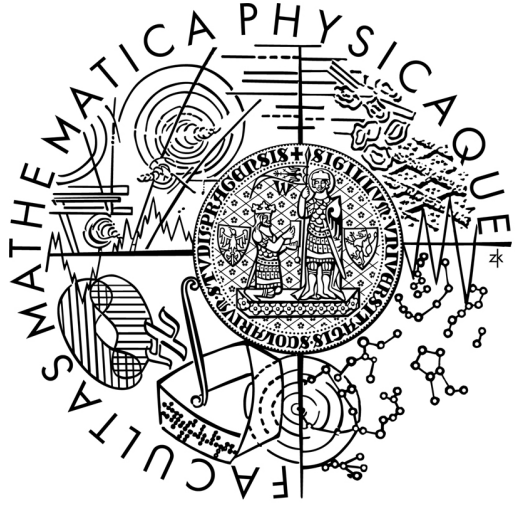
Modeling Dependencies in Claims Reserving with GEE

Šárka Hudecová and Michal Pešta
Charles University in Prague
Faculty of Mathematics and Physics



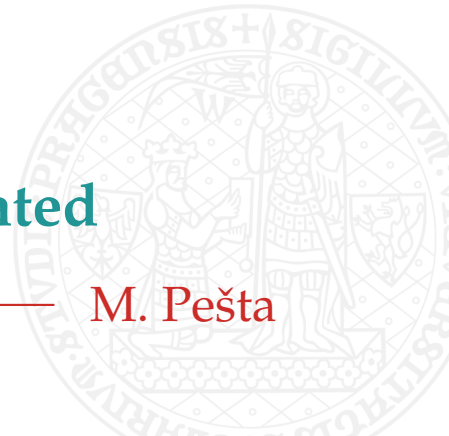
Overview

- Motivated by claims reserving in **non-life insurance**
- Joint work with Šárka Hudecová
- **Triangular** data models
- Published in *Insurance: Mathematics and Economics*, 53(3): 786–794.

Triangular data

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

- Incomplete **2-way** contingency table
- Comprise by **ordered-in-time** observations
- Natural to suspect the observations to be **correlated**

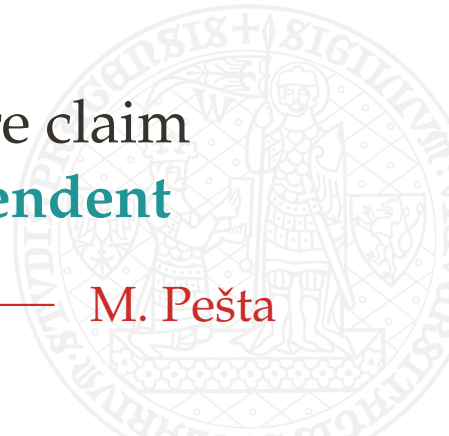


Terminology and goals

- $X_{i,j}$ ... **incremental claims** in accident year i for development period j (accounting year $i+j$)
- n ... current year – corresponds to the most recent accident year and development period
- Our data history consists of **right-angled isosceles triangles** $X_{i,j}$, where $i+j \leq n+1$
- **Predict** $Y_{i,n} = \sum_{j=1}^{n} X_{i,j}$ and $R_i = \sum_{j=n+2-i}^{n} X_{i,j}$ (**claims reserve**)
- Total reserve $R = \sum_{i=1}^{n} R_i$

Objectives

- A common approach to the claims reserving problem is based on **generalized linear models (GLM)**, where the claims in different origin and development years are assumed to be **independent** variables
- If this is **violated**, the classical techniques may provide incorrect predictions of the claims reserves or even misleading estimates of the prediction error
- The application of **generalized estimating equations (GEE)** for estimation of the claims reserves
- Claim triangles are handled as **panel data**, where claim amounts within the same accident year are **dependent**

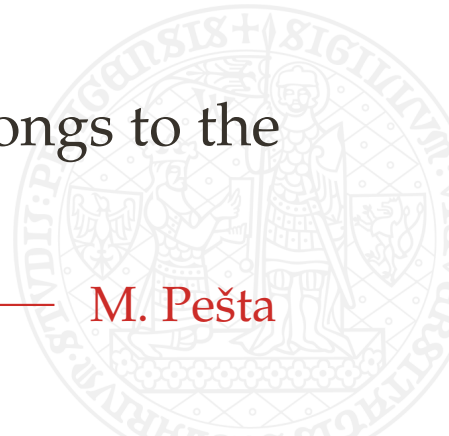


Dependencies

- An assumption that the claim amounts in different years are independent variables can be sometimes **unrealistic or at least questionable**
- It has been pointed out that methods, which enable modeling the dependencies, are needed, cf. Antonio et al. (2006) or Antonio and Beirlant (2007)
- These papers suggest the **generalized linear mixed models (GLMM)** (extension of GLM)
- Generalized estimating equations (GEE) were introduced by Liang and Zeger (1986) as a method for estimating model parameters if the independence assumption is violated

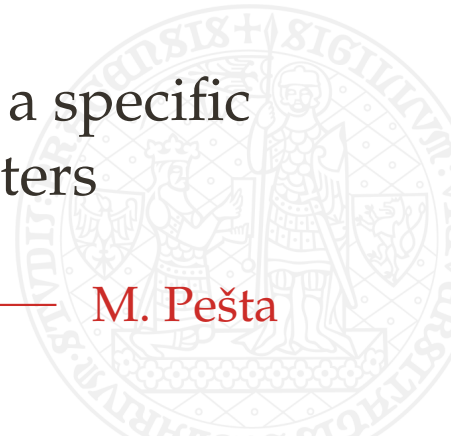
Generalized estimating equations

- The primary interest of the GEE is to **model the marginal expectation of the response** variable given the covariates
- In contrast to GLMM, this method does **not explicitly model the correlation structure**
- The associations are treated as nuisance parameters and modeled using so called "**working correlation matrices**"
- **No additional distributional assumptions** are required, compared to a specific probability distribution for the outcome in the GLM/GLMM framework
- The GEE solely assume that the distribution belongs to the **exponential family**



Assumptions

- Assume that the observations of a **common accident year** are **correlated**
- Incremental claims for accident year $i \in \{1, \ldots, n\}$ form an $(n - i + 1) \times 1$ vector $\mathbf{X}_i = [X_{i,1}, \ldots, X_{i,n-i+1}]^\top$
- Observations of **different accident years** are supposed to be **independent**
- I.e, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent, but the components of $\mathbf{X}_i$ are allowed to be correlated
- Hence, the run-off triangle can be considered as a specific type of **panel data** with accident year (row) clusters



Pillar I – link function

- Denote the expectation of $\mathbf{X}_i$ as

$$\mathsf{E}\mathbf{X}_i = \boldsymbol{\mu}_i = [\mu_{i,1}, \ldots, \mu_{i,n-i+1}]^\top$$

- Suppose that accident year i and development year j influence the expectation via **link function** g

$$\mu_{i,j} = g^{-1}(\eta_{i,j}) = g^{-1}(\mathbf{z}_{i,j}^\top \boldsymbol{\theta}),$$

  where $\mathbf{z}_{i,j}$ is a $p \times 1$ vector of dummy covariates that arranges the impact of accident and development year on the claim amounts through model parameters $\boldsymbol{\theta} \in \mathbb{R}^{p \times 1}$

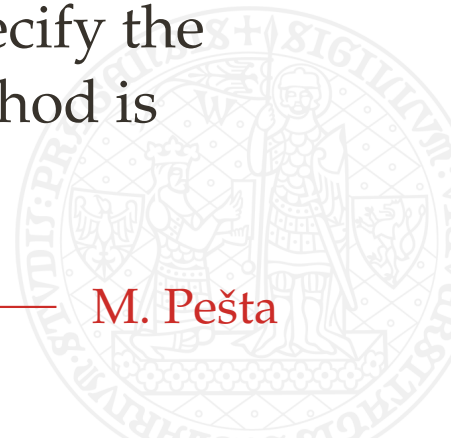
- The **linear predictor** $\eta_i = \mathbf{z}_{i,j}^\top \boldsymbol{\theta}$ together with the link function g fully specifies the **mean structure** $\boldsymbol{\mu}_i$

Pillar II – variance function

- Assume that the variance of the incremental claim can be expressed as a **variance function** h of its expectation

$$\mathrm{Var}\, X_{i,j} = \phi h(\mu_{i,j}),$$

  where $\phi > 0$ is a **dispersion (scale) parameter**
- In connection to the GLM, if $X_{i,j}$ followed the gamma distribution, then $h(x) = x^2$
- In the GEE framework, it is **not necessary** to specify the **whole distribution** of the data, because the method is quasi-likelihood based



Pillar III – correlation structure

- Correlation between the components of $\mathbf{X}_i$ is modeled using a **working correlation matrix** $\mathbf{C}_i(\boldsymbol{\vartheta}) \in \mathbb{R}^{(n-i+1)\times(n-i+1)}$
- Depends only on an $s \times 1$ vector of unknown parameters $\boldsymbol{\vartheta}$, which is the same for all the accident years i
- Consequently, the **working covariance matrix** of the incremental claims is

$$\boldsymbol{V}_i = \mathrm{Cov}\mathbf{X}_i = \phi \mathbf{A}_i^{1/2} \mathbf{C}_i(\boldsymbol{\vartheta}) \mathbf{A}_i^{1/2},$$

  where $\mathbf{A}_i$ is an $(n-i+1) \times (n-i+1)$ diagonal matrix with $h(\mu_{i,j})$ as the jth diagonal element
- The name "working" comes from the fact that the structure of $\mathbf{C}_i$ does not need to be correctly specified

Choice of mean structure

- **Logarithm link** function with mean structure

$$\log(\mu_{i,j}) = \gamma + \alpha_i + \beta_j,$$

  where α_i is the effect of accident year i, β_j is the effect of the development year j, and γ is baseline ($\alpha_1 = 0 = \beta_1$)
- In this case $\boldsymbol{\theta} = [\gamma, \alpha_2, \ldots, \alpha_n, \beta_2, \ldots, \beta_n]$ and

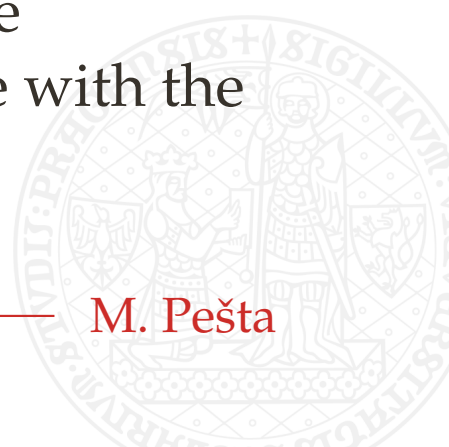
$$\mathbf{z}_{i,j} = [1, \delta_{2,i}, \ldots, \delta_{n,i}, \delta_{2,j}, \ldots, \delta_{n,j}]^\top$$

- **Hoerl curve** with the logarithmic link function

$$\mathbf{z}_{i,j} = [1, \delta_{2,i}, \ldots, \delta_{n,i}, 1 \times \delta_{2,j}, \ldots, n \times \delta_{n,j},$$
$$\delta_{2,j} \times \log 2, \ldots, \delta_{n,j} \times \log n]^\top$$
$$\boldsymbol{\theta} = [\gamma, \alpha_2, \ldots, \alpha_n, \beta_2, \ldots, \beta_n, \lambda_2, \ldots, \lambda_n]^\top$$

Choice of variance function

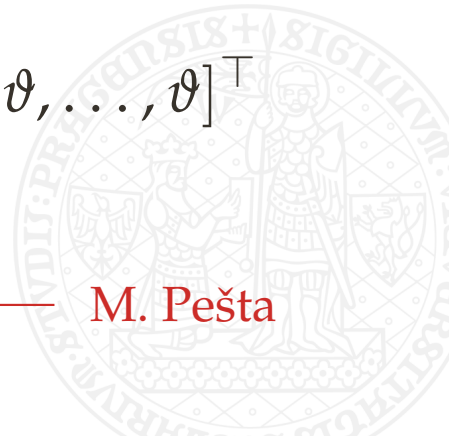
- The choice of the variance function is somehow **analogous** to the specification of the distribution in the GLM
- Suitable variance functions for the claims reserving purposes are: **linear** (its quasi-score vector corresponds to the score vector of the overdispersed Poisson distribution) or **quadratic** (gamma distribution)
- The variance function as a **non-integer power** of the mean (multiplied by the scaling parameter) can also be a practical choice if one realizes the concordance with the Tweedie distribution



Choice of correlation structure

- The simplest case **uncorrelated** (or **independent**) incremental claims, i.e., $\mathbf{C}_i(\boldsymbol{\vartheta}) = \mathbf{I}_{n-i+1} = \{\delta_{j,k}\}_{j,k=1}^{n-i+1,n-i+1}$
- The opposite extreme case is an **unstructured** correlation matrix $\mathbf{C}_i(\boldsymbol{\vartheta}) = \{\vartheta_{j,k}\}_{j,k=1}^{n-i+1,n-i+1}$ such that $\vartheta_{j,j} = 1$ and $\mathbf{C}_i(\boldsymbol{\vartheta})$ is positive definite
- As a compromise to these two extreme cases, one can consider an **exchangeable** correlation structure

$$\mathbf{C}_i(\boldsymbol{\vartheta}) = \{\delta_{j,k} + (1 - \delta_{j,k})\vartheta\}_{j,k=1}^{n-i+1,n-i+1}, \boldsymbol{\vartheta} = [\vartheta, \ldots, \vartheta]^\top$$



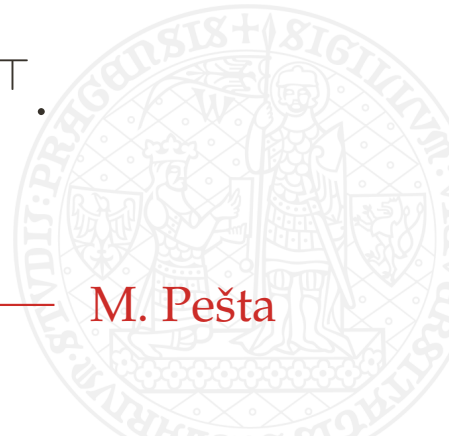
Choice of correlation structure (cont.)

- m-**dependent** correlation structure
  $\mathbf{C}_i(\boldsymbol{\vartheta}) = \{c_{j,k}\}_{j,k=1}^{n-i+1,n-i+1}$,

$$c_{j,k} = \begin{cases} 1, & j = k, \\ \vartheta_{|j-k|}, & 0 < |j-k| \leq m, \\ 0, & |j-k| > m; \end{cases} \quad \boldsymbol{\vartheta} = \{\vartheta_l\}_{l=1}^{m},$$

- **Autoregressive AR(1)** correlation structure

$$\mathbf{C}_i(\boldsymbol{\vartheta}) = \{\vartheta^{|j-k|}\}_{j,k=1}^{n-i+1,n-i+1}, \quad \boldsymbol{\vartheta} = [\vartheta, \ldots, \vartheta]^\top.$$



Estimation

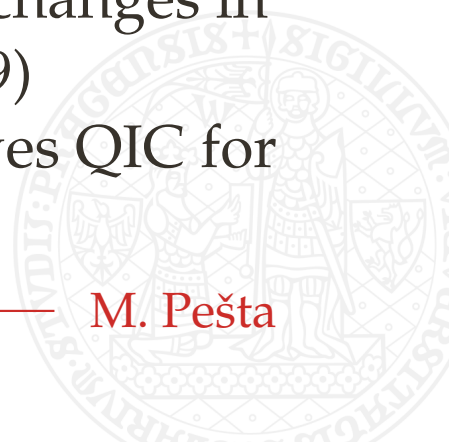
- The generalized estimating equations are formed via **quasi-score vector**

$$\mathbf{u}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{D}_i^\top \boldsymbol{V}_i^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_i),$$

  where $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\theta} \equiv \left\{ \partial \mu_{i,j} / \partial \theta_k \right\}_{j,k=1}^{n-i+i,p}$
- For given estimates $(\widehat{\phi}, \widehat{\boldsymbol{\vartheta}})$ of $(\phi, \boldsymbol{\vartheta})$, the estimate of parameter $\boldsymbol{\theta}$ solves the equation $\mathbf{u}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$
- The parameters $(\phi, \boldsymbol{\vartheta})$ are usually estimated by the moment estimates
- The fitting algorithm is iterative: updating the estimate of $\boldsymbol{\theta}$ in one step and re-estimating $(\phi, \boldsymbol{\vartheta})$ in the second step

Model selection

- Analogy of the AIC by Pan (2001): **quasi-likelihood under the independence model criterion** (QIC)
- A **smaller QIC** value indicates a **better fit** to the data
- The QIC equals to AIC (up to a constant) under the independence
- Hardin and Hilbe (2003): a modified version of QIC
- The main advantage of this modification is computational efficiency
- Simulations have shown that QIC tends to be more sensitive to changes in the mean structure than changes in the covariance structure, cf. Hin and Wang (2009)
- **Correlation information criterion** (CIC) improves QIC for selecting the appropriate working correlation



Mean square error of prediction

- Prediction $\widehat{R}_i^{(n)}$ for the ith claims reserve R_i

$$\mathsf{MSE}\left[\widehat{R}_i^{(n)}\right] := \mathsf{E}\left[\widehat{R}_i^{(n)} - R_i^{(n)}\right]^2 = \mathsf{E}\left[\sum_{j=n+2-i}^{n}\left(\widehat{X}_{i,j}^{(n)} - X_{i,j}\right)\right]^2$$

$$= \sum_{j=n+2-i}^{n}\mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]^2 + \sum_{\substack{j,k=n+2-i \\ j\neq k}}^{n}\mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]\left[\widehat{X}_{i,k}^{(n)} - X_{i,k}\right]$$

where $\widehat{X}_{i,j}^{(n)} = g^{-1}\left(\mathbf{z}_{i,j}^{\top}\widehat{\boldsymbol{\theta}}\right)$ is the **plug-in prediction** of $X_{i,j}$ based on the GEE estimate $\widehat{\boldsymbol{\theta}}$

Drawbacks of current approaches

- Elaborating the expected value from the first sum

$$\begin{aligned}
&\mathrm{MSE}\left[\widehat{X}_{i,j}^{(n)}\right] := \mathrm{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]^2 \\
&= \mathrm{Var}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right] + \left(\mathrm{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]\right)^2 \\
&= \mathrm{Var}\,\widehat{X}_{i,j}^{(n)} - 2\mathrm{Cov}\left(\widehat{X}_{i,j}^{(n)}, X_{i,j}\right) + \mathrm{Var}\,X_{i,j} + \left(\mathrm{E}\widehat{X}_{i,j}^{(n)} - \mathrm{E}X_{i,j}\right)^2
\end{aligned}$$

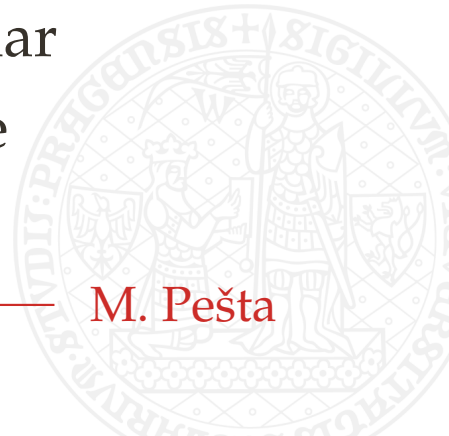
- Many authors directly and inappropriately assume the unbiasedness or approximate unbiasedness of estimator $\widehat{X}_{i,j}^{(n)}$ for $\mathrm{E}X_{i,j}$, that is $\mathrm{E}\widehat{X}_{i,j}^{(n)} = \mathrm{E}X_{i,j}$ or $\mathrm{E}\widehat{X}_{i,j}^{(n)} \approx \mathrm{E}X_{i,j}$

Drawbacks of current approaches (cont.)

- The **conjecture** of the (approximately) unbiased predictor then implies that the MSE of prediction for incremental claims is given as

$$\mathsf{MSE}\left[\widehat{X}_{i,j}^{(n)}\right] \equiv \mathsf{E}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]^2 \approx \mathsf{Var}\left[\widehat{X}_{i,j}^{(n)} - X_{i,j}\right]$$

- However, the unbiasedness of $\widehat{X}_{i,j}^{(n)}$ can be arguable (or even unrealistic), for smaller samples in particular
- MSEP $\neq$ process variance + estimation variance



Estimation of the MSE of prediction

- A first-order stochastic Taylor expansion and some algebra

$$\widehat{\mathsf{MSE}}\left[\widehat{R}^{(n)}\right] = \sum_{i=2}^{n} \underbrace{[1,\ldots,1]}_{(i-1)\times 1} \widehat{\mathsf{MSE}}\left[\widehat{\vec{\mathbf{X}}}_i^{(n)}\right] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

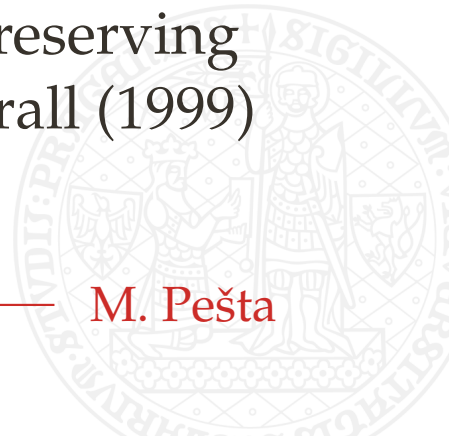
$$\widehat{\mathsf{MSE}}\left[\widehat{\vec{\mathbf{X}}}_i^{(n)}\right] = \widehat{\phi}\widehat{\vec{\mathbf{A}}}_i^{1/2}\widehat{\vec{\mathbf{C}}}_i\widehat{\vec{\mathbf{A}}}_i^{1/2} - 2\widehat{\phi}\widehat{\vec{\mathbf{A}}}_i^{1/2}\bar{\widehat{\mathbf{C}}}_i\widehat{\mathbf{A}}_i^{1/2}\widehat{\mathbf{H}}_{ii}^{\top} + \widehat{\vec{\mathbf{D}}}_i\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}\widehat{\vec{\mathbf{D}}}_i^{\top}$$

$$\widehat{\mathbf{H}}_{ij} = \widehat{\vec{\mathbf{D}}}_i\left(\sum_{l=1}^{n}\widehat{\mathbf{D}}_l^{\top}\widehat{\mathbf{V}}_l^{-1}\widehat{\mathbf{D}}_l\right)^{-1}\widehat{\mathbf{D}}_j^{\top}\widehat{\mathbf{V}}_j^{-1}$$

- For detailed notation, see Hudecová and Pešta (2013)

Real data

- Data set from Taylor and Ashe (1983)
- $n = 10$ accident years are available
- The Pearson residuals of the classical GLM suggest that there might be some **small or moderate correlation** of the incremental claims within the same accident year (for the first and second development years, we get -0.22 for the overdispersed Poisson and -0.29 for the gamma model)
- Six competing GEE models **versus** well-known reserving methods taken from Table 1 in England and Verrall (1999)



Prediction errors as % of reserve estimate

Method	CoV [%]	GEE	CoV [%]
Mack (1993)	13	Ind linear	5.1
Poisson GLM	16	Ind quadratic	5.6
Gamma GLM	15	Exch linear	6.9
Verrall (1991)	15	Exch quadratic	7.5
Renshaw/Christofides	16	AR(1) linear	4.8
Zehnwirth	16	AR(1) quadratic	5.5

- Based on Correlation information criterion (**CIC**)
- GEE approach is a distribution free, the estimate of the MSEP does not neglect the bias of prediction, and that the dependencies are allowed within each accident year

Summary

- **GEE modeling technique** as a suitable stochastic method for claims reserving
- GEE approach **enables modeling dependencies** between claim amounts of the development years
- Model **selection criteria** are available for the GEE and, thus, the competing models can be compared directly by one number, which is often a practical advantage
- **Estimate for the mean square error of prediction** for the claims reserves is derived in a non-traditional way—incorporating the sandwich (robust) covariance matrix estimate, not neglecting the bias of prediction, and not ignoring dependencies between claim amounts

Thank you !

Michal.Pesta@mff.cuni.cz