelu využívající metodiky SUR zůstává stejný, pouze je nutné upravit definice matic X_n a vektorů $\boldsymbol{\beta}_n$ tak, aby zahrnovaly intercepty. Pro model s intercepty definujeme rovností $X_n = (\mathbf{1}, \mathbf{Y}_{<,k}^{(1)}, \dots, \mathbf{Y}_{<,k}^{(N)})$ matici pozorování typu $(L-k) \times (N+1)$, kde $\mathbf{1}$ je $(L-k)$ -rozměrný vektor $(1, 1, \dots, 1)$. $\boldsymbol{\beta}_n$ je $(N+1)$ -rozměrný vektor $(\beta_{n,0}, \beta_{n,1}, \dots, \beta_{n,N})$ odhadovaných parametrů n -tého vývojového trojúhelníku. Dále již postupujeme obdobně jako v případě výpočtu odhadu vektoru $\boldsymbol{\beta}$ bez uvažování interceptů. S napočítanými odhady vektorů $\boldsymbol{\beta}$ nakonec již můžeme přejít k výpočtu samotných očekávaných celkových úhrnů škod. Pro výpočet tedy využijeme postup popsany v (3.10), nejprve je ale nutné

upravit model (3.11) tak, abychom mohli daný přístup využít bez porušení předpokladů. Model s interceptry (3.11) lze přepsat jako

$$\begin{pmatrix} 1 \\ \mathbf{Y}_{i,k+1} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{A}_k & B_k \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{Y}_{i,k} \end{pmatrix} + \begin{pmatrix} 0 \\ \boldsymbol{\varepsilon}_{i,k} \end{pmatrix}.$$

Nyní můžeme vhodně označit jednotlivé prvky soustavy rovnic a přepsat je do tvaru požadovaného modelu

$$\mathbf{Z}_{i,k} = \begin{pmatrix} 1 \\ \mathbf{Y}_{i,k} \end{pmatrix}, \quad E_k = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{A}_k & B_k \end{pmatrix}, \quad \mathbf{e}_{i,k} = \begin{pmatrix} 0 \\ \boldsymbol{\varepsilon}_{i,k} \end{pmatrix},$$

$$\mathbf{Z}_{i,k+1} = E_k \mathbf{Z}_{i,k} + \mathbf{e}_{i,k}. \quad (3.12)$$

Obdrželi jsme tedy tvar potřebný pro rekurzivní výpočet celkové očekávané hodnoty škod, který ale navíc zahrnuje i interceptry. Můžeme je nyní zapsat pomocí obdobného vzorce jako (3.10). Dostáváme tedy výsledný odhad

$$\widehat{\mathbf{Z}}_{i,L} = \left(\prod_{k=L-1}^{L-i+1} \widehat{E}_{i,L-i+1} \right) \mathbf{Z}_{i,L-i+1}. \quad (3.13)$$

Jestliže z odhadu $\widehat{\mathbf{Z}}_{i,L}$ odebereme první prvek vektoru, tedy 1, obdržíme odhad celkových očekávaných škod $\widehat{\mathbf{Y}}_{i,L}$. Celý výše popsany postup navíc nijak neovlivňuje vlastnosti odhadu, jelikož do soustavy nebyly přidány žádné náhodné proměnné. Stále se tedy jedná o nestranný odhad očekávaných celkových úhrnů škod.

3.3 Porovnání odhadů v GMCL modelu

V této podkapitole se zabýváme porovnáním odhadů získaných pomocí metod OLS, GLS a FGLS pro GMCL model.

Nejprve předpokládejme, že matice Σ_k je známá. V takovém případě je nejlepším lineárním nestranným odhadem $\widehat{\boldsymbol{\beta}}_G$, odhad získaný metodou GLS. Tento fakt vyplývá z Gauss-Markovovy věty (viz [2]), konkrétněji její modifikace, která uvádí, že pokud Σ_k je známé, pak $\widehat{\boldsymbol{\beta}}_G$ je alespoň tak eficientní odhad jako $\widehat{\boldsymbol{\beta}}_O$. Tyto dva odhady jsou shodně eficientní, jestliže matice Σ_k je diagonální, nebo prvních $(L - k)$ škod ve vývojovém roce k jsou vzájemně uměrné přes jednotlivé trojúhelníky, dle věty 1 v [8].

Nyní uvažujme, že matice Σ_k není známá, a proto musíme pracovat s jejím odhadem $\widehat{\Sigma}_k$. Jak již bylo zmíněno v minulé kapitole, tento fakt značně zvýší variabilitu při použití metody FGLS. I přesto, že vztah mezi odhadem $\widehat{\boldsymbol{\beta}}_O$ a $\widehat{\boldsymbol{\beta}}_{FG}$ není přesně daný, lze na základě asymptotických výsledků obecně tvrdit, že metoda FGLS dosáhne více eficientního odhadu pouze tehdy, máme-li dostatek pozorování a jsou-li korelace mezi trojúhelníky nezanedbatelné (alespoň 0.3). V opačném případě může být naopak odhad $\widehat{\boldsymbol{\beta}}_O$ přesnější, protože nevyžaduje výpočet odhadu matice Σ_k . Doplnující informace ve formě odhadu matice Σ_k , může mít tedy za následek naopak zhoršení výsledného odhadu z hlediska eficeince.

4. Aplikace na reálných datech

V poslední části této práce se věnujeme aplikaci modelu GMCL na reálných datech. V souvislosti s tím dále uvažujeme různé variace modelu GMCL, které využíváme pro porovnání kvality odhadů a pro lepší představu o tom, jak GMCL model funguje. Kromě samotného GMCL modelu ukážeme i odlišnosti mezi GMCL modelem a klasickým chain ladder modelem na zvolených datech. Jednotlivé výpočty jsou zpracovány v softwaru R a jsou přiloženy k práci.

4.1 Vstupní data

Za vstupní data jsme si zvolili tři různé kumulativní vývojové trojúhelníky. První dva trojúhelníky popisují škody vzniklé v pojištění soukromých vozidel. Poslední trojúhelník pak popisuje škody vzniklé v pojištění firemních automobilů. Zároveň první a poslední trojúhelník popisuje uhrazené škody a druhý trojúhelník udává kumulativní nastalé škody. Pro ilustraci uvádíme v následující tabulce vývojové trojúhelníky, které dále považujeme za naše vstupní data.

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	
3	114932	227704	298120	345542	367760	377999	383611	385224		
4	114452	227761	301072	340669	359979	369248	373325			
5	115597	243611	315215	354490	372376	382738				
6	127760	259416	326975	365780	386725					
7	135616	262294	327086	367357						
8	127177	244249	317972							
9	128631	246803								
10	126288									

Tabulka 4.1: Vývojový trojúhelník kumulativních uhrazených škod v soukromém pojištění vozidel, viz [8]

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	325423	336426	346061	347726	350995	353598	354797	355025	354986	355363
2	323627	339267	344507	349295	351038	351583	352050	352231	352193	
3	358410	386330	385684	384699	387678	387954	388540	389436		
4	405319	396641	391833	384819	380914	380163	379706			
5	434065	429311	422181	409322	394154	392802				
6	417178	422307	413486	406711	406503					
7	398929	398787	398020	400540						
8	378754	361097	369328							
9	351081	335507								
10	329236									

Tabulka 4.2: Vývojový trojúhelník kumulativních nastalých škod v soukromém pojištění vozidel, viz [8]

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	19827	44449	61205	77398	88079	95695	99853	104789	105427	106690
2	22331	48480	68789	92356	104958	112399	115638	117415	118571	
3	22533	44484	65691	88435	102044	112672	115973	118359		
4	23128	51328	81542	98063	113149	121515	124347			
5	25053	57220	84607	104936	117663	126180				
6	30136	64767	92288	108835	121326					
7	34764	69125	91354	111987						
8	31803	63471	92439							
9	40559	77667								
10	46285									

Tabulka 4.3: Vývojový trojúhelník kumulativních uhrazených škod ve firemním pojištění vozidel, viz [8]

4.2 Aplikace modelů

Nyní, když již máme vstupní data, můžeme přikročit k samotnému výpočtu a následné interpretaci výsledků. Postup výpočtu popisujeme v příloze obsahující implementace jednotlivých modelů, komentáře k použitým funkcím a také stručně okomentované výsledky, kterým se podrobněji věnujeme zde.

Nejprve uvedeme jednotlivé modely, které využíváme pro odhady výše budoucích škod. Prvně demonstrováným modelem je klasický chain ladder model (značíme SCL neboli standard chain ladder). Tento model zpracováváme zejména proto, abychom v následující podkapitole ukázali, v čem spočívá výhoda GMCL modelu a jak se liší použití modelu zanedbávající korelace mezi trojúhelníky oproti modelům, které je naopak uvažují. Zbylé tři uvažované modely, kterým se dále budeme věnovat, jsou různé variace GMCL modelu.

Na úvod samotných výpočtů s použitím GMCL modelu je vhodné poznamenat, že poslední vývojová období se díky malému počtu dat těžko odhadují

za pomoci metody odhadu, která uvažuje kovariance chybových členů Σ_k . V posledních obdobích, kde se již realizuje minimum napozorovaných dat, jsou vývojové faktory blízké 1 a zároveň jsou odhadnuté kovariance malé. V důsledku toho pak software vrací nepřesné nebo žádné výsledky. Ve snaze předejít tomuto negativnímu jevu využíváme v pokročilejších vývojových letech (dle volby 3-4 poslední roky) výhradně metodu odhadu OLS, která tyto kovariance zanedbává. Díky tomu se nemusíme obávat nepřesných výsledků z důvodu nedostatku dat. V praxi je tedy vhodné rozdělit vývojová období na dvě části a nejprve odhadnout první část metodou FGLS nebo IFGLS a poté odhadnout zbyvající část pomocí metody OLS.

Nejjednodušším GMCL modelem uvedeným v této práci, je *multivariate chain ladder* (MCL) model, který je specifickým případem GMCL modelu. Tento model budeme dále v textu označovat jako zjednodušený nebo jednoduchý model GMCL (využíváme zkratky MCL+int). Jedná se o přímé zobecnění klasického chain ladder modelu pro více vývojových trojúhelníků. O zjednodušený GMCL model jde pouze z toho důvodu, že matice vývojových faktorů je pouze diagonální. Postup při výpočtu je stejný jako při užití GMCL modelu, jen se od začátku omezíme pouze na diagonální matici vývojových faktorů. Navíc oproti běžnému MCL modelu ale budeme uvažovat intercepty v prvních pěti letech. Díky tomu budeme schopni dosáhnout větší přesnosti výsledků než v případě MCL modelu bez interceptů, což bude později v práci detailněji rozebráno.

Nakonec budeme uvažovat dva GMCL modely předpokládající existenci matice vývojových faktorů libovolné struktury (nejde již pouze o diagonální matici jako v předchozím případě). Rozdíl v těchto dvou uvažovaných GMCL modelech je ten, že jeden z nich zahrnuje intercepty (dále GMCL+int) v prvních pěti letech a druhý nikoli (značíme GMCL). V navazující části *Porovnání modelů* ukážeme, jak se výsledky získané aplikací těchto dvou modelů odlišují a jak důležitá je volba interceptů při odhadech výše budoucích škod. Pro ilustraci uvádíme odhadnuté vývojové trojúhelníky kumulativních škod v pojištění vozidel pomocí modelu GMCL s intercepty v tabulkách 4.4, 4.5 a 4.6.

Jednotlivé modely jsou implementovány za využití softwaru R, jak již bylo zmíněno na začátku této kapitoly. Převážná většina využívaných příkazů pochází z knihovny *ChainLadder* (odkaz na dokumentaci lze najít v samotné příloze obsahující výpočty). Stěžejními funkcemi této knihovny jsou funkce *MultiChainLadder* a *MultiChainLadder2*, které umožňují výpočty pomocí různých chain ladder modelů (včetně GMCL modelu). Obě tyto funkce mají několik parametrů. Námi nejvyužívanější jsou parametry označující typ modelu a druh metody odhadu. Za metodu odhadu volíme OLS, jestliže nechceme uvažovat korelace mezi vývojovými trojúhelníky, nebo naopak SUR, pokud chceme do výpočtu zahrnout i odhadnuté korelace mezi vývojovými trojúhelníky. Dále můžeme volbou parametru „int“ určovat, v jakých vývojových letech jsou v modelu uvažovány intercepty. Za využití těchto základních parametrů ve funkcích *MultiChainLadder* a *MultiChainLadder2* obdržíme odhady vývojových trojúhelníku pro jednotlivé uvažované modely a z nich následně můžeme vypočítat technické rezervy, které by byla pojišťovna na základě těchto modelů povinna udržovat. Odhadnuté vývo-

ové trojúhelníky pro model GMCL uvažující intercepty lze nalést v následujících tabulkách.

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	350752
3	114932	227704	298120	345542	367760	377999	383611	385224	386801	387054
4	114452	227761	301072	340669	359979	369248	373325	375696	377234	377481
5	115597	243611	315215	354490	372376	382738	387384	389844	391441	391696
6	127760	259416	326975	365780	386725	395317	399175	401711	403355	403619
7	135616	262294	327086	367357	390742	398131	400709	403254	404905	405170
8	127177	244249	317972	356405	383785	390889	393057	395553	397173	397433
9	128631	246803	308037	342039	373505	379998	380464	382881	384449	384700
10	126288	241995	297080	325543	363061	368447	366553	368881	370391	370633

Tabulka 4.4: Odhadnutý vývojový trojúhelník kumulativních uhrazených škod v soukromém pojištění vozidel pomocí GMCL modelu s intercepty

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	325423	336426	346061	347726	350995	353598	354797	355025	354986	355363
2	323627	339267	344507	349295	351038	351583	352050	352231	352193	352567
3	358410	386330	385684	384699	387678	387954	388540	389436	389394	389807
4	405319	396641	391833	384819	380914	380163	379706	380158	380117	380521
5	434065	429311	422181	409322	394154	392802	392637	393105	393062	393479
6	417178	422307	413486	406711	406503	404406	404024	404506	404462	404891
7	398929	398787	398020	400540	408778	405858	405007	405490	405446	405876
8	378754	361097	369328	379891	400817	398059	397103	397576	397533	397955
9	351081	335507	346978	363869	390118	386017	384130	384588	384546	384955
10	329236	295335	314087	341499	378574	372690	369595	370036	369995	370388

Tabulka 4.5: Odhadnutý vývojový trojúhelník kumulativních již nastalých škod v soukromém pojištění vozidel pomocí GMCL modelu s intercepty

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	19827	44449	61205	77398	88079	95695	99853	104789	105427	106690
2	22331	48480	68789	92356	104958	112399	115638	117415	118571	119991
3	22533	44484	65691	88435	102044	112672	115973	118359	119315	120744
4	23128	51328	81542	98063	113149	121515	124347	127760	128792	130335
5	25053	57220	84607	104936	117663	126180	129049	132592	133662	135264
6	30136	64767	92288	108835	121326	131358	133942	137618	138729	140391
7	34764	69125	91354	111987	126663	137655	139692	143527	144686	146419
8	31803	63471	92439	109953	124926	136429	138274	142070	143217	144933
9	40559	77667	104796	121717	137283	147795	148807	152892	154127	155973
10	46285	90652	120535	136387	152904	162650	162535	166996	168345	170362

Tabulka 4.6: Odhadnutý vývojový trojúhelník kumulativních uhrazených škod ve firemním pojištění vozidel pomocí GMCL modelu s intercepty

Výsledky získané aplikací jednotlivých modelů na vstupní data lze dále shrnout do tabulky 4.7, která obsahuje odhadnuté celkové úhrny škod pro jednotlivé roky vzniku ve všech třech vývojových trojúhelnících dohromady. V návaznosti na tyto výsledky můžeme odečtením posledních pozorovaných hodnot v jednotlivých letech vzniku všech vývojových trojúhelníků od příslušných celkových úhrnů škod získat hodnoty IBNR rezerv. Shrnutí těchto rezerv pro jednotlivé aplikované modely nalezneme v tabulce 4.8.

	SCL	MCL+int	GMCL	GMC+int
1	815637.00	815637.00	815637.00	815637.00
2	823310.65	823310.65	823310.65	823310.65
3	897605.12	897605.12	897605.12	897605.12
4	888336.57	888336.57	888336.57	888336.57
5	923767.95	920117.71	920439.34	920439.34
6	959659.06	948241.16	949043.50	948901.49
7	962732.46	942177.07	940601.70	957464.99
8	927081.25	911468.56	886783.91	940320.23
9	925217.48	915577.49	815118.69	925627.55
10	971320.73	937361.48	713035.20	911383.05
Total	9094668.26	8999832.80	8649911.69	9029025.99

Tabulka 4.7: Celkové úhrny škod v jednotlivých letech pro všechny vývojové trojúhelníky

	SCL	MCL+int	GMCL	GMC+int
1	0	0	0	0
2	2023.65	2023.65	2023.65	2023.65
3	4586.12	4586.12	4586.12	4586.12
4	10958.57	10958.57	10958.57	10958.57
5	22047.95	18397.71	18719.34	18719.34
6	45105.06	33687.16	34489.50	34347.49
7	82848.46	62293.07	60717.70	77580.99
8	147342.25	131729.56	107044.91	160581.23
9	265240.48	255600.49	155141.69	265650.55
10	469511.73	435552.48	211226.20	409574.05
Total	1049664.26	954828.8	604907.69	984021.99

Tabulka 4.8: Odhady IBNR rezerv pro všechny vývojové trojúhelníky

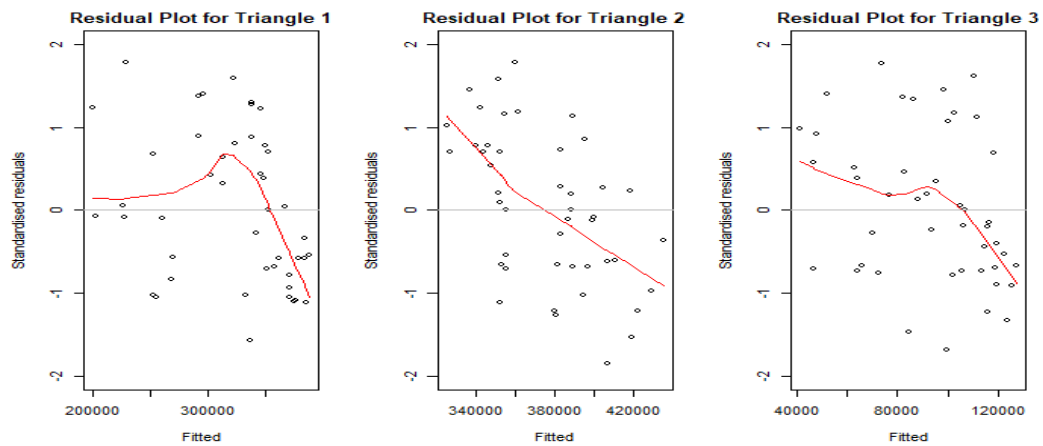
Obdrželi jsme tedy celkové hodnoty IBNR rezerv pro všechny tři vývojové trojúhelníky na základě čtyř popsaných modelů. V následující kapitole se budeme věnovat tomu, které odhady jsou pro námi zvolená data nejpřesnější.

4.3 Porovnání modelů

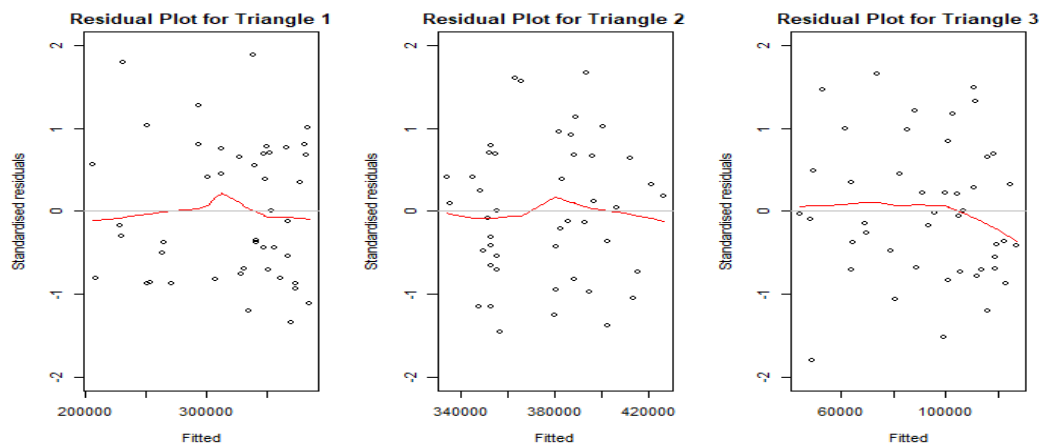
V této části využijeme výsledky modelů z předchozí podkapitoly a pokusíme se rozebrat některé jejich odlišnosti a porovnat kvalitu jednotlivých odhadů.

Nejprve budeme analyzovat standardizovaná rezidua, tedy rezidua vydělená jejich směrodatnou odchylkou. Pomocí grafů zobrazující standardizovaná rezidua

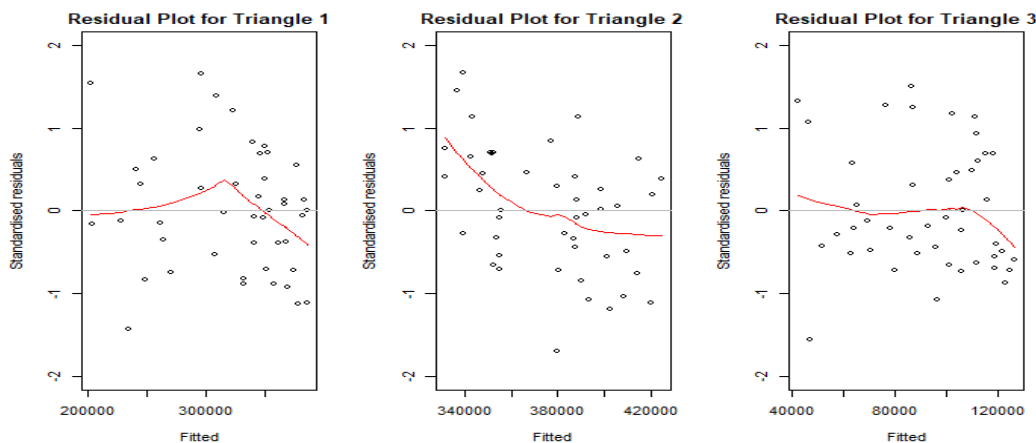
pozorujeme, jak přibližně vypadá jejich rozdělení. Čím více symetrické a blízké nule hodnoty standardizovaných reziduí jsou, tím kvalitnější odhad získáváme. Pro tyto účely vykreslujeme grafy zobrazující standardizovaná rezidua jednotlivých trojúhelníků pro dané modely vyhlazené křivkou, abychom získali lepší představu o rozděleních, kterými se tato rezidua přibližně řídí. Takto popsané grafy se nejjednodušším způsobem dají získat přímo z objektu, který vznikne při odhadování jednotlivých modelů v softwaru R. Není tedy nutné vše samostatně vypočítávat z výsledných odhadů, jelikož i námi využívané grafy jsou součástí funkcí *MultiChainLadder* a *MultiChainLadder2*. Zmiňované grafy můžeme pozorovat na obrázcích 4.1, 4.2, 4.3 a 4.4. Každá sekce grafů přísluší jednomu z modelů a jednotlivé grafy odpovídají daným vývojovým trojúhelníkům.



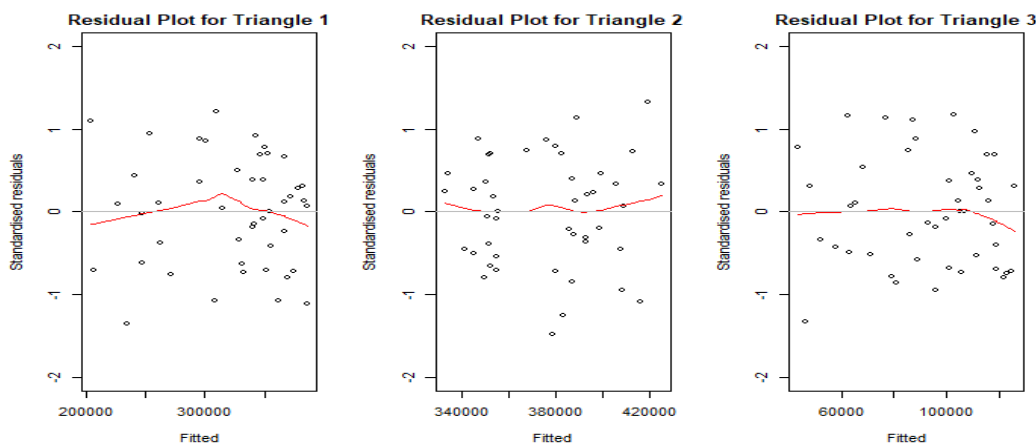
Obrázek 4.1: SCL - standardizovaná rezidua



Obrázek 4.2: MCL s intercepty - standardizovaná rezidua



Obrázek 4.3: GMCL bez interceptů - standardizovaná rezidua



Obrázek 4.4: GMCL s intercety - standardizovaná rezidua

Z grafů je možné usoudit, že nejvíce nesouměrné a obecně nerovnoměrně rozptýlené hodnoty se objevují při použití klasického chain ladder modelu, což se dalo očekávat vzhledem k tomu, že žádným způsobem nezohledňuje korelace mezi jednotlivými trojúhelníky. Dále se také ukazuje důležitost použití interceptů. I přesto, že model GMCL bez interceptů uvažuje složitější strukturu matice vývojových faktorů než MCL model, nevykazuje stejně malé hodnoty standardizovaných reziduí jako právě zmíněný zjednodušený GMCL model (MCL+int). Podobně i GMCL model zahrnující intercepty dopadá z tohoto hlediska mnohem lépe než GMCL model neobsahující intercepty. Právě posledně zmiňovaný GMCL model s intercepty má hodnoty reziduí obecně nejvyrovnanější, a proto se alespoň z tohoto hlediska jedná o nejlepší odhad v porovnání s ostatními modely.

Dále se věnujeme porovnání modelů na základě P/I poměru celkového úhrnu škod za jednotlivé roky vzniku. P/I poměr zde značí podíl celkových úhrnů škod za jednotlivé roky vzniku uhrazených (P) a vzniklých (I) škod v jednom pojistném odvětví. U tohoto kritéria vyžadujeme, aby P/I poměr byl v čase co nejvíce stabilní a zároveň, aby hodnoty tohoto podílu byly blízké hodnotě sta procent. Vycházíme totiž z toho, že odhady trojúhelníků P a I by měly být podobné vzhledem k tomu, že odhadujeme data pro stejná pojistná odvětví. P/I poměry pro

jednotlivé modely jsou rozpracovány v následující tabulce 4.9.

	SCL	MCL+int	GMCL	GMCL+int
1	99.50	99.50	99.50	99.50
2	99.49	99.49	99.49	99.49
3	99.29	99.29	99.29	99.29
4	99.20	99.20	99.20	99.20
5	99.82	99.46	99.55	99.55
6	100.43	99.64	99.69	99.69
7	103.53	102.08	99.96	99.83
8	111.23	106.73	100.29	99.87
9	122.15	111.17	100.87	99.93
10	126.15	111.37	101.88	100.07
Total	105.56	102.67	99.91	99.64

Tabulka 4.9: P/I poměry pro jednotlivé modely

Z tabulky 4.9 je zřejmé, že zatímco u klasického chain ladder modelu dochází k poměrně zásadnímu rozdílu v hodnotách odhadů P a I trojúhelníků, zbylé modely lépe reflektují určitý vztah mezi trojúhelníky. Můžeme si všimnout, že jak GMCL model zahrnující intercepty, tak GMCL model bez interceptů je schopný odhadovat hodnoty P/I poměrů velmi blízké jedné. Lze tedy konstatovat, že z hlediska P/I poměru dopadly modely GMCL výrazně lépe, než zjednodušený GMCL (MCL+int) a klasický chain ladder (SCL) model.

Na základě předcházející analýzy reziduí a P/I poměru můžeme tvrdit, že model, který dle výše popsaných kritérií nejlépe vystihuje budoucí vývoj dat a tím pádem dává nejspolehlivější odhady pro námi zvolená vstupní data, je právě GMCL model, který kromě korelací jednotlivých trojúhelníků zahrnuje i intercepty. Právě variabilita GMCL modelu se ukazuje být zásadní v samotných výpočtech, kdy je možné na základě vstupních dat rozhodnout jakou metodu a nástroje při výpočtu optimálně využít.

Závěr

V práci jsme odvodili GMCL model, který slouží pro odhady rezerv v neživotním pojištění ve více pojistných odvětví najednou. To znamená, že můžeme uvažovat několik vývojových trojúhelníků a kromě samostatných škod v jednotlivých trojúhelnících můžeme do výpočtu zahrnout i závislosti mezi těmito trojúhelníky prostřednictvím jejich korelací. Pro tyto účely nejprve uvádíme klasický chain ladder model, na jehož základu následně stojí model GMCL využívající metodiky SUR, která umožňuje vytvářet odhady zahrnující právě zmíněné korelace mezi vývojovými trojúhelníky. Samotné metodice SUR je také věnována podstatná část práce, jelikož tvoří jádro modelu GMCL.

Dále se věnujeme zejména odvození praktického zápisu obecného GMCL modelu, na který již můžeme přímo aplikovat metodiku SUR. Pomocí takto definovaného modelu jsme schopni obdržet několik různých druhů odhadů v závislosti na způsobu jejich výpočtu. Tomuto tématu byla věnována převážná část kapitoly zabývající se modelem GMCL a odhady získaných pomocí tohoto modelu. Na tato pozorování navazuje kapitola zabývající se praktickým použitím GMCL modelu.

Nakonec věnujeme významnou část práce právě aplikaci GMCL modelu na reálná data, kde provádíme výpočty na třech různých vývojových trojúhelnících a analyzujeme výsledné celkové úhrny škod a z nich plynoucí rezervy IBNR. K těmto výpočtům využíváme různé variace modelu GMCL, které následně porovnáváme mezi sebou a zároveň s výsledky obdrženými prostřednictvím klasického chain ladder modelu popsáném v první kapitole. Pomocí několika kritérií jsme zkoumali jednotlivé modely a z výsledků této analýzy jsme zjistili, že nej přesnější odhady pro námi zvolená data obdržíme při použití GMCL modelu, který zahrnuje kromě korelací mezi trojúhelníky také intercepty v prvních pěti vývojových obdobích.

Seznam použité literatury

[1] BADOUNAS, Ioannis a Georgios PITSELIS. Loss Reserving Estimation With Correlated Run-Off Triangles in a Quantile Longitudinal Model. *Risks*, 2020, 8(1), 588-599. ISSN 2227-9091. Dostupné z: doi:10.3390/risks8010014.

[2] BALAKRISHNAN, N., Theodore COLTON, Brian EVERITT, Walter PIEGORSCH, Fabrizio RUGGERI a Jozef L. TEUGELS, eds. *Wiley StatsRef: Statistics Reference Online* [online]. Chichester, UK: John Wiley & Sons, 2014 [cit. 2021-03-05]. ISBN 9781118445112.

[3] BALTAGI, Badi H. Seemingly Unrelated Regressions. BALTAGI, Badi H. *Econometrics* [online]. Berlin, Heidelberg: Springer, 2011, 2011-4-9, s. 241-256 [cit. 2020-12-16]. ISBN 978-3-642-20058-8. Dostupné z: doi: 10.1007/978-3-642-20059-5_10.

[4] BESTOVÁ, Jana. Run-off analýzy v neživotním pojištění [online]. Praha, 2007 [cit. 2020]. Vedoucí práce Mgr. Marcela Vítková. Dostupné z: <https://docplayer.cz/36759689-Run-off-analyzy-v-nezivotnim-pojisteni.html>. Diplomová práce. Univerzita Karlova.

[5] MACK, Thomas. A Simple Parametric Model For Rating Automobile Insurance or Estimating IBNR Claims Reseres. *Astin Bulletin*, 1991, 21(1), 93-111.

[6] SRIVASTAVA, Virendera K. a David E.A. GILES. Seemingly Unrelated Regression Equations Models [online]. CRC Press, 2020 [cit. 2020-12-20]. ISBN 9781003065654. Dostupné z: doi:10.1201/9781003065654

[7] ZELLNER, Arnold. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 1962, 57(298), 348- 368.

[8] ZHANG, Yanwei. A general multivariate chain ladder model. *Insurance: Mathematics and Economics*, 2010, 46(3), 588-599.

Seznam tabulek

1.1	Vývojový trojúhelník	3
1.2	Vývojový trojúhelník doplněný na obdélník	5
4.1	Vývojový trojúhelník kumulativních uhrazených škod v soukromém pojištění vozidel, viz [8]	17
4.2	Vývojový trojúhelník kumulativních nastalých škod v soukromém pojištění vozidel, viz [8]	18
4.3	Vývojový trojúhelník kumulativních uhrazených škod ve firemním pojištění vozidel, viz [8]	18
4.4	Odhadnutý vývojový trojúhelník kumulativních uhrazených škod v soukromém pojištění vozidel pomocí GMCL modelu s intercepty	20
4.5	Odhadnutý vývojový trojúhelník kumulativních již nastalých škod v soukromém pojištění vozidel pomocí GMCL modelu s intercepty	20
4.6	Odhadnutý vývojový trojúhelník kumulativních uhrazených škod ve firemním pojištění vozidel pomocí GMCL modelu s intercepty .	20
4.7	Celkové úhrny škod v jednotlivých letech pro všechny vývojové trojúhelníky	21
4.8	Odhady IBNR rezerv pro všechny vývojové trojúhelníky	21
4.9	P/I poměry pro jednotlivé modely	24

A. Příloha

A.1 Zdrojový kód spolu s výstupy

Tato příloha je tvořena skriptem, ve kterém ilustrujeme použití GMCL modelu na reálných datech. Pro výpočty založené na GMCL modelu využíváme softwaru R, konkrétněji knihovnu *ChainLadder*. Základní dokumentace k této knihovně je dostupná z <https://www.rdocumentation.org/packages/ChainLadder/versions/0.1.5-1/topics/MultiChainLadder> a rozsáhlejší dokumentaci je možné nalézt na stránce <https://CRAN.R-project.org/package=ChainLadder>.

```
library(ChainLadder)
```

Welcome to ChainLadder version 0.2.12

To cite package 'ChainLadder' in publications use:

```
Markus Gesmann, Daniel  
Murphy, Yanwei (Wayne)  
Zhang, Alessandro Carrato,  
Mario Wuthrich, Fabio  
Concina and Eric Dal Moro  
(2021). ChainLadder:  
Statistical Methods and  
Models for Claims Reserving  
in General Insurance. R  
package version 0.2.12.  
https://CRAN.R-project.org/package=ChainLadder
```

To suppress this message use:

```
suppressPackageStartupMessages(library(ChainLadder))
```

Pro exportování matic a tabulek dále využíváme knihovnu *xtable*, pomocí které převádíme data ze softwaru R do podoby kódu v LaTexu, což umožňuje jednoduché přenesení výsledků do jednodušeji čitelné podoby.

Data a použité funkce

Nejprve načteme vstupní data, která budeme dále v práci využívat pro veškeré výpočty. Tato námi zvolená data jsou součástí již zmínované knihovny *ChainLadder* a jsou totožná s těmi, jež byly použity v práci [8]. Samotná data nabývají tvar tří rozdílných vývojových trojúhelníků. První dva trojúhelníky se týkají pojištění soukromých vozidel a poslední trojúhelník zaznamenává škody realizované v pojištění firemních automobilů. První a poslední trojúhelník udává počet zaplacených ("paid", P) škod a trojúhelník druhý škod nastalých ("incurred", I).


```

## Načítaná data "auto" z ChainLadder knihovny

# Kumulativní vývojový trojúhelník uhrazených škod
# z pojištění soukromých vozidel
auto[[1]]
# Kumulativní vývojový trojúhelník nastalých škod
# z pojištění soukromých vozidel
auto[[2]]
# Kumulativní vývojový trojúhelník uhrazených škod
# z pojištění firemních vozidel
auto[[3]]

## Export tabulek (matic) do Latexu
print(xtable(round(cbind(auto[[1]]),0)), file = "tbl1.tex")
print(xtable(round(cbind(auto[[2]]),0)), file = "tbl2.tex")
print(xtable(round(cbind(auto[[3]]),0)), file = "tbl3.tex")

```

	Vývojové roky									
Roky vzniku	1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	
3	114932	227704	298120	345542	367760	377999	383611	385224		
4	114452	227761	301072	340669	359979	369248	373325			
5	115597	243611	315215	354490	372376	382738				
6	127760	259416	326975	365780	386725					
7	135616	262294	327086	367357						
8	127177	244249	317972							
9	128631	246803								
10	126288									

1. vývojový trojúhelník

	Vývojové roky									
Roky vzniku	1	2	3	4	5	6	7	8	9	10
1	325423	336426	346061	347726	350995	353598	354797	355025	354986	355363
2	323627	339267	344507	349295	351038	351583	352050	352231	352193	
3	358410	386330	385684	384699	387678	387954	388540	389436		
4	405319	396641	391833	384819	380914	380163	379706			
5	434065	429311	422181	409322	394154	392802				
6	417178	422307	413486	406711	406503					
7	398929	398787	398020	400540						
8	378754	361097	369328							
9	351081	335507								
10	329236									

2. vývojový trojúhelník

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	19827	44449	61205	77398	88079	95695	99853	104789	105427	106690
2	22331	48480	68789	92356	104958	112399	115638	117415	118571	
3	22533	44484	65691	88435	102044	112672	115973	118359		
4	23128	51328	81542	98063	113149	121515	124347			
5	25053	57220	84607	104936	117663	126180				
6	30136	64767	92288	108835	121326					
7	34764	69125	91354	111987						
8	31803	63471	92439							
9	40559	77667								
10	46285									

3. vývojové trojúhelník

Stěžejními funkcemi dané knihovny, které budeme dále využívat pro naše výpočty, jsou *MultiChainLadder()* a *MultiChainLadder2()*. Tyto dvě funkce mají několik vstupních parametrů. Prvním z nich je list vývojových trojúhelníků, pro které využíváme námi zvolený model. Model vybereme nastavením parametru *model*. Primárně využíváme model *GMCL*, ale pro srovnání také uvádíme modely *MCL* (pouze diagonální matice vývojových faktorů), jakožto speciální případ *GMCL* modelu, nebo při vynechání parametru *model* metodu `<i>SCL</i>` (klasický chain ladder model), který zanedbává závislosti mezi jednotlivými trojúhelníky. Dále je možné upravit metodu odhadování pomocí parametru *fit.method*. Tento parametr je však automaticky již vhodně nastaven pro model *GMCL*, tj. na metodu *SUR*. Z toho důvodu není nezbytně nutné ho předefinovávat pro použití `<i>GMCL</i>` modelu. Při aplikaci modelů jiných je tento parametr velmi důležitý, např. z důvodu záměrného zanedbávání korelací. Zbylými parametry je možné např. nastavit různá omezení parametrů nebo zadat konkrétní vývojové roky, ve kterých se vyskytují intercepty. Některé z těchto parametrů využijeme i v této práci, ale samotný komentář k nim doplníme až při použití ve výpočtech.

Model Chain ladder

Nejprve odhadneme dané vývojové trojúhelníky pomocí klasického chain ladder modelu (značíme *SCL*, standard chain ladder). Využijeme k tomu funkci *MultiChainLadder2* s argumentem *OLS*, který zajišťuje, že odhadujeme pomocí metody *OLS* a tedy budeme predikovat hodnoty škod pro jednotlivé vývojové trojúhelníky odděleně bez uvažování jakýchkoliv závislostí.

```
## Implementace SCL modelu
scl <- MultiChainLadder2(auto, "OLS")
scl.trg <- scl$FullTriangles

## Shrnutí
sumup.scl <- summary(scl) # Obecné shrnutí
x <- sumup.scl$Ultimate # Celkové škody pro jednotlivé
# roky a trojúhelníky

y <- sumup.scl$IBNR # IBNR
```

```

## Export tabulek (matic) do Latexu
# Doplněné trojúhelníky
print(xtable(round(cbind(scl.trg[[1]]),0)), file = "tblscl1.tex")
print(xtable(round(cbind(scl.trg[[2]]),0)), file = "tblscl2.tex")
print(xtable(round(cbind(scl.trg[[3]]),0)), file = "tblscl3.tex")
print(xtable(cbind(x)), file = "ultSCL.tex")
print(xtable(cbind(y)), file = "ibnrSCL.tex")

```

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	350752
3	114932	227704	298120	345542	367760	377999	383611	385224	386801	387054
4	114452	227761	301072	340669	359979	369248	373325	375696	377234	377481
5	115597	243611	315215	354490	372376	382738	389054	391525	393128	393385
6	127760	259416	326975	365780	386725	398772	405352	407927	409597	409865
7	135616	262294	327086	367357	390804	402978	409628	412229	413917	414188
8	127177	244249	317972	361356	384420	396395	402937	405496	407156	407422
9	128631	246803	317291	360583	383598	395547	402074	404628	406285	406550
10	126288	250861	322508	366512	389905	402050	408685	411281	412965	413235

Odhadnutý 1. trojúhelník - SCL model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	325423	336426	346061	347726	350995	353598	354797	355025	354986	355363
2	323627	339267	344507	349295	351038	351583	352050	352231	352193	352567
3	358410	386330	385684	384699	387678	387954	388540	389436	389394	389807
4	405319	396641	391833	384819	380914	380163	379706	380158	380117	380521
5	434065	429311	422181	409322	394154	392802	393253	393721	393678	394096
6	417178	422307	413486	406711	406503	406781	407248	407733	407688	408121
7	398929	398787	398020	400540	398480	398752	399210	399685	399642	400066
8	378754	361097	369328	366718	364832	365081	365500	365936	365896	366285
9	351081	335507	335591	333219	331505	331732	332113	332508	332472	332825
10	329236	330202	330284	327950	326263	326486	326861	327250	327215	327562

Odhadnutý 2. trojúhelník - SCL model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	19827	44449	61205	77398	88079	95695	99853	104789	105427	106690
2	22331	48480	68789	92356	104958	112399	115638	117415	118571	119991
3	22533	44484	65691	88435	102044	112672	115973	118359	119315	120744
4	23128	51328	81542	98063	113149	121515	124347	127760	128792	130335
5	25053	57220	84607	104936	117663	126180	130025	133595	134673	136287
6	30136	64767	92288	108835	121326	131167	135164	138875	139996	141673
7	34764	69125	91354	111987	127154	137468	141657	145546	146721	148478
8	31803	63471	92439	115680	131347	142001	146328	150345	151559	153374
9	40559	77667	112007	140168	159152	172061	177304	182171	183642	185842
10	46285	96340	138937	173868	197416	213429	219933	225971	227795	230524

Odhadnutý 3. trojúhelník - SCL model

Rok vzniku	Vývojový trojúhelník			
	1	2	3	1+2+3
1	353584.00	355363.00	106690.00	815637.00
2	350752.15	352567.03	119991.46	823310.65
3	387054.02	389807.14	120743.96	897605.12
4	377481.05	380520.67	130334.85	888336.57
5	393384.94	394096.37	136286.64	923767.95
6	409864.70	408121.44	141672.92	959659.06
7	414187.75	400066.25	148478.46	962732.46
8	407422.23	366284.53	153374.49	927081.25
9	406550.25	332825.15	185842.09	925217.48
10	413234.51	327562.13	230524.09	971320.73
Total	3913515.59	3707213.70	1473938.97	9094668.26

Celkové odhadnuté úhrny škod - SCL model

Rok vzniku	Vývojový trojúhelník			
	1	2	3	1+2+3
1	0.00	0.00	0.00	0.00
2	229.15	374.03	1420.46	2023.65
3	1830.02	371.14	2384.96	4586.12
4	4156.05	814.67	5987.85	10958.57
5	10646.94	1294.37	10106.64	22047.95
6	23139.70	1618.44	20346.92	45105.06
7	46830.75	-473.75	36491.46	82848.46
8	89450.23	-3043.47	60935.49	147342.25
9	159747.25	-2681.85	108175.09	265240.48
10	286946.51	-1673.87	184239.09	469511.73
Total	622976.59	-3400.30	430087.97	1049664.26

Odhadnuté IBNR rezervy - SCL model

Model GMCL

Tato podkapitola praktické části práce se soustřeďuje právě na výpočet rezerv za pomoci GMCL modelu. Ilustrujeme tři různé varianty. První nejjednodušší

model uvažuje pouze diagonální matici vývojových faktorů, zbylé dva uvažují libovolnou čtvercovou matici vývojových faktorů.

Na základě předcházející analýzy reziduí a P/I poměru můžeme tvrdit, že model, který dle výše popsaných kritérií nejlépe vystihuje budoucí vývoj dat a tím pádem dává nejspolehlivější odhady pro námi zvolená vstupní data, je právě GMCL model, který kromě korelací jednotlivých trojúhelníku zahrnuje i intercepty. Právě variabilita GMCL modelu se ukazuje být zásadní v samotných výpočtech, kdy je možné na základě vstupních dat rozhodnout jakou metodu a nástroje při výpočtu optimálně využít.

Poznámka:

Dle dokumentace dochází v některých případech k selhání funkce *MultiChainLadder*. Důvodem většinou bývá nedostatek dat v posledních vývojových obdobích. Za takového předpokladu vrátí funkce s největší pravděpodobností chybu, jelikož kovarianční matice Σ_k je singulární. Abychom se takovému scénáři vyhnuli, rozdělíme si vývojové trojúhelníky na dvě oddělené matice. Na první část aplikujeme model GMCL a na zbylá data poté klasickou metodu chain ladder, která neuvažuje korelace mezi chybovými členy a tudíž tomuto problému předchází. To lze v praxi udělat manuálně nebo s využitím funkce *MultiChainLadder2*, která obsahuje navíc oproti funkci *MultiChainLadder* argument *last*. Tento parametr označuje, kolik posledních let chceme odhadovat odděleně. Ve výpočtech ukážeme postup jak manuální, tak pomocí funkce *MultiChainLadder2*.

Zjednodušený GMCL model s intercepty

Prvním a nejjednodušším GMCL modelem je takový model, který uvažujeme pouze diagonální matici vývojových faktorů. Jedná se o přímé zobecnění klasického chain ladder modelu. K tomuto modelu je ale možné navíc přidat intercepty a odhadovat vývojové faktory za pomoci metodiky SUR. Využijeme těchto možností a zvolíme intercepty ve vývojových letech 1, 2, 3, 4 a 5. Tento model získáme volbou parametru *MCL+int* ve funkci *MultiChainLadder2*.

```
## Implementace MCL Modelu s intercepty (MCL+int)
mcl <- MultiChainLadder2(auto,
                        last = 3,
                        type = "MCL+int",
                        fit.method = "SUR")
mcl.trg <- mcl$FullTriangles

## Shrnutí
sumup.mcl <- summary(mcl)
x <- sumup.mcl$Ultimate      # Celkové škody

y <- sumup.mcl$IBNR         # IBNR

## Export tabulek (matic) do Latexu
```

```

print(xtable(round(cbind(mcl.trg[[1]]),0)), file = "tblmcl1.tex")
print(xtable(round(cbind(mcl.trg[[2]]),0)), file = "tblmcl2.tex")
print(xtable(round(cbind(mcl.trg[[3]]),0)), file = "tblmcl3.tex")
print(xtable(cbind(x)), file = "ultMCL.tex")
print(xtable(cbind(y)), file = "ibnrMCL.tex")

```

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	350752
3	114932	227704	298120	345542	367760	377999	383611	385224	386801	387054
4	114452	227761	301072	340669	359979	369248	373325	375696	377234	377481
5	115597	243611	315215	354490	372376	382738	387201	389660	391256	391511
6	127760	259416	326975	365780	386725	395514	399228	401763	403408	403672
7	135616	262294	327086	367357	385946	394782	398539	401070	402712	402975
8	127177	244249	317972	358763	378135	387447	391634	394121	395735	395993
9	128631	246803	316262	357065	376592	385997	390269	392748	394356	394614
10	126288	249184	319139	359923	379189	388436	392565	395059	396676	396935

Odhadnutý 1. trojúhelník - MCL+int model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	325423	336426	346061	347726	350995	353598	354797	355025	354986	355363
2	323627	339267	344507	349295	351038	351583	352050	352231	352193	352567
3	358410	386330	385684	384699	387678	387954	388540	389436	389394	389807
4	405319	396641	391833	384819	380914	380163	379706	380158	380117	380521
5	434065	429311	422181	409322	394154	392802	392805	393273	393231	393648
6	417178	422307	413486	406711	406503	404457	404249	404730	404686	405116
7	398929	398787	398020	400540	395199	393930	393913	394382	394339	394758
8	378754	361097	369328	369258	369290	369802	370224	370665	370624	371018
9	351081	335507	344288	348117	351781	353495	354214	354636	354597	354974
10	329236	338267	346554	350029	353365	354970	355662	356086	356047	356425

Odhadnutý 2. trojúhelník - MCL+int model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	19827	44449	61205	77398	88079	95695	99853	104789	105427	106690
2	22331	48480	68789	92356	104958	112399	115638	117415	118571	119991
3	22533	44484	65691	88435	102044	112672	115973	118359	119315	120744
4	23128	51328	81542	98063	113149	121515	124347	127760	128792	130335
5	25053	57220	84607	104936	117663	126180	128758	132292	133360	134958
6	30136	64767	92288	108835	121326	130702	133046	136699	137802	139453
7	34764	69125	91354	111987	126097	135722	137808	141591	142734	144444
8	31803	63471	92439	111999	126109	135735	137820	141604	142747	144457
9	40559	77667	111604	131184	146692	157396	158364	162711	164025	165990
10	46285	88894	127634	147231	163908	175513	175548	180366	181823	184001

Odhadnutý 3. trojúhelník - MCL+int model

Vývojový trojúhelník				
	1	2	3	1+2+3
Rok vzniku				
1	353584.00	355363.00	106690.00	815637.00
2	350752.15	352567.03	119991.46	823310.65
3	387054.02	389807.14	120743.96	897605.12
4	377481.05	380520.67	130334.85	888336.57
5	391511.44	393648.22	134958.04	920117.71
6	403672.01	405115.97	139453.18	948241.16
7	402975.37	394757.77	144443.94	942177.07
8	395993.45	371017.96	144457.14	911468.56
9	394613.57	354973.76	165990.16	915577.49
10	396935.30	356425.27	184000.91	937361.48
Total	3854572.37	3754196.79	1391063.64	8999832.80

Celkové odhadnuté úhrny škod - MCL+int model

Vývojový trojúhelník				
	1	2	3	1+2+3
Rok vzniku				
1	0.00	0.00	0.00	0.00
2	229.15	374.03	1420.46	2023.65
3	1830.02	371.14	2384.96	4586.12
4	4156.05	814.67	5987.85	10958.57
5	8773.44	846.22	8778.04	18397.71
6	16947.01	-1387.03	18127.18	33687.16
7	35618.37	-5782.23	32456.94	62293.07
8	78021.45	1689.96	52018.14	131729.56
9	147810.57	19466.76	88323.16	255600.49
10	270647.30	27189.27	137715.91	435552.48
Total	564033.37	43582.79	347212.64	954828.80

Odhadnuté IBNR rezervy - MCL+int model

GMCL model bez interceptů

Volba *GMCL-int* označuje model GMCL bez interceptů a parametr *last* poslední vývojové roky, které odhadujeme zvlášť metodou OLS. Následující volbou výše zmíněných argumentů dostaneme GMCL model bez interceptů.

```
## Implementace Modelu GMCL bez interceptů
full.gmcl1 <- MultiChainLadder2(auto,
                                last = 3,
                                type = "GMCL-int",
                                fit.method = "SUR")

gmcl1.trg <- full.gmcl1$FullTriangles
gmcl1.trg

## Shrnutí
sumup.gmcl1 <- summary(full.gmcl1)
x <- sumup.gmcl1$Ultimate      # Celkové škod
y <- sumup.gmcl1$IBNR        # IBNR
```

```

## Export tabulek (matic) do Latexu
print(xtable(round(cbind(gmcl1.trg[[1]]),0)), file = "tblgmcl11.tex")
print(xtable(round(cbind(gmcl1.trg[[2]]),0)), file = "tblgmcl12.tex")
print(xtable(round(cbind(gmcl1.trg[[3]]),0)), file = "tblgmcl13.tex")
print(xtable(cbind(x)), file = "ultGMCL1.tex")
print(xtable(cbind(y)), file = "ibnrGMCL1.tex")

```

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	350752
3	114932	227704	298120	345542	367760	377999	383611	385224	386801	387054
4	114452	227761	301072	340669	359979	369248	373325	375696	377234	377481
5	115597	243611	315215	354490	372376	382738	387384	389844	391441	391696
6	127760	259416	326975	365780	386725	395353	399212	401747	403392	403656
7	135616	262294	327086	367357	386190	391609	393219	395717	397337	397597
8	127177	244249	317972	353291	368618	370249	369609	371957	373480	373724
9	128631	246803	306294	331349	341651	337690	332725	334838	336209	336429
10	126288	237048	287822	300366	303940	292640	282253	284045	285208	285395

Odhadnutý 1. trojúhelník - GMCL-int model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	325423	336426	346061	347726	350995	353598	354797	355025	354986	355363
2	323627	339267	344507	349295	351038	351583	352050	352231	352193	352567
3	358410	386330	385684	384699	387678	387954	388540	389436	389394	389807
4	405319	396641	391833	384819	380914	380163	379706	380158	380117	380521
5	434065	429311	422181	409322	394154	392802	392637	393105	393062	393479
6	417178	422307	413486	406711	406503	404448	404064	404545	404501	404931
7	398929	398787	398020	400540	402208	397977	396909	397382	397338	397760
8	378754	361097	369328	373402	379491	373412	371853	372296	372255	372650
9	351081	335507	338050	340613	345831	335820	332823	333220	333184	333538
10	329236	290855	290805	293245	299055	284224	279528	279861	279830	280127

Odhadnutý 2. trojúhelník - GMCL-int model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	19827	44449	61205	77398	88079	95695	99853	104789	105427	106690
2	22331	48480	68789	92356	104958	112399	115638	117415	118571	119991
3	22533	44484	65691	88435	102044	112672	115973	118359	119315	120744
4	23128	51328	81542	98063	113149	121515	124347	127760	128792	130335
5	25053	57220	84607	104936	117663	126180	129049	132592	133662	135264
6	30136	64767	92288	108835	121326	131422	134004	137683	138794	140457
7	34764	69125	91354	111987	126166	136947	138572	142376	143525	145245
8	31803	63471	92439	108264	121847	133311	133959	137636	138747	140410
9	40559	77667	104362	115631	128651	139784	138484	142285	143434	145152
10	46285	87843	115978	121422	133532	144472	140736	144599	145767	147513

Odhadnutý 3. trojúhelník - GMCL-int model

Vývojový trojúhelník				
	1	2	3	1+2+3
Rok vzniku				
1	353584.00	355363.00	106690.00	815637.00
2	350752.15	352567.03	119991.46	823310.65
3	387054.02	389807.14	120743.96	897605.12
4	377481.05	380520.67	130334.85	888336.57
5	391696.42	393479.27	135263.65	920439.34
6	403655.89	404930.54	140457.07	949043.50
7	397596.58	397760.37	145244.75	940601.70
8	373724.00	372650.37	140409.54	886783.91
9	336428.77	333537.58	145152.34	815118.69
10	285394.71	280127.48	147513.01	713035.20
Total	3657367.59	3660743.45	1331800.64	8649911.69

Celkové odhadnuté úhrny škod - GMCL-int model

Vývojový trojúhelník				
	1	2	3	1+2+3
Rok vzniku				
1	0.00	0.00	0.00	0.00
2	229.15	374.03	1420.46	2023.65
3	1830.02	371.14	2384.96	4586.12
4	4156.05	814.67	5987.85	10958.57
5	8958.42	677.27	9083.65	18719.34
6	16930.89	-1572.46	19131.07	34489.50
7	30239.58	-2779.63	33257.75	60717.70
8	55752.00	3322.37	47970.54	107044.91
9	89625.77	-1969.42	67485.34	155141.69
10	159106.71	-49108.52	101228.01	211226.20
Total	366828.59	-49870.55	287949.64	604907.69

Odhadnuté IBNR rezervy - GMCL-int model

GMCL model s intercepty

Nakonec uvažujeme model GMCL s intercepty v prvních pěti vývojových letech. I v tomto případě využíváme pro výpočet metodiku SUR, konkrétněji metodu odhadu FGLS. Tentokrát ukážeme postup s využitím funkce *MultiChainLadder*. Nejprve odhadneme parametry pro obě skupiny a poté modely spojíme a vytvoříme společný odhad. K tomu využívámá funkcí *predict()*, *Mse()*, *JoinFitMse()*, přičemž kromě funkce *predict* jsou zbylé dvě zmíněné součástí knihovny *ChainLadder*. Níže popsany postup vychází právě z doporučení v dokumentaci k této knihovně.

```

auto <- as(auto,"triangles") # Nutné přetypovat data na třídu
                                # 'triangles'
intercepts <- 1:5                # Vybrané vývojové roky
                                # obsahující intercepty
cast1 <- auto[, 1:7]             # První část dat, vývojové roky 1 - 7
cast2 <- auto[, 7:10]           # Druhá část dat, vývojové roky 7 - 10

## Implementace modelu

```

```

gmcl21 <- MultiChainLadder(cast1,
                          "SUR",
                          int = intercepts,
                          model = "GMCL")
gmcl22 <- MultiChainLadder(cast2, "OLS")
# Na tuto část je z důvodu
# neregularity matice \Sigma_k
# nutné využít metodu OLS
fit.gmcl2 <- Join2Fits(gmcl21, gmcl22)
# Spojení dvou odhadů za různé
# časové úseky
pred.gmcl2 <- predict(fit.gmcl2)
# Lineární fitování dat (Predikce
# pomocí lineárního modelu)
mse.gmcl2 <- Mse(fit.gmcl2, pred.gmcl2)
# Funkce, která počítá střední
# čtvercovou odchylku pro doplněný
# vývojový trojúhelník a daný model
full.gmcl2 <- JoinFitMse(fit.gmcl2, mse.gmcl2)
# Konstrukce objektu,
# pro který jsou
# definované další
# operace jako například
# různá shnutí a grafy
gmcl2.trg <- full.gmcl2$FullTriangles
# Doplněné vývojové trojúhelníky
# - GMCL model s intercepty

sumup.gmcl2 <- summary(full.gmcl2) # Obecné shrnutí
x <- sumup.gmcl2$Ultimate          # Celkové škody
y <- sumup.gmcl2$IBNR             # IBNR

## Export tabulek (matic) do latexu
print(xtable(round(cbind(gmcl2.trg[[1]]),0)), file = "tblgmcl21.tex")
print(xtable(round(cbind(gmcl2.trg[[2]]),0)), file = "tblgmcl22.tex")
print(xtable(round(cbind(gmcl2.trg[[3]]),0)), file = "tblgmcl23.tex")
print(xtable(cbind(x)), file = "ultGMCL2.tex")
print(xtable(cbind(y)), file = "ibnrGMCL2.tex")

```

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	101125	209921	266618	305107	327850	340669	348430	351193	353353	353584
2	102541	203213	260677	303182	328932	340948	347333	349813	350523	350752
3	114932	227704	298120	345542	367760	377999	383611	385224	386801	387054
4	114452	227761	301072	340669	359979	369248	373325	375696	377234	377481
5	115597	243611	315215	354490	372376	382738	387384	389844	391441	391696
6	127760	259416	326975	365780	386725	395317	399175	401711	403355	403619
7	135616	262294	327086	367357	390742	398131	400709	403254	404905	405170
8	127177	244249	317972	356405	383785	390889	393057	395553	397173	397433
9	128631	246803	308037	342039	373505	379998	380464	382881	384449	384700
10	126288	241995	297080	325543	363061	368447	366553	368881	370391	370633

Odhadnutý 1. trojúhelník - GMCL model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	325423	336426	346061	347726	350995	353598	354797	355025	354986	355363
2	323627	339267	344507	349295	351038	351583	352050	352231	352193	352567
3	358410	386330	385684	384699	387678	387954	388540	389436	389394	389807
4	405319	396641	391833	384819	380914	380163	379706	380158	380117	380521
5	434065	429311	422181	409322	394154	392802	392637	393105	393062	393479
6	417178	422307	413486	406711	406503	404406	404024	404506	404462	404891
7	398929	398787	398020	400540	408778	405858	405007	405490	405446	405876
8	378754	361097	369328	379891	400817	398059	397103	397576	397533	397955
9	351081	335507	346978	363869	390118	386017	384130	384588	384546	384955
10	329236	295335	314087	341499	378574	372690	369595	370036	369995	370388

Odhadnutý 2. trojúhelník - GMCL model

Roky vzniku	Vývojové roky									
	1	2	3	4	5	6	7	8	9	10
1	19827	44449	61205	77398	88079	95695	99853	104789	105427	106690
2	22331	48480	68789	92356	104958	112399	115638	117415	118571	119991
3	22533	44484	65691	88435	102044	112672	115973	118359	119315	120744
4	23128	51328	81542	98063	113149	121515	124347	127760	128792	130335
5	25053	57220	84607	104936	117663	126180	129049	132592	133662	135264
6	30136	64767	92288	108835	121326	131358	133942	137618	138729	140391
7	34764	69125	91354	111987	126663	137655	139692	143527	144686	146419
8	31803	63471	92439	109953	124926	136429	138274	142070	143217	144933
9	40559	77667	104796	121717	137283	147795	148807	152892	154127	155973
10	46285	90652	120535	136387	152904	162650	162535	166996	168345	170362

Odhadnutý 3. trojúhelník - GMCL model

Rok vzniku	Vývojový trojúhelník			
	1	2	3	1+2+3
1	353584.00	355363.00	106690.00	815637.00
2	350752.15	352567.03	119991.46	823310.65
3	387054.02	389807.14	120743.96	897605.12
4	377481.05	380520.67	130334.85	888336.57
5	391696.42	393479.27	135263.65	920439.34
6	403618.95	404891.15	140391.40	948901.49
7	405169.66	405876.20	146419.13	957464.99
8	397432.57	397954.80	144932.86	940320.23
9	384699.84	384954.63	155973.09	925627.55
10	370633.31	370388.23	170361.51	911383.05
Total	3822121.97	3835802.11	1371101.91	9029025.99

Celkové odhadnuté úhrny škod - GMCL model

Rok vzniku	Vývojový trojúhelník			
	1	2	3	1+2+3
1	0.00	0.00	0.00	0.00
2	229.15	374.03	1420.46	2023.65
3	1830.02	371.14	2384.96	4586.12
4	4156.05	814.67	5987.85	10958.57
5	8958.42	677.27	9083.65	18719.34
6	16893.95	-1611.85	19065.40	34347.49
7	37812.66	5336.20	34432.13	77580.99
8	79460.57	28626.80	52493.86	160581.23
9	137896.84	49447.63	78306.09	265650.55
10	244345.31	41152.23	124076.51	409574.05
Total	531582.97	125188.11	327250.91	984021.99

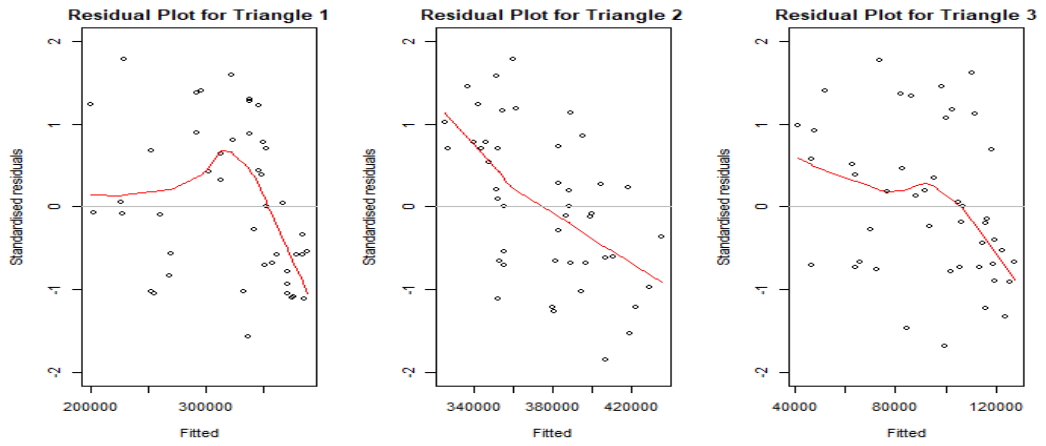
Odhadnuté IBNR rezervy - GMCL model

Porovnání výsledků

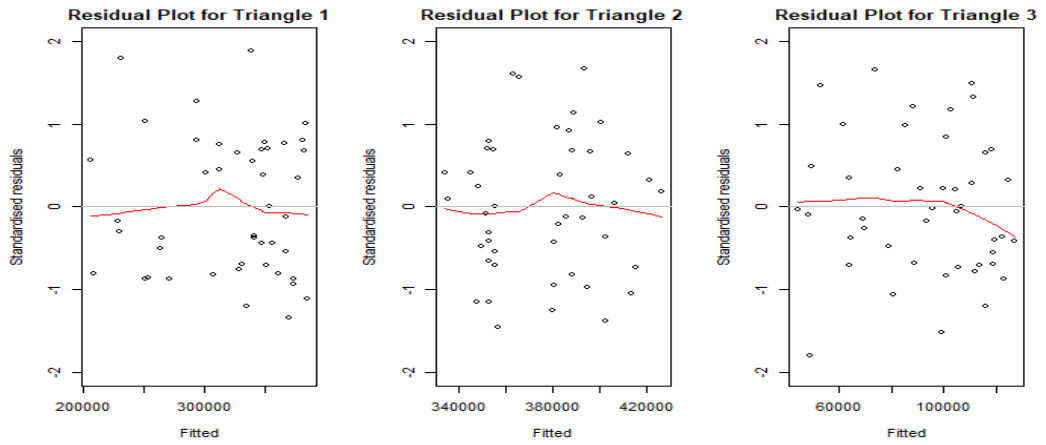
Grafy standardizovaných reziduí

Vykreslíme standardizovaná rezidua a proložíme křivkou, abychom mohli detailněji pozorovat odlišnosti jednotlivých modelů. Standardizovaná rezidua označují rezidua vydělená jejich směrodatnou odchylkou.

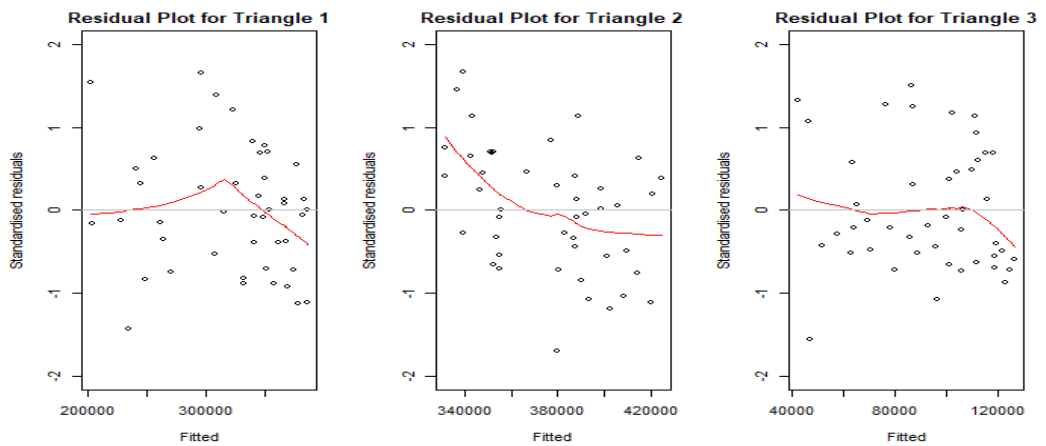
```
## Grafy standardizovaných reziduí
par(mfrow = c(1, 3))          # Volba tří grafů do řádku
# SCL model
plot(scl, which.plot = 3:3, ylim = c(-2,2))
# Zjednodušený GMCL model
plot(mcl, which.plot = 3:3, ylim = c(-2,2))
# GMCL model bez interceptů
plot(full.gmcl1, which.plot = 3:3, ylim = c(-2,2))
# GMCL model
plot(full.gmcl2, which.plot = 3:3, ylim = c(-2,2))
```



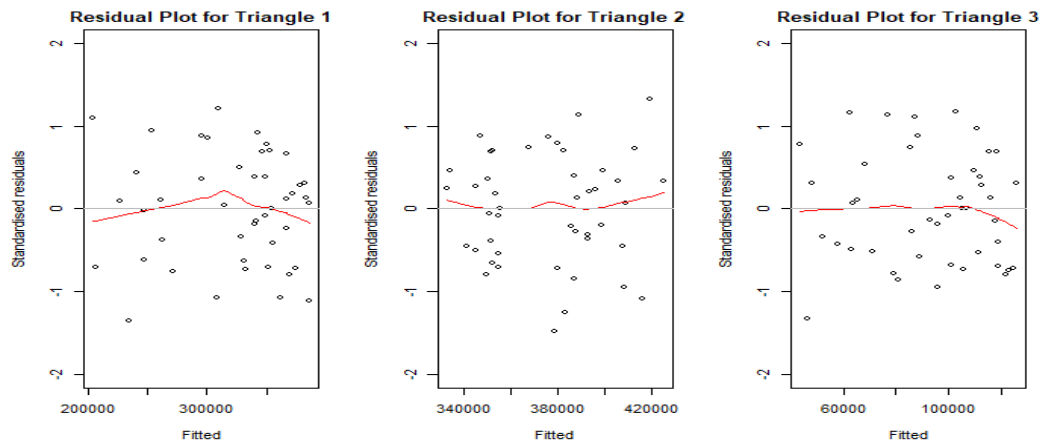
SCL model - standardizovaná rezidua



MCL+int model - standardizovaná rezidua



GMCL-int model - standardizovaná rezidua



GMCL model - standardizovaná rezidua

Porovnání P/I poměrů

V této části vytvoříme tabulku odhadnutých P/I poměrů celkových výší škod pro jednotlivé modely. Náš odhad bude tím kvalitnější, čím blíže si tyto dvě hodnoty (škod uhrazených a vzniklých) budou. Jinými slovy lepší metoda odhadu z tohoto hlediska je taková, jejíž P/I poměr je blíže 1 (100 %). Přičemž jako P chápeme odhadnutý trojúhelník škod uhrazených a jakožto I trojúhelník škod vzniklých.

```
## P/I poměr - Klasický Chain ladder model
loss.scl <- sumup.scl$Ultimate
      # Celkové úhrn škod (pro všechny
      # trojúhelníky)
pi.scl <- loss.scl[,1]/loss.scl[,2]*100
      # Podíl výší škod trojúhelníků P a I
## P/I poměr - Zjednodušený GMCL model
loss.mcl <- sumup.mcl$Ultimate
      # Celkové úhrn škod (pro všechny
      # trojúhelníky)
pi.mcl <- loss.mcl[, 1]/loss.mcl[, 2]*100
      # Podíl výší škod trojúhelníků P a I
## P/I poměr - GMCL model bez interceptů
loss.gmcl1 <- sumup.gmcl1$Ultimate
      # Celkové úhrn škod (pro všechny
      # trojúhelníky)
pi.gmcl1 <- loss.gmcl1[, 1]/loss.gmcl1[, 2]*100
      # Podíl výší škod trojúhelníků P a I
## P/I poměr - GMCL model bez interceptů
loss.gmcl2 <- sumup.gmcl2$Ultimate
      # Celkové úhrn škod (pro všechny
      # trojúhelníky)
pi.gmcl2 <- loss.gmcl2[, 1]/loss.gmcl2[, 2]*100
      # Podíl výší škod trojúhelníků P a I
## Vytvoření tabulky
all.ratios <- cbind(pi.scl, pi.mcl, pi.gmcl1, pi.gmcl2)
```

```

dimnames(all.ratios)[[2]] <- c("SCL",
                              "MCL+int",
                              "GMCL-int",
                              "GMCL")
ratios.table <- round(all.ratios ,2)
## Export tabulky
print(xtable(cbind(ratios.table)), file = "PI_ratios.tex")

```

	SCL	MCL+int	GMCL-int	GMCL
1	99.50	99.50	99.50	99.50
2	99.49	99.49	99.49	99.49
3	99.29	99.29	99.29	99.29
4	99.20	99.20	99.20	99.20
5	99.82	99.46	99.55	99.55
6	100.43	99.64	99.69	99.69
7	103.53	102.08	99.96	99.83
8	111.23	106.73	100.29	99.87
9	122.15	111.17	100.87	99.93
10	126.15	111.37	101.88	100.07
Total	105.56	102.67	99.91	99.64

P/I poměry pro jednotlivé modely