

Katedra pravděpodobnosti a matematické statistiky



MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

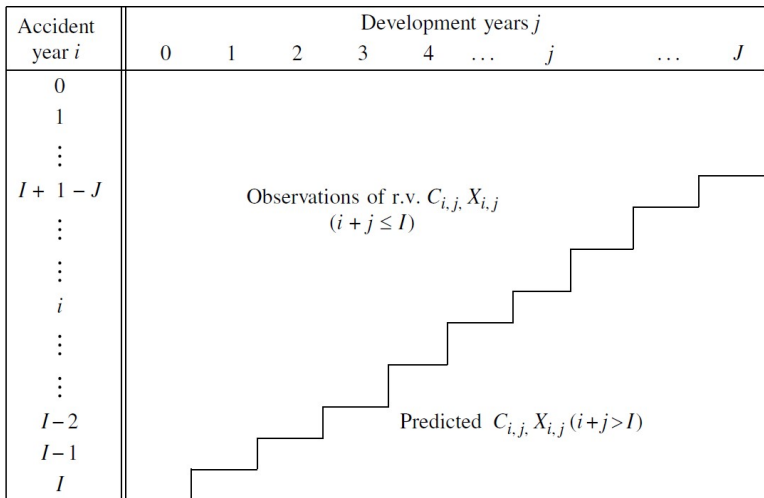
Petr Vejmelka

**Stavové modelování rezerv v neživotním
pojištění**

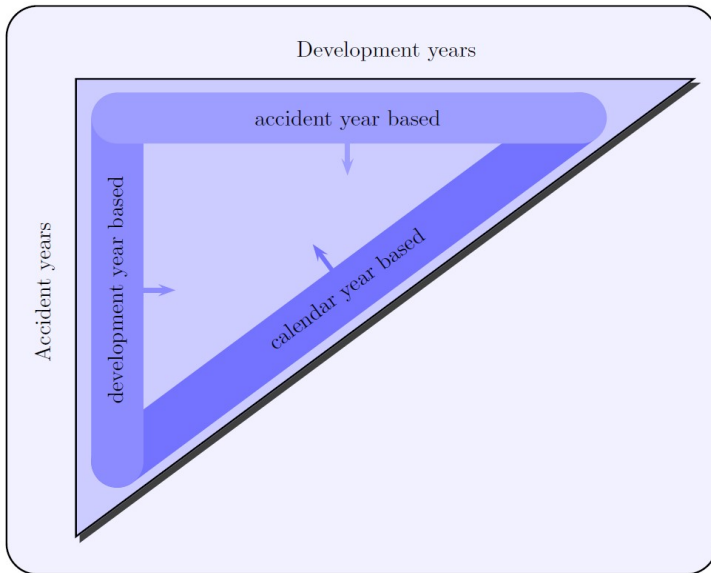
8.12.2023

- 1 Vývojové trojúhelníky ve formě časových řad
- 2 Stavový model
- 3 Logaritmicko-normální model
- 4 Odhad kvantilů rezerv
- 5 Využití clusterování
- 6 Generátor/replikátor portfolií
- 7 Porovnání modelů

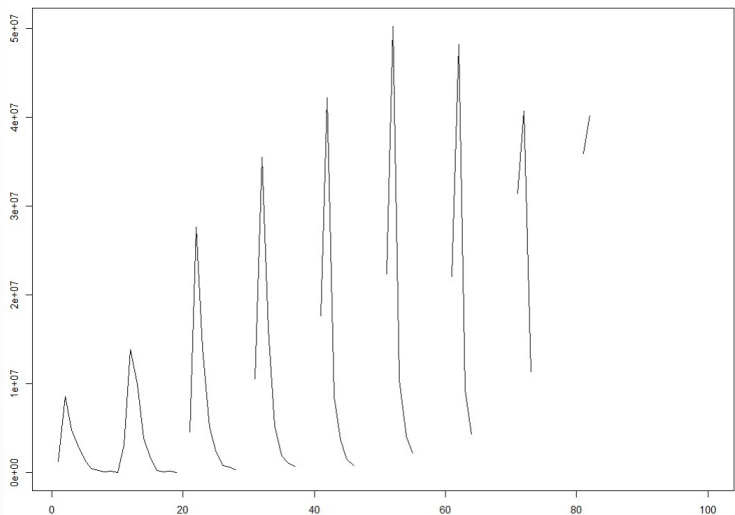
Vývojový trojúhelník



Uspořádání dat



Vývojový trojúhelník ve formátu časové řady



Lineární Gaussovský stavový model

Lineární stavový model (dynamický lineární model) je dán následujícím systémem rovnic:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \end{aligned}$$

- \mathbf{y}_t je p -rozměrný vektor pozorování v čase t ,
- $\boldsymbol{\alpha}_t$ je m -rozměrný stavový vektor v čase t ,
- \mathbf{Z}_t , \mathbf{T}_t a \mathbf{R}_t jsou matice parametrů typů $(p \times m)$, $(m \times m)$ a $(m \times k)$,
- $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{H}_t)$ a $\boldsymbol{\eta}_t \sim N(0, \mathbf{Q}_t)$ nezávislé a vzájemně nekorelované
- \mathbf{H}_t a \mathbf{Q}_t jsou rozptylové matice typů $(p \times p)$ a $(k \times k)$.

Log-normální model

Pro řádkově uspořádaná data uvažujeme:

$$Y_t - Y_{i,0} = \beta_t + \varepsilon_t,$$
$$\beta_{t+1} = \beta_{t-s+1} + \eta_t,$$

kde $Y_t = \log X_t$.

Tvar lineárního stavového modelu:

$$Y_t - Y_{i,0} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \alpha_t + \varepsilon_t,$$
$$\alpha_{t+1} = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \eta_t,$$

kde $\alpha_t = (\beta_t, \dots, \beta_{t-s+1})'$, $\varepsilon_t \stackrel{\text{iid.}}{\sim} N(0, \sigma_\varepsilon^2)$ a $\eta_t \stackrel{\text{iid.}}{\sim} N(0, \sigma_\eta^2)$.

- Nulové nebo záporné hodnoty
- Počáteční odhady
- Inverzní logaritmická transformace

$$\hat{X}_{ij} = \exp \left\{ \hat{Y}_{ij} + \sigma_{\hat{Y}_{ij}}^2 / 2 \right\}$$

Řádkově uspořádaný tvar modelu:

$$\begin{aligned}y_t(n) - y_t^0(n) &= \alpha_t(n) + \varepsilon_t(n), \\ \alpha_{t+1} &= \alpha_{t-s+1}(n) + \eta_t(n),\end{aligned}$$

kde $\varepsilon_t(n) \stackrel{\text{ind.}}{\sim} N(0, \sigma_\varepsilon(n, n))$ a $\eta_t(n) \stackrel{\text{ind.}}{\sim} N(0, \sigma_\eta(n, n))$,
 $\sigma_\varepsilon(m, n) = \text{Cov}(\varepsilon_t(m), \varepsilon_t(n))$ a obdobně pro $\sigma_\eta(m, n)$, kde
 $m, n = 1, \dots, N$.

DLM tvar:

$$\mathbf{y}_t - \mathbf{y}_t^0 = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} \alpha_t + \varepsilon_t,$$

Vícerozměrný log-normální model

$$\alpha_{t+1} = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ & & \ddots & & \\ & & & 0 & 0 & \dots & 0 & 1 \\ & & & 1 & 0 & \dots & 0 & 0 \\ & & & \vdots & \vdots & & \vdots & \vdots \\ & & & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ & & \ddots & & \\ & & & 1 & 0 & \dots & 0 & 0 \\ & & & 0 & 0 & \dots & 0 & 0 \\ & & & \vdots & \vdots & & \vdots & \vdots \\ & & & 0 & 0 & \dots & 0 & 0 \end{pmatrix} \eta_t,$$

kde

$$\varepsilon_t = (\varepsilon_t(1), \dots, \varepsilon_t(N))', \eta_t = (\eta_t(1), 0, \dots, 0, \dots, \eta_t(N), 0, \dots, 0)' \text{ a}$$

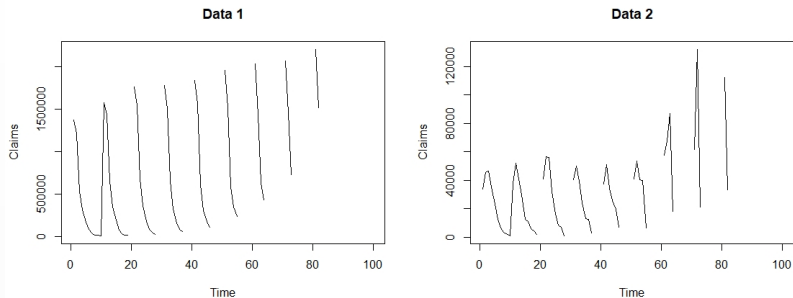
$$\alpha_t = (\beta_t(1), \dots, \beta_{t-s+1}(1), \dots, \beta_t(N), \dots, \beta_{t-s+1}(N))'.$$

Škodní rok	Zpoždění									
	0	1	2	3	4	5	6	7	8	9
0	1,376,384	1,211,168	535,883	313,790	168,142	79,972	39,235	15,030	10,865	4,086
1	1,576,278	1,437,150	652,445	342,694	188,799	76,856	35,042	17,089	12,507	
2	1,763,277	1,540,231	678,959	364,199	177,108	78,169	47,391	25,288		
3	1,779,698	1,498,531	661,401	321,434	162,578	84,581	53,449			
4	1,843,224	1,573,604	613,095	299,473	176,842	106,296				
5	1,962,385	1,520,298	581,932	347,434	238,375					
6	2,033,371	1,430,451	633,500	432,257						
7	2,072,061	1,458,541	727,098							
8	2,210,754	1,517,501								
9	2,206,886									

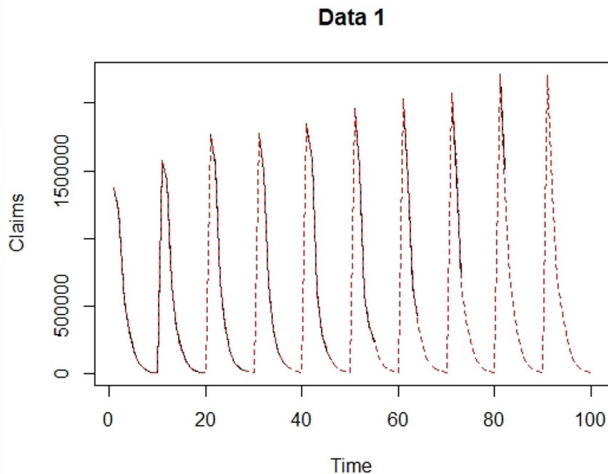
Tabulka: Vývojový trojúhelník - data 1 (auta pro osobní využití)

Škodní rok	Zpoždění									
	0	1	2	3	4	5	6	7	8	9
0	33,810	45,318	46,549	35,206	23,360	12,502	6,602	3,373	2,373	778
1	37,663	51,771	40,998	29,496	12,669	11,204	5,785	4,220	1,910	
2	40,630	56,318	56,182	32,473	15,828	8,409	7,120	1,125		
3	40,475	49,967	39,313	24,044	13,156	12,595	2,908			
4	37,127	50,938	34,154	25,455	19,421	6,728				
5	41,125	53,302	40,289	39,912	6,650					
6	57,515	67,881	86,734	18,109						
7	61,553	132,208	20,923							
8	112,103	33,250								
9	37,554									

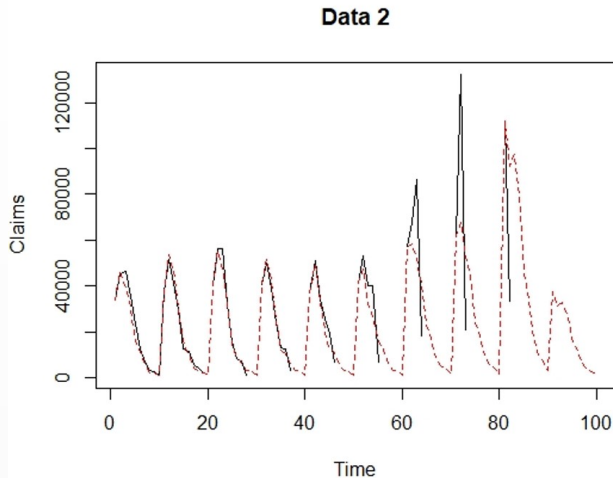
Tabulka: Vývojový trojúhelník - data 2 (auta pro komerční využití)



Obrázek: Převedení vývojových trojúhelníků na časové řady



Obrázek: Známé škody a projekce



Obrázek: Známé škody a projekce

- Rezidua:

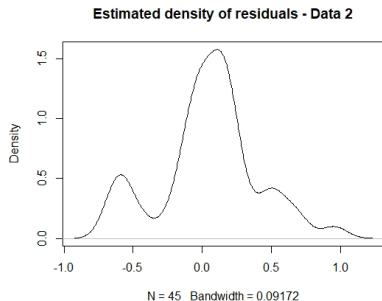
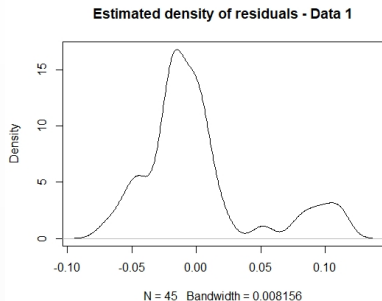
$$r_{ij}^p = \frac{X_{ij} - \hat{X}_{ij}}{\hat{X}_{ij}}.$$

- Jádrový odhad hustoty:

$$\hat{f}_w(x) = \frac{1}{wh_w} \sum_{t=1}^w K\left(\frac{x - r_t^p}{h_w}\right),$$

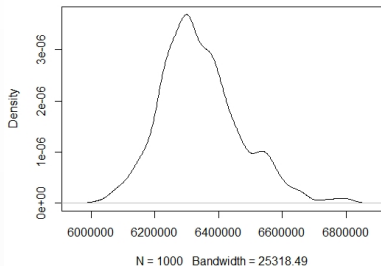
kde w je počet dostupných r_{ij}^p a h_w je šířka vyhlazovacího okna.

- Resamplování B výběrů reziduí a jejich využití pro konstrukci B sad vývojových trojúhelníků
- Výpočet rezerv pro každou resamplovanou sadu vývojových trojúhelníků
- Kvantily a odhad hustoty rezerv

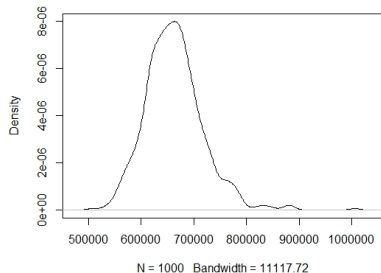


Obrázek: Jádrový odhad hustoty reziduí

Estimated density of IBNR - Data 1



Estimated density of IBNR - Data 2



Obrázek: Jádrový odhad hustoty rezerv

- Anticlust (R balíček *anticlust*)
- K-means (R balíček *stats*)
- Model based clustering (R balíček *mclust*)
- Clustering large applications (R balíček *ClusterR*)

- Funkce *balanced_clustering*
- Clustery mají stejnou velikost
- Procedura
 - 1 Spočítá se vektor průměrů všech sloupců dané matice (centroid)
 - 2 Nejvzdálenější pozorování od centroidu je přiřazeno do clusteru spolu s jeho $\frac{n}{k} - 1$ nejbližšími sousedy
 - 3 Toto se opakuje se zbylými pozorováními dokud nejsou všechna přiřazena do clusterů

- Funkce *Kmeans*
- Procedura
 - 1 k pozorování je náhodně vybráno
 - 2 V závislosti na vzdálenosti zbylých pozorování od vybraných jsou rozřazeny do clusterů
 - 3 Jsou spočteny centroidy a všechna pozorování jsou přeřazena v závislosti na jejich vzdálenostech od centroidů
 - 4 Předchozí krok se opakuje do doby než jsou pozorování ustálena v clusterech

- Funkce *Mclust*
- Procedura
 - 1 Předpokládá se, že data pochází ze směsi hustot
 - 2 Pro různé parametrizace kovarianční matice je spočítán maximálně věrohodný odhad
 - 3 Nejlepší model je zvolen pomocí informačních kritérií
 - 4 Každé pozorování je přiřazeno do clusteru s největší pravděpodobností

- Funkce *Clara_Medoids*
- Lze použít i pro velká portfolia
- Procedura
 - 1 Je zvolen malý výběr dat
 - 2 Aplikuje se PAM algoritmus
 - 3 Zbylá pozorování jsou rozřazena do clusterů na základě jejich vzdálenosti od medoidů
 - 4 Tento postup je proveden několikrát a je vybrán případ dosahující nejlepších výsledků

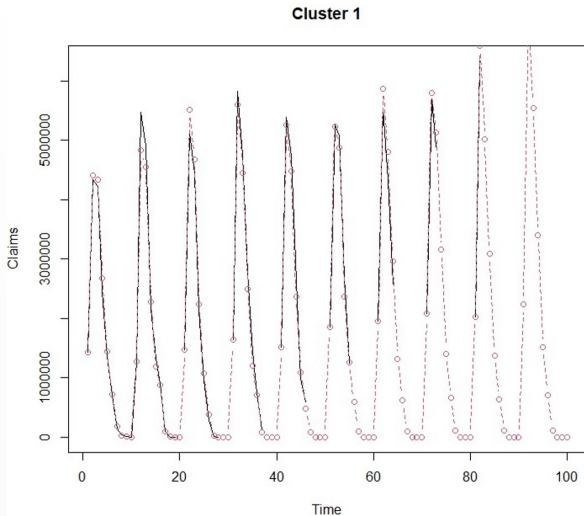
Generátor portfolií škod

- Uvažován generátor navržený ve Wang, M. and Wüthrich, M. (2022)
- Generováno 500 portfolií škod
- V průměru 30 000 škod v každém portfoliu

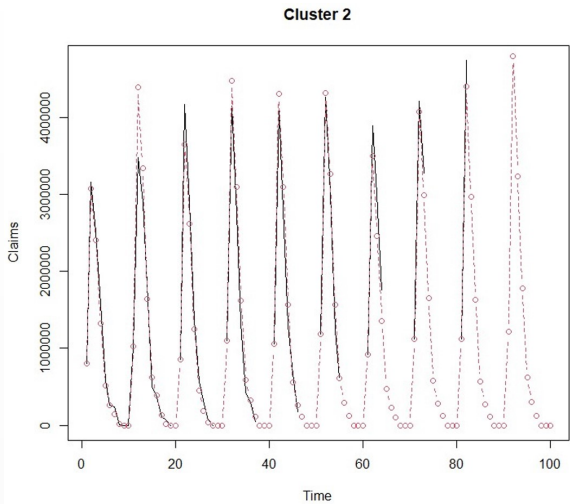
Id	Typ	Počet plateb	Škodní měsíc	Měsíc platby	Výše platby	Celková škoda
1	2	1	1	42	6,175	6,175
2	1	2	1	9	3,609	6,521
2	1	2	1	16	2,912	6,521
3	2	2	1	7	64,314	122,690
3	2	2	1	12	28,376	122,690
4	5	1	1	20	2,330	2,330

Tabulka: Ilustrace dat

- Generovaná pozorování rozdělena do dvou clusterů pomocí uvažovaných clusterovacích metod



Obrázek: Ilustrace známých a vyhlazených hodnot náhodně vybraného portfolia clusterovaného pomocí CLARA metody

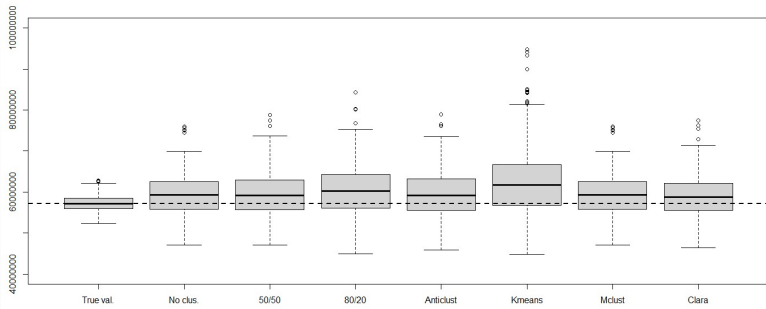


Obrázek: Ilustrace známých a vyhlazených hodnot náhodně vybraného portfolia clusterovaného pomocí CLARA metody

Numerická analýza - porovnání clusterovacích metod

Bez clusterování	50/50	80/20	Anticlust	Kmeans	Mclust	CLARA
$2.131 \cdot 10^9$	$2.366 \cdot 10^9$	$2.754 \cdot 10^9$	$2.452 \cdot 10^9$	$3.517 \cdot 10^9$	$2.752 \cdot 10^9$	$2.074 \cdot 10^9$

Tabulka: Porovnání clusterovacích metod - součet absolutních odchylek rezerv od skutečných hodnot



Obrázek: Boxploty rezerv a skutečných hodnot

- Data poskytnutá Českou kanceláří pojistitelů
- Pojistná plnění hrazená z Garančního fondu ČKP
- Kompenzace za škody způsobené provozem neidentifikovaného vozidla nebo vozidla bez povinného ručení

Id	Typ	Datum škody	Datum nahlášení škody	Datum platby	Výše platby
35121	Škoda na zdraví	14/01/2001	23/12/2002	10/08/2005	3,246
13500	Materiální škoda	15/01/2001	01/03/2001	06/02/2004	27,442
16030	Materiální škoda	15/01/2001	29/05/2001	25/06/2001	1,000
16030	Materiální škoda	15/01/2001	29/05/2001	25/01/2002	9,237
16094	Škoda na zdraví	15/01/2001	30/05/2001	10/08/2001	1,000
16094	Škoda na zdraví	15/01/2001	30/05/2001	26/07/2002	1,845

Tabulka: Ilustrace dostupných dat

- Uvažované roky: 2001-2019 (2011-2019 predikce nevypořádaných škod nastalých v letech 2001-2010)

Generátor replikovaných portfolií

Pro každé replikované portfolio:

- Počet škod



Pro každou škodu:

- Škodní typ



Na základě škodního typu:

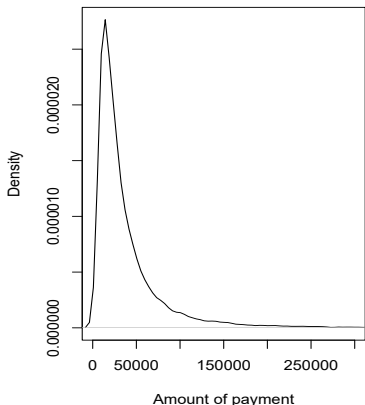
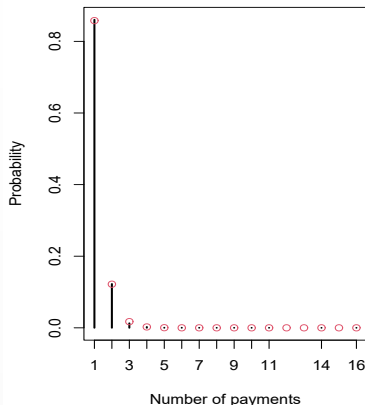
- Počet plateb
- Rok výskytu škod



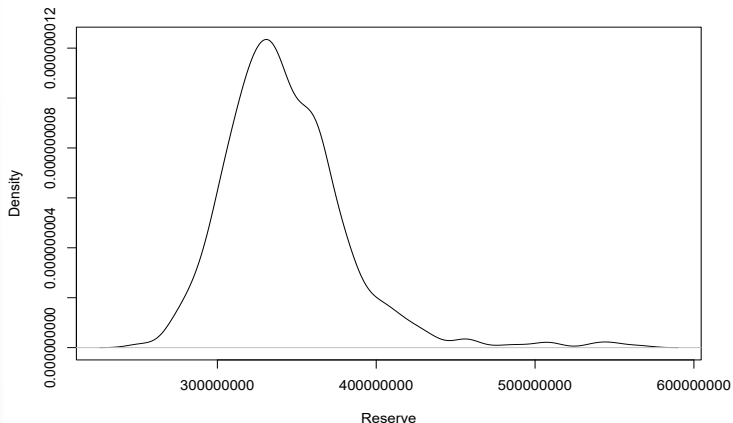
Na základě počtu plateb

- Výše platby
- Zpoždění platby

Generátor replikovaných portfolií



Obrázek: Odhadnuté pravděpodobnosti počtu plateb a odhad hustoty výší plateb pro škody vypořádané jednou platbou (materiální škody)



Obrázek: Odhadnuté rozdělení rezervy

Předpoklady:

- Kumulativní škody $C_{i,j}$ pro různé roky i jsou nezávislé.
- Existují vývojové faktory $f_0, \dots, f_{J-1} > 0$ takové, že pro všechna $0 \leq i \leq I$ a všechna $1 \leq j \leq J$ platí

$$E[C_{i,j}|C_{i,0}, \dots, C_{i,j-1}] = E[C_{i,j}|C_{i,j-1}] = f_{j-1} C_{i,j-1}.$$

Za těchto předpokladů:

$$E[C_{i,j}|D_I] = E[C_{i,j}|C_{i,I-i}] = C_{i,I-i} f_{I-i} \cdots f_{J-1} \quad \text{pro } 1 \leq i \leq I.$$

Odhad vývojových faktorů:

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} \quad \text{pro } j = 0, \dots, J-1.$$

Chain-ladder odhad:

$$\hat{C}_{i,j}^{CL} = \hat{E}[C_{i,j}|D_I] = C_{i,I-i} \hat{f}_{I-i} \cdots \hat{f}_{J-1}.$$

- Tweedieho složené Poissonovo rozdělení s rozptylovou funkcí
 $V(\mu) = \mu^p$
- p je uvažováno v rozmezí $(1, 2)$ a je odhadováno z dat

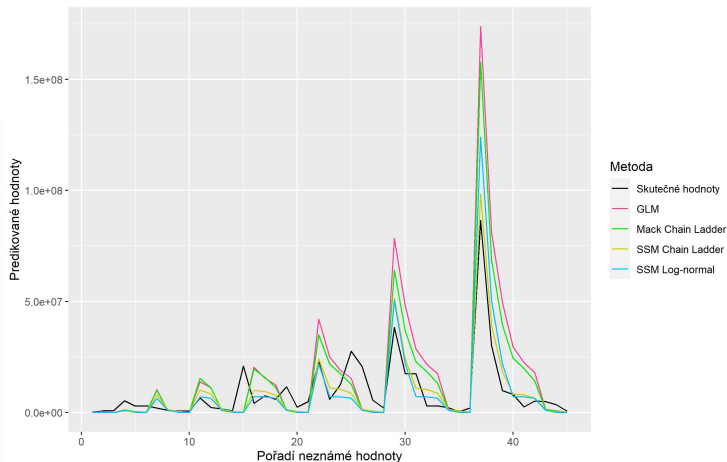
CDR (Claims development result) pro škodní rok i a kalendářní rok k :

$$\begin{aligned} CDR_i^k &= C_{i,J}^k - C_{i,J}^{k-1} \\ &= R_{i,j^*}^k - R_{i,j^*}^{k-1} + AvE_{i,j^*}^k \end{aligned}$$

CDR score pro kalendářní rok k :

$$CDR_{score}^k = \frac{\sum_{i=1}^I |X_{i,j^*}| (CDR_i^k)^2}{\sum_{i=1}^I |X_{i,j^*}|}$$

Porovnání predikcí



Obrázek: Předpovědi neznámých hodnot ve vývojovém trojúhelníku

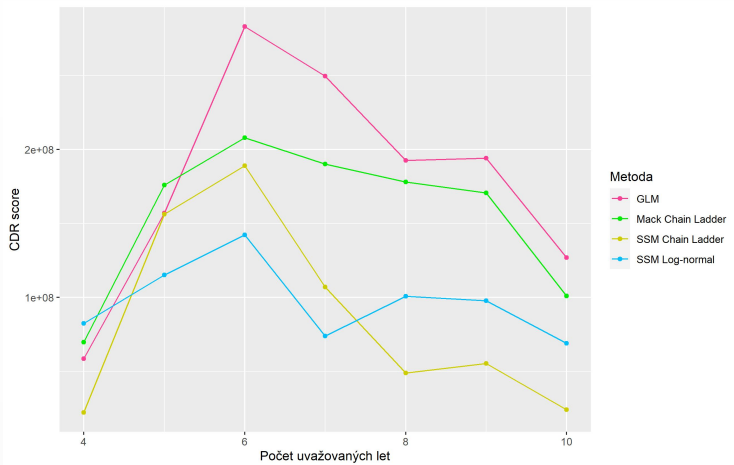
Skutečná	GLM	Chain-ladder	SSM Log-normální	SSM Chain-ladder
4.19×10^8	7.67×10^8	6.64×10^8	4.02×10^8	4.06×10^8

Tabulka: Predikce rezerv







Model	2.5% kvantil	25% kvantil	50% kvantil	75% kvantil	97.5% kvantil
Chain-ladder	4.35×10^8	5.77×10^8	6.66×10^8	7.67×10^8	9.57×10^8
SM Log-normální	2.13×10^8	3.34×10^8	4.21×10^8	5.62×10^8	9.66×10^8

Tabulka: Vybrané kvantily rezerv

Porovnání CDR



Obrázek: Vývoj CDR score pro jednotlivé kalendářní roky

-  Helske, J. (2017): KFAS: Exponential family state space models in R. *Journal of Statistical Software*, 78(10), 1–38.
-  Hendrych, R. & Cipra, T. (2020): Applying state space models to stochastic claims reserving. *ASTIN Bulletin*, 51(1), 267–301.
-  Mouselimis, L. (2023): ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering. *R package version 1.3.1*.
-  Papenberg, M. & Klau, G. W. (2021): Using anticlustering to partition data sets into equivalent parts. *Psychological Methods*, 26(2), 161–174.
-  Scrucca, L. et al. (2016): Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 289–317.
-  Wang, M. & Wüthrich, M. (2022): Individual claims generator for claims reserving studies: Data Simulation.R. *SSRN Electronic Journal*.

Děkuji za pozornost