Conditional Least Squares and Copulae in Claims Reserving for a Single Line of Business

Michal Pešta
Charles University in Prague
Faculty of Mathematics and Physics

Overview

- Motivated by claims reserving in **non-life insurance**
- Joint work with Ostap Okhrin (HU Berlin)
- **Triangular** data models
- published in *Insurance: Mathematics and Economics*, 56 (May 2014): 28-37.

Triangular data

$i \backslash j$	1	2		$\cdots$		$n-1$	n
1	$Y_{1,1}$	$Y_{1,2}$		$\cdots$		$Y_{1,n-1}$	$Y_{1,n}$
2	$Y_{2,1}$	$Y_{2,2}$		$\cdots$		$Y_{2,n-1}$	
			$\ddots$				
$\vdots$	$\vdots$	$\vdots$		$Y_{i,n+1-i}$			
$n-1$	$Y_{n-1,1}$	$Y_{n-1,2}$					
n	$Y_{n,1}$						

- n copies of stochastic process
- The **first** realization consists of n observations
- The **last** one has only one observation

Terminology and goals

- $Y_{i,j} \ldots$ **cumulative payments** in origin year i after j development periods (accounting year $i + j$)
- $n \ldots$ current year – corresponds to the most recent accident year and development period
- Our data history consists of **right-angled isosceles triangles** $Y_{i,j}$, where $i + j \leq n + 1$
- **Predict** $Y_{i,n}$ and $R_i = Y_{i,n} - Y_{i,n+1-i}$ (**claims reserve**)
- **Estimate distribution** of the reserves

Conditional mean and variance (CMV) model

- **CMV model**

$$Y_{i,j} = \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) + \sigma(Y_{i,j-1}, \boldsymbol{\beta}, j)\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

- $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown parameters, which dimensions do not depend on n
- μ is a continuous function in $\boldsymbol{\alpha}$
- σ is a positive and continuous function in $\boldsymbol{\beta}$
- Errors $\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})$

CMV model's errors

- Disturbances $\{\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})\}_{j=1}^{n+1-i}$ are **independent sample copies** of a **stationary first-order Markov process** for all i
- All $\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})$ have the **common true invariant distribution** $G_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ which is absolutely continuous with respect to Lebesgue measure on the real line
- Filtration $\mathcal{F}_{i,j} = \sigma(Y_{k,l} : l \leq j, k \leq i+1-j)$ denotes the information set generated by that trapezoid

$$\mathsf{E}[\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})|\mathcal{F}_{i,j-1}] = 0$$
$$\mathsf{var}[\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})|\mathcal{F}_{i,j-1}] = s(\boldsymbol{\alpha},\boldsymbol{\beta})$$

Properties of the CMV model

- **Unknown true values** $[\boldsymbol{\alpha}^{*\top}, \boldsymbol{\beta}^{*\top}]^\top$ of parameters $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$ set (due to **identifiability** purposes): $s(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = 1$
- Model's name come from the fact that

$$\mathsf{E}[Y_{i,j}|\mathcal{F}_{i,j-1}] = \mu(Y_{i,j-1}, \boldsymbol{\alpha}, j)$$
$$\mathsf{var}[Y_{i,j}|\mathcal{F}_{i,j-1}] = \sigma^2(Y_{i,j-1}, \boldsymbol{\beta}, j)s(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

- **Conditional mean models**: types of ARMA models, vector autoregressions, linear and nonlinear regressions, ...
- **Conditional variance models**: ARCH and any of its numerous parametric extensions (GARCH, EGARCH, GJR-GARCH, etc.), stochastic volatility models, ...

Candidates for the mean and variance function

- From the nature of data: $Y_{i,j} \nearrow C_i \in \mathbb{R}^+$ almost surely as $j \to \infty, \forall i$ (**stabilizing property**)
- One may propose, e.g.,

$$\mu(Y_{i,j-1}, \boldsymbol{\alpha}, j) = \eta(\boldsymbol{\alpha}, j) Y_{i,j-1}$$
$$\sigma(Y_{i,j-1}, \boldsymbol{\beta}, j) = \nu(\boldsymbol{\beta}, j)\sqrt{Y_{i,j-1}}$$

- $\eta(\boldsymbol{\alpha}, j)$ should be **decreasing** in j with **limit** 1 as $j \to \infty$
- $\nu(\boldsymbol{\beta}, j)$ should be **decreasing** in j with **limit** 0 as $j \to \infty$

Dependence modeling

- Since the **mean and variance trends are removed** by the CMV model, the **rest of the relationship among claim amounts** $Y_{i,j}$ can be additionally captured by **modeling dependent errors**
- $\{\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})\}_{j=1}^{n+1-i}$ are independent sample copies of a **stationary first-order Markov process** for all i generated from $(G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\cdot), C(\cdot, \cdot; \gamma))$
- $C(\cdot, \cdot; \gamma)$ is the **true parametric copula** for $[\varepsilon_{i,j-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})]$, which is given and fixed up to unknown parameter γ

Copula-based model

- It is believed that there exist a kind of **information overlap** between the claims from consecutive development periods
- Joint **bivariate distribution** of $[\varepsilon_{i,j-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta})]$ has distribution function

$$H(e_1, e_2) = C(G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_1), G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_2); \boldsymbol{\gamma})$$

- **Conditional copula density** can be derived as

$$h(e_2|e_1) = g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_2)c(G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_1), G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(e_2); \boldsymbol{\gamma})$$

  where c is the **copula density** and $g_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is the **marginal density** corresponding to $G_{\boldsymbol{\alpha},\boldsymbol{\beta}}$
- Play an important role in “making" the dependent errors **conditionally independent**

Parameter estimation

- CMV model with copula assume **three vector parameters** to be estimated
- Estimation process consists of **two stages**
- In the first one, mean and variance parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are estimated in a **distribution-free** fashion, since no specific distributional assumptions are proposed nor required for the claims
- The second stage concerns estimation of the dependence structure, mainly the copula parameter γ, in a **likelihood based** way

Conditional least squares (CLS)

- Denote

$$M_n(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{n-1}\sum_{j=2}^{n}\frac{1}{n+1-j}\sum_{i=1}^{n+1-j} \frac{\left[Y_{i,j}-\mu(Y_{i,j-1},\boldsymbol{\alpha},j)\right]^2}{\sigma^2(Y_{i,j-1},\boldsymbol{\beta},j)}$$

$$V_n(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{n-1}\sum_{j=2}^{n}\frac{1}{n+1-j}\sum_{i=1}^{n+1-j} \left\{\left[Y_{i,j}-\mu(Y_{i,j-1},\boldsymbol{\alpha},j)\right]^2-\sigma^2(Y_{i,j-1},\boldsymbol{\beta},j)\right\}^2$$

CLS estimates

- CLS estimate of the **mean parameter** $\boldsymbol{\alpha}$ for a fixed value of parameter $\boldsymbol{\beta} \in \boldsymbol{\Theta}_2$ is defined as

$$\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\alpha} \in \boldsymbol{\Theta}_1} M_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

  and CLS estimate of the **variance parameter** $\boldsymbol{\beta}$ for a fixed value of parameter $\boldsymbol{\alpha} \in \boldsymbol{\Theta}_1$ is defined as

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \arg\min_{\boldsymbol{\beta} \in \boldsymbol{\Theta}_2} V_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

- **Computationally not feasible** to find the global minimum of M_n and V_n with respect to $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$ simultaneously

Consistency

- Under **regularity conditions**

$$\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) \xrightarrow[n\to\infty]{\mathsf{P}} \boldsymbol{\alpha}^*(\boldsymbol{\beta}),\ \forall\boldsymbol{\beta}; \quad \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \xrightarrow[n\to\infty]{\mathsf{P}} \boldsymbol{\beta}^*(\boldsymbol{\alpha}),\ \forall\boldsymbol{\alpha}$$

- **Mixingales** are to mixing processes as martingale differences are to independent processes

Iterative CLS

- What is the connection between the true unknown parameter values $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ of the CMV model and true unknown parameter values $\boldsymbol{\alpha}^*(\boldsymbol{\beta})$ and $\boldsymbol{\beta}^*(\boldsymbol{\alpha})$?

$$\begin{bmatrix} \widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}^*) \\ \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}^*) \end{bmatrix} \xrightarrow[n\to\infty]{\mathsf{P}} \begin{bmatrix} \boldsymbol{\alpha}^* \\ \boldsymbol{\beta}^* \end{bmatrix}$$

- **Iteratively estimate** $\boldsymbol{\alpha}$ given the fixed value of $\boldsymbol{\beta}$ and, consequently, estimate $\boldsymbol{\beta}$ given the fixed value of $\boldsymbol{\alpha}$ (obtained from previous step)
- **Repeat in turns** until almost no change in consecutive estimates of $[\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$

Estimation of dependence structure

- Estimate the unknown **marginal distribution** function $G_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ of CMV model errors $\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})$ non-parametrically by the **empirical distribution function**

$$\widehat{G}_n(e) = \frac{1}{n(n-1)/2+1} \sum_{i=1}^{n-1} \sum_{j=2}^{n+1-i} \mathcal{I}\{\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}) \leq e\}$$

of the **fitted residuals**

$$\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}) = \frac{Y_{i,j} - \mu(Y_{i,j-1},\widehat{\boldsymbol{\alpha}},j)}{\sigma(Y_{i,j-1},\widehat{\boldsymbol{\beta}},j)}$$

Likelihood for copula

- **Full log-likelihood** for copula parameter γ

$$\mathcal{L}(\gamma) = \sum_{i=1}^{n-2} \sum_{j=2}^{n+1-i} \log g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})) + \sum_{i=1}^{n-2} \sum_{j=3}^{n+1-i} \log c(G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\varepsilon_{i,j-1}(\boldsymbol{\alpha},\boldsymbol{\beta})), G_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\varepsilon_{i,j}(\boldsymbol{\alpha},\boldsymbol{\beta})); \gamma)$$

Psuedo (quasi) likelihood

- Ignoring the first term in $\mathcal{L}(\gamma)$ and replacing ε's and $G_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ by their estimated counterparts $\widehat{\varepsilon}$'s and $\widehat{G}_n$, parameter γ can be estimated by the so-called canonical maximum likelihood, i.e., maximizing the **partial (pseudo) log-likelihood**

$$\widehat{\gamma} = \arg\max_{\gamma} \widetilde{\mathcal{L}}(\gamma)$$

$$\widetilde{\mathcal{L}}(\gamma) = \sum_{i=1}^{n-2} \sum_{j=3}^{n+1-i} \log c\left(\widehat{G}_n(\widehat{\varepsilon}_{i,j-1}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})), \widehat{G}_n(\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})); \gamma\right)$$

Prediction

- Predictor for reserve $R_i^{(n)}$ can be defined as

$$\widehat{R}_i^{(n)} = \widehat{Y}_{i,n} - Y_{i,n+1-i}$$

- Prediction of unobserved claims may be done in a **telescopic** way based on the CMV model formulation: start with the diagonal element $Y_{i,n+1-i}$ and predict $Y_{i,j}$, $j > n+1-i$ stepwise in each row

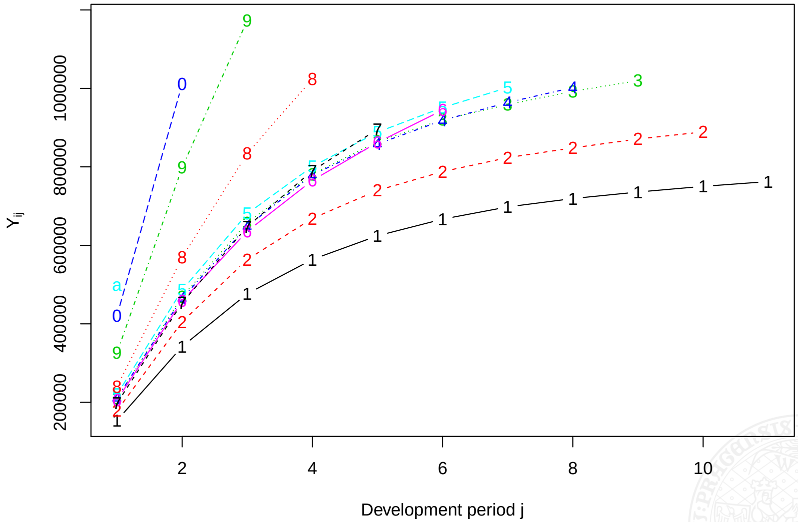
$$\widehat{Y}_{i,j} = Y_{i,j}, \quad i+j \leq n+1$$
$$\widehat{Y}_{i,j} = \mu(\widehat{Y}_{i,j-1}, \widehat{\boldsymbol{\alpha}}, j) + \sigma(\widehat{Y}_{i,j-1}, \widehat{\boldsymbol{\beta}}, j)\widetilde{\varepsilon}_j, \quad i+j > n+1$$

Semiparametric bootstrap

- Errors $\widetilde{\varepsilon}_j$ are simulated from the fitted residuals
- Takes **advantage of the fact** that $\varepsilon_{i,j}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = G^{-1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(X_j)$ for all i (due to the independent rows), where $\{X_j\}_{j=2}^{n}$ is a stationary first-order Markov process with the copula $C(x_1, x_2; \gamma)$ being the joint distribution of $[X_{j-1}, X_j]$

Resampling algorithm

- Generate $n-1$ independent $Un(0,1)$ rvs $\{X_j\}_{j=2}^n$
- Repeat $b = 1, \ldots, B$
- ${}_{(b)}U_2 \leftarrow X_2$
- ${}_{(b)}\widehat{\varepsilon}_2 \leftarrow \widehat{G}_n^{-}({}_{(b)}U_2)$
- ${}_{(b)}U_j \leftarrow C_{2|1}^{-1}(X_j|{}_{(b)}U_{j-1}; \widehat{\gamma}),\ j = 3, \ldots, n$
- ${}_{(b)}\widehat{\varepsilon}_j \leftarrow \widehat{G}_n^{-}({}_{(b)}U_j),\ j = 3, \ldots, n$
- Center bootstrap residuals ${}_{(b)}\widetilde{\varepsilon}_j \leftarrow {}_{(b)}\widehat{\varepsilon}_j - \frac{1}{n-1}\sum_{l=2}^n {}_{(b)}\widehat{\varepsilon}_l$
- ${}_{(b)}\widehat{Y}_{i,n+1-j} \leftarrow Y_{i,n+1-i}$
- ${}_{(b)}\widehat{Y}_{i,j} \leftarrow \mu({}_{(b)}\widehat{Y}_{i,j-1}, \widehat{\boldsymbol{\alpha}}, j) + \sigma({}_{(b)}\widehat{Y}_{i,j-1}, \widehat{\boldsymbol{\beta}}, j)\,{}_{(b)}\widetilde{\varepsilon}_j$,
  $j = n+2-i, \ldots, n$
- ${}_{(b)}\widehat{R}_i^{(n)} \leftarrow {}_{(b)}\widehat{Y}_{i,n} - Y_{i,n+1-i}$


Y_ij
200000
400000
600000
800000
1000000
2
4
6
8
10
Development period j

Real data

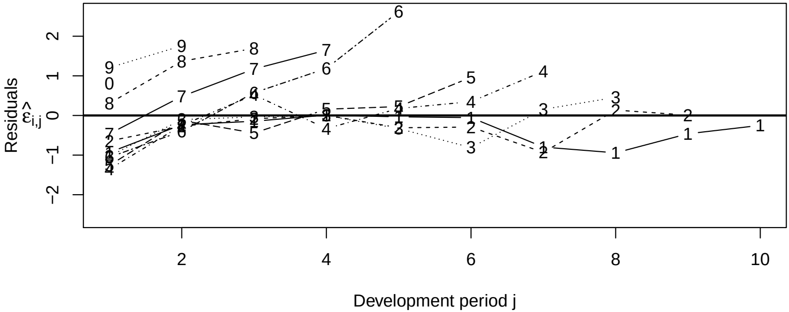
- Data set from Zehnwirth and Barnett (2000)

$$\mu(Y_{i,j-1},\boldsymbol{\alpha},j) = \left(1+\alpha_1\alpha_2 j^{-1-\alpha_2}\exp\left\{\alpha_1 j^{-\alpha_2}\right\}\right) Y_{i,j-1}$$

$$\sigma(Y_{i,j-1},\boldsymbol{\beta},j) = \beta_1 \exp\{-\beta_2 j\}\sqrt{Y_{i,j-1}}$$

- CLS estimates:

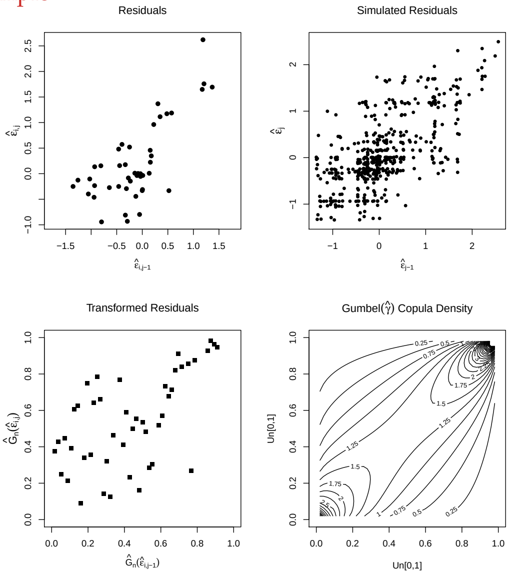
$$\widehat{\alpha}_1 = 2.033,\ \widehat{\alpha}_2 = 1.106,\ \widehat{\beta}_1 = 109.8,\ \widehat{\beta}_2 = 0.4053$$



- Still some slight **pattern** (trend) not captured by mean and variance parametric part

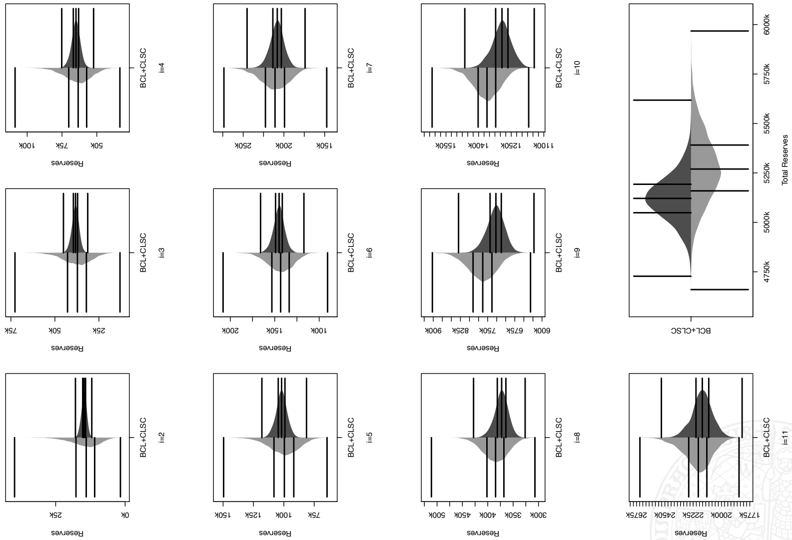
Copula godness-of-fit

- Kendall τ for the pairs of consecutive residuals $\{[\widehat{\varepsilon}_{i,j-1}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}),\widehat{\varepsilon}_{i,j}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})]\}_{i=1,j=3}^{n-2,n+1-i}$ equals 0.43, which indicates **at least mild dependence**
- Three Archimedean copulae (Clayton, Frank, and Gumbel) together with Gaussian and Student t_5-copula considered
- $S_n^{(C)}$ goodness-of-fit test proposed by Genest et al. (2009)
- **Gumbel copula** ($\widehat{\gamma} = 1.776$) was chosen
- Exhibits strong right tail dependence and relatively weak left tail dependence
- Transformed residuals (by the residuals' marginal edf $\widehat{G}_n$; having uniform margins) seem to be strongly correlated at high values but less correlated at low values


Residuals
$\hat{\varepsilon}_{i,j}$
$\hat{\varepsilon}_{i,j-1}$
Simulated Residuals
$\hat{\varepsilon}_{j}$
$\hat{\varepsilon}_{j-1}$
Transformed Residuals
$\hat{G}_n(\hat{\varepsilon}_{i,j})$
$\hat{G}_n(\hat{\varepsilon}_{i,j-1})$
Gumbel($\hat{\gamma}$) Copula Density
Un[0,1]
Un[0,1]

Results

- Benchmark: traditional **bootstrapped chain ladder** (BCL)
- **Disadvantages**, which can be overcome by our approach:
  - Number of parameters depending on the sample size
  - Some parameters estimated by just ratio of two numbers (yielding zero sample variance)
  - Questionable consistency of the estimates
  - Non-realistic assumption of independence of the residuals
- Our approach:
  - Slightly smaller predictions of reserves
  - But even more important is that the estimates of the reserves' distribution are **less volatile**


BCL+CLSC
Reserves
Total Reserves
i=2
i=3
i=4
i=5
i=6
i=7
i=8
i=9
i=10
i=11
4750k
5000k
5250k
5500k
5750k
6000k

Summary

- **Conditional mean and variance (CMV)** time series model for triangular data with innovations being a **stationary first-order Markov process**
- Framework is demonstrated to be suitable for stochastic claims reserving in general insurance
- Very **flexible modeling approach**, relatively **smaller number of model parameters** not depending on the number of development periods, and time series **innovations not** considered as **independent**
- Increase in precision of the claims reserves' prediction
- Theoretical **justification** of the proposed approached shown

Thank you !

Michal.Pesta@mff.cuni.cz