March 20, 2024 CONVENTION A EUROPE

Automatic Segmentation of Insurance Rating Classes via Group Fused Lasso

Shunichi Nomura

Waseda University

Joint work with Atsumori Takahashi (ACCESS CO., LTD.)

Our paper (open access): https://doi.org/10.1515/apjri-2022-0012

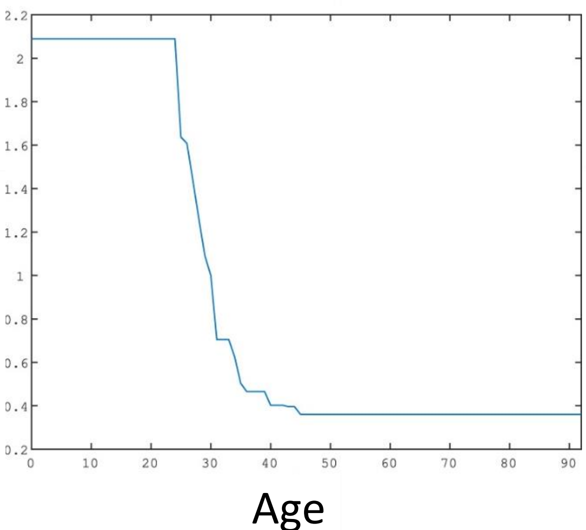
Motivation

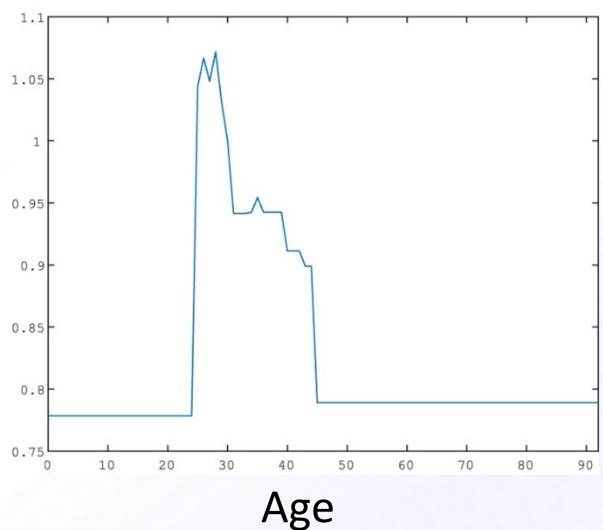
- Premium tariffs have long been used to reference general insurance premiums according to “rating factors”.
- Some rating factors (e.g. insured’s age, region, occupation) have too many categories to determine reliable pure premiums for their individual categories. Complex tariffs with many categories may increase the operational risk of applying incorrect premiums.
- Categories in such risk factors are often integrated into fewer groups which have similar risk levels by actuaries.
- However, finding the optimal grouping of the rating categories is difficult due to the enormous number of grouping combinations.
- Thus, we employ a sparse regularisation technique to find a good grouping of the rating categories efficiently.

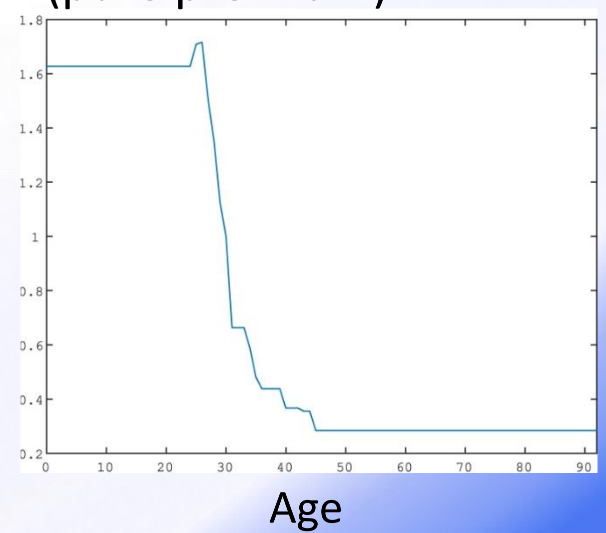
Aim of our study

- Sparse regularisation techniques can integrate categories in rating factors into some groups with similar risk levels (e.g. Accurate GLM by Fujita, Kondo, and Iwasawa, 2020). However, most of these methods are applied to claim frequency and severity, separately.
- We propose sparse regularisation techniques to provide consistent segmentation of rating classes between claim frequency and severity.

Expected claim frequency × Expected claim severity = Expected total claim cost (pure premium)







Settings

- We have claim data of T policies. The t th policy has exposure w_t, number of claims z_t, and mean severity of claims y_t.
- Consider p rating factors whose numbers of categories are $n_1, \dots, n_p$, respectively.
- Let $x_{t1}, \dots, x_{tp}$ denote the categories to which the t th policy belong for the 1st to p th rating factors.
- Our aim is to estimate the expected claim frequency $\mu_t^{(1)}$ and severity $\mu_t^{(2)}$, leading to the expected total claim cost or pure premium $\mu_t^{(1)} \times \mu_t^{(2)}$.

Distribution for claim frequency/severity

- Poisson distribution for claim frequency z_t :

$$f^{(1)}(z_t|\mu_t^{(1)}, w_t) = \frac{\left(w_t\mu_t^{(1)}\right)^{z_t}}{z_t!} e^{-w_t\mu_t^{(1)}}$$

  - z_t : number of the claims from the t-th policy.
  - w_t : total exposure of the t-th policy.
  - $\mu_t^{(1)}$: expected claim frequency per exposure of the t-th policy, i.e. $\mathrm{E}(z_t) = w_t\mu_t^{(1)}$.

- Gamma distribution for claim severity y_t given the number of claims z_t :

$$f^{(2)}(y_t|\mu_t^{(2)}, z_t, \phi) = \frac{1}{y_t\Gamma\left(\frac{z_t}{\phi}\right)}\left(\frac{y_t z_t}{\mu_t^{(2)}\phi}\right)^{\frac{z_t}{\phi}} \exp\left(-\frac{y_t z_t}{\mu_t^{(2)}\phi}\right)$$

  - y_t : mean severity of the claims from the t-th policy.
  - $\mu_t^{(2)}$: expected claim severity of the t-th policy, i.e. $\mathrm{E}(y_t) = \mu_t^{(2)}$.
  - ϕ : dispersion parameter, i.e. $\mathrm{Var}(y_t) \propto \phi$.

Generalised linear models for claim frequency/severity

- Log-link GLMs (multiplicative models) for claim frequency/severity:

$$\log \mu_t^{(j)} = \beta_0^{(j)} + \beta_{1x_{t1}}^{(j)} + \beta_{2x_{t2}}^{(j)} + \cdots + \beta_{px_{tp}}^{(j)} \qquad (j = 1,2)$$

$$\Leftrightarrow \mu_t^{(j)} = \exp\left(\beta_0^{(j)}\right) \times \exp\left(\beta_{1x_{t1}}^{(j)}\right) \times \exp\left(\beta_{2x_{t2}}^{(j)}\right) \cdots \times \exp\left(\beta_{px_{tp}}^{(j)}\right) \ (j = 1,2)$$

  - $\beta_0^{(j)}$: intercept for claim frequency ($j = 1$) or severity ($j = 2$).
  - x_{ti} : category of the i-th factor to which the t-th policy belongs.
  - $\beta_{ik}^{(j)}$: regression coefficient for the k-th category of the i-th factor (see below).

Relative difference between the categories in the *i*-th factor

	Categories in the i-th factor				
	1st category	⋯	k th category	⋯	n_i th category
Expected claim frequency	$\exp\left(\beta_{i1}^{(1)}\right)$	⋯	$\exp\left(\beta_{ik}^{(1)}\right)$	⋯	$\exp\left(\beta_{in_i}^{(1)}\right)$
Expected claim severity	$\exp\left(\beta_{i1}^{(2)}\right)$	⋯	$\exp\left(\beta_{ik}^{(2)}\right)$	⋯	$\exp\left(\beta_{in_i}^{(2)}\right)$

Fused lasso

- Log-likelihood function:

$$l(\boldsymbol{\beta},\phi)=\sum_{t=1}^{T}\left\{\log f^{(1)}(y_t|\mu_t^{(1)}(\boldsymbol{\beta}),z_t,\phi)+\log f^{(2)}(z_t|\mu_t^{(2)}(\boldsymbol{\beta}),w_t)\right\}$$

- Fused lasso is a sparse regularisation method to shrink absolute difference in regression coefficients between adjacent categories:

$$(\widehat{\boldsymbol{\beta}},\hat{\phi})=\underset{\boldsymbol{\beta},\phi}{\operatorname{argmin}}-l(\boldsymbol{\beta},\phi)+\underbrace{\kappa\sum_{i=1}^{p}\sum_{\substack{u,v\text{ are}\\ \text{adjacent}}}\sum_{j=1}^{2}\left|\beta_{iu}^{(j)}-\beta_{iv}^{(j)}\right|}_{\text{regularisation terms}}$$

- Non-significant differences in adjacent coefficients will be shrunk to exact zero, which means those coefficients have the same expected claim frequency or severity.

Fused lasso

- Adjacent categories with the same expected claim frequency and severity, simultaneously (i.e. $\beta_{iu}^{(1)} = \beta_{iv}^{(1)}$ and $\beta_{iu}^{(2)} = \beta_{iv}^{(2)}$), can be regarded as one category group to which the same premium should be applied.
- However, due to the separate penalty for frequency and severity, the fused lasso integrates adjacent categories inconsistently between frequency and severity, leading to too many groups for insurance rating.

Penalty term for the difference in expected claim frequency

$$\left|\beta_{iu}^{(1)} - \beta_{iv}^{(1)}\right|$$

	Categories in the i-th factor			
	⋯	u th category	v th category	⋯
Expected claim frequency	⋯	$\exp\left(\beta_{iu}^{(1)}\right)$	$\exp\left(\beta_{iv}^{(1)}\right)$	⋯
Expected claim severity	⋯	$\exp\left(\beta_{iu}^{(2)}\right)$	$\exp\left(\beta_{iv}^{(2)}\right)$	⋯

Penalty term for the difference in expected claim severity

$$\left|\beta_{iu}^{(2)} - \beta_{iv}^{(2)}\right|$$

Group fused lasso

- Group Fused lasso shrinks differences in expected claim frequency and severity simultaneously to zero between adjacent categories.

$$(\hat{\boldsymbol{\beta}}, \hat{\phi}) = \underset{\boldsymbol{\beta},\phi}{\operatorname{argmin}} -l(\boldsymbol{\beta}, \phi) + \kappa \sum_{i=1}^{p} \sum_{\substack{u,v \text{ are} \\ \text{adjacent}}} \left\| \begin{pmatrix} \beta_{iu}^{(1)} \\ \beta_{iu}^{(2)} \end{pmatrix} - \begin{pmatrix} \beta_{iv}^{(1)} \\ \beta_{iv}^{(2)} \end{pmatrix} \right\|_2$$

Relative difference between the categories in the i-th factor

	Categories in the i-th factor			
	⋯	u th category	v th category	⋯
Expected claim frequency	⋯	$\exp\left(\beta_{iu}^{(1)}\right)$	$\exp\left(\beta_{iv}^{(1)}\right)$	⋯
Expected claim severity	⋯	$\exp\left(\beta_{iu}^{(2)}\right)$	$\exp\left(\beta_{iv}^{(2)}\right)$	⋯

Penalty term for the difference between adjacent categories

$$\left\| \begin{pmatrix} \beta_{iu}^{(1)} \\ \beta_{iu}^{(2)} \end{pmatrix} - \begin{pmatrix} \beta_{iv}^{(1)} \\ \beta_{iv}^{(2)} \end{pmatrix} \right\|_2 = \sqrt{\left(\beta_{iu}^{(1)} - \beta_{iv}^{(1)}\right)^2 + \left(\beta_{iu}^{(2)} - \beta_{iv}^{(2)}\right)^2}$$

Optimisation problem

- Gradient-based optimisation methods cannot be applied for (Group) fused lasso because the objective functions cannot always be differentiable (e.g. when $\beta_{iu}^{(1)} = \beta_{iv}^{(1)}$ and $\beta_{iu}^{(2)} = \beta_{iv}^{(2)}$).
- The optimal solution $(\hat{\boldsymbol{\beta}}, \hat{\phi})$ can be obtained by solving the following equivalent optimisation problem with dummy variables.

$$\min_{\boldsymbol{\beta},\phi,\xi} \quad -l(\boldsymbol{\beta}, \phi) + \kappa \sum_{i=1}^{p} \sum_{\substack{u,v \text{ are} \\ \text{adjacent}}} \left\| \boldsymbol{\xi}_{i,(u,v)} \right\|_2$$

$$\text{s.t.} \quad \boldsymbol{\xi}_{i,(u,v)} = \boldsymbol{\beta}_{iu} - \boldsymbol{\beta}_{iv} = \begin{pmatrix} \beta_{iu}^{(1)} \\ \beta_{iu}^{(2)} \end{pmatrix} - \begin{pmatrix} \beta_{iv}^{(1)} \\ \beta_{iv}^{(2)} \end{pmatrix}$$

for $i = 1, \ldots, p$, and all adjacent pairs (u, v)

Algorithm

- The constrained optimisation problem can be solved by iterating the following updating equations, which is called the alternating direction method of multipliers (ADMM) :

(Transpose of) Lagrange multiplier

Constant to adjust convergence speed

$$(\boldsymbol{\beta}^{new}, \phi^{new}) = \underset{\boldsymbol{\beta},\phi}{\operatorname{argmin}} -l(\boldsymbol{\beta}, \phi) + \kappa \sum_{i=1}^{p} \sum_{\substack{u,v \text{ are} \\ \text{adjacent}}} \left\{ -\boldsymbol{\lambda}'_{i,(u,v)} \left(\boldsymbol{\beta}_{iu} - \boldsymbol{\beta}_{iv} - \boldsymbol{\xi}_{i,(u,v)} \right) + \frac{\rho}{2} \left\| \boldsymbol{\beta}_{iu} - \boldsymbol{\beta}_{iv} - \boldsymbol{\xi}_{i,(u,v)} \right\|_2^2 \right\}$$

Gradient-based optimisation methods are available

$$\boldsymbol{\xi}_{i,(u,v)}^{new} = \begin{cases} \left(1 - \frac{\kappa}{\left\|\boldsymbol{\eta}_{i,(u,v)}\right\|_2}\right) \frac{\boldsymbol{\eta}_{i,(u,v)}}{\rho} & \text{if } \left\|\boldsymbol{\eta}_{i,(u,v)}\right\|_2 > \kappa \\ 0 & \text{otherwise} \end{cases}$$

where $\boldsymbol{\eta}_{i,(u,v)} = \rho(\boldsymbol{\beta}_{iu}^{new} - \boldsymbol{\beta}_{iv}^{new}) - \boldsymbol{\lambda}_{i,(u,v)}$

for $i = 1, \dots, p$, and all adjacent pairs (u, v)

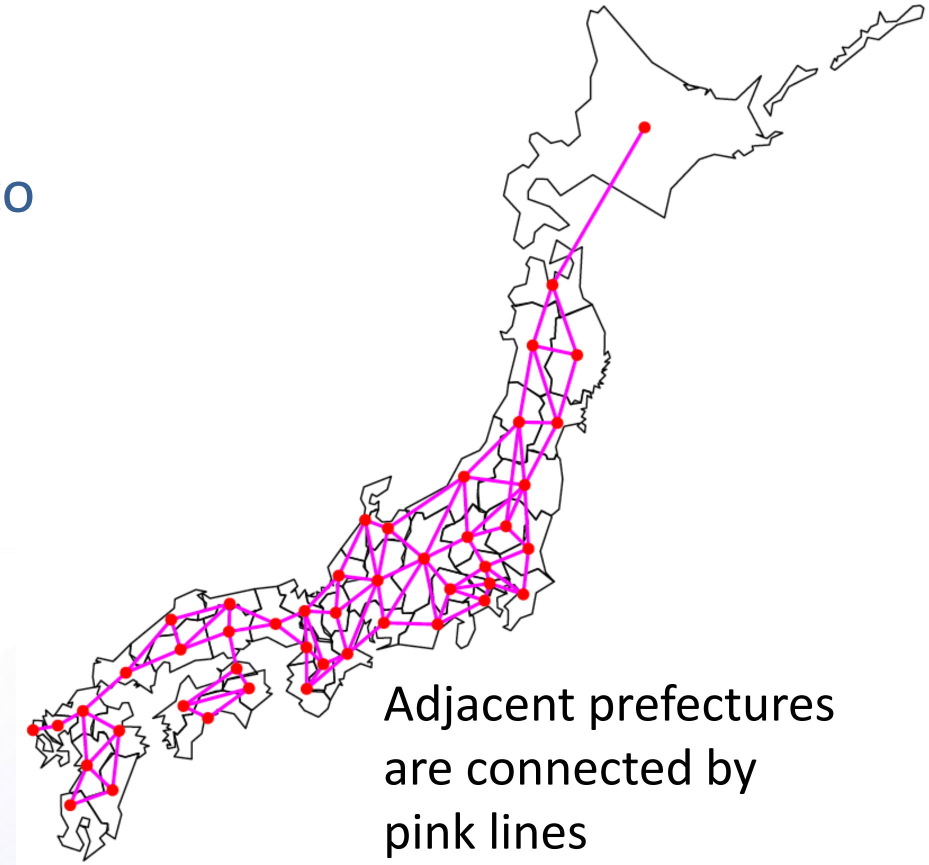
$$\boldsymbol{\lambda}_{i,(u,v)}^{new} = \boldsymbol{\lambda}_{i,(u,v)} - \rho\left(\boldsymbol{\beta}_{iu}^{new} - \boldsymbol{\beta}_{iv}^{new} - \boldsymbol{\xi}_{i,(u,v)}^{new}\right)$$

for $i = 1, \dots, p$, and all adjacent pairs (u, v)

Example: compulsory automobile insurance in Japan

- Nomura (2017) applied the proposed model to claim data of microbuses from compulsory automobile liability insurance aggregated by prefectures in Japan and estimated expected claim frequency and severity.

- We defined the prefectures whose roads connects directly to each other as the adjacent prefectures (see the right map).

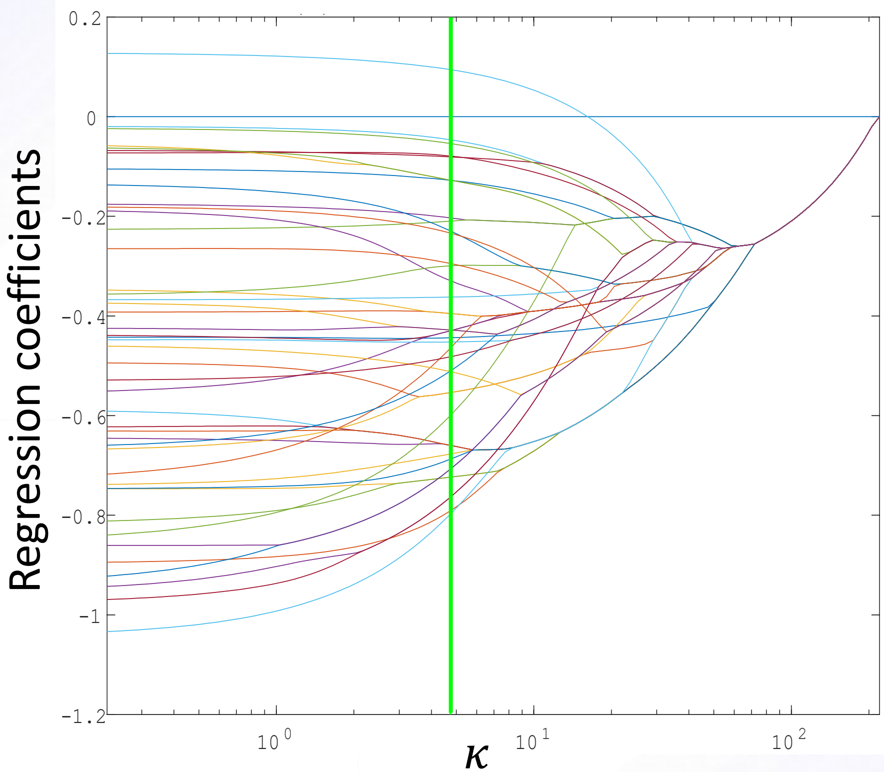


Adjacent prefectures are connected by pink lines

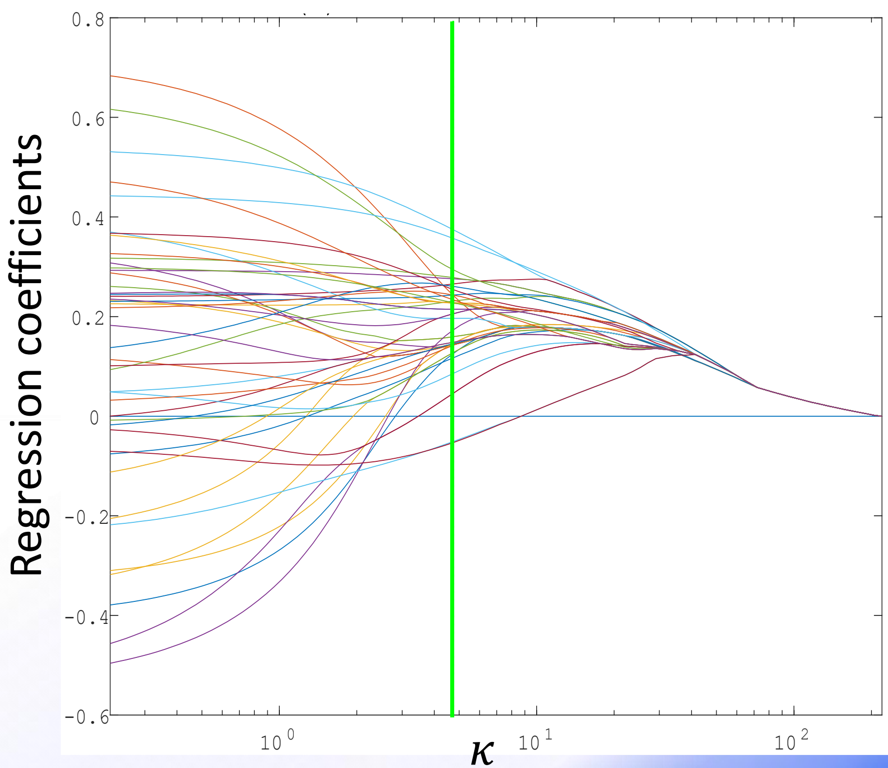
Regression coefficients w.r.t. penalty weight κ

- Regression coefficients (relative difference among the prefectures) are integrated as the penalty weight κ increases.
- The optimal κ (green vertical line) selected by 5-fold cross validation integrates 46 prefectures into 35 groups.

Expected claim frequency



Expected claim severity



Group fused lasso with ordinal constraints

- In practice, some ordinal constraints are often imposed to insurance premiums such as monotonic constraints on bonus–malus classes in automobile insurance.
- We enhanced to the group fused lasso by adding such constraints , whose optimal solution $(\widehat{\boldsymbol{\beta}}, \hat{\phi})$ can be obtained by solving the following optimisation problem.

$$\min_{\boldsymbol{\beta},\phi,\xi} \quad -l(\boldsymbol{\beta},\phi) + \kappa \sum_{i=1}^{p} \sum_{\substack{u,v \text{ are} \\ \text{adjacent}}} \left\| \begin{pmatrix} \beta_{iu}^{(1)} \\ \beta_{iu}^{(2)} \end{pmatrix} - \begin{pmatrix} \beta_{iv}^{(1)} \\ \beta_{iv}^{(2)} \end{pmatrix} \right\|_2$$

$$\text{s. t.} \quad \begin{pmatrix} \beta_{iu}^{(1)} \\ \beta_{iu}^{(2)} \end{pmatrix} \geq \begin{pmatrix} \beta_{iv}^{(1)} \\ \beta_{iv}^{(2)} \end{pmatrix}$$

for sets of factor i and adjacent pairs (u, v) under ordinal constraints

Component-wise inequality

Algorithm modified for ordinal constraints

- The constrained problem can be solved by the previous algorithm where the update formula of $\boldsymbol{\xi}_{i,(u,v)}$ for constrained set of factor i and adjacent pairs (u,v) are modified to:

$$\boldsymbol{\xi}_{i,(u,v)}^{new} = \begin{cases} \left(1 - \dfrac{\kappa}{\left\|\boldsymbol{\eta}_{i,(u,v)}^{+}\right\|_2}\right) \dfrac{\boldsymbol{\eta}_{i,(u,v)}^{+}}{\rho} & \text{if } \left\|\boldsymbol{\eta}_{i,(u,v)}^{+}\right\|_2 > \kappa \\ 0 & \text{otherwise} \end{cases}$$

where $\boldsymbol{\eta}_{i,(u,v)} = \rho(\boldsymbol{\beta}_{iu}^{new} - \boldsymbol{\beta}_{iv}^{new}) - \boldsymbol{\lambda}_{i,(u,v)}$ and

$$\boldsymbol{\eta}_{i,(u,v)}^{+} = \begin{pmatrix} \max\left\{\eta_{i,(u,v)}^{(1)}, 0\right\} \\ \max\left\{\eta_{i,(u,v)}^{(2)}, 0\right\} \end{pmatrix}.$$

Component-wise positive part of $\boldsymbol{\eta}_{i,(u,v)}$

Example: Swedish motorcycle insurance

- We applied the modified method to claim data of the Swedish motorcycle insurance in Ohlsson and Johansson (2010).
- We used the following variables as rating factors :

$i = 1$: The age of insured's owner having $n_1 = 100$ categories (0-99 years old).

$i = 2$: The EV-rate class taking values from 1 to $n_2 = 7$, and classified by so called the EV ratio (=engine output (kW) ÷ (vehicle weight (kg) + 75) × 100).

EV Class	1	2	3	4	5	6	7
EV ratio	0-5	6-8	9-12	13-15	16-19	20-24	25-

$i = 3$: The city-size class taking values from 1 to $n_3 = 7$, and classified by the scale and location of cities and towns (see details in the next slide).

$i = 4$: The Bonus–malus class taking values from 1 to $n_4 = 7$. The class starts from 1 for a new driver, increases by 1 for each claim-free year, and decreases by 2 for each claim.

City-size classes

Class 1: central and semi-central parts of Sweden’s three largest cities,

Class 2: suburbs plus middle-sized cities,

Class 3: lesser towns (except those in 5 or 7),

Class 4: small towns and countryside (except 5–7),

Class 5: Northern towns,

Class 6: Northern countryside,

Class 7: Gotland (Sweden’s largest island).

Example: Swedish motorcycle insurance

- We applied the proposed model with the following constraints:

$$\min_{\boldsymbol{\beta},\phi,\xi} \quad -l(\boldsymbol{\beta},\phi) + \kappa \sum_{i=1}^{4} \sum_{k=2}^{n_i} \|\boldsymbol{\beta}_{ik} - \boldsymbol{\beta}_{i\,k-1}\|_2$$

$$\text{s.t.} \quad \boldsymbol{\beta}_{21} \le \boldsymbol{\beta}_{22} \le \boldsymbol{\beta}_{23} \le \boldsymbol{\beta}_{24} \le \boldsymbol{\beta}_{25} \le \boldsymbol{\beta}_{26} \le \boldsymbol{\beta}_{27},$$

monotonicity for EV-rate classes

$$\boldsymbol{\beta}_{41} \ge \boldsymbol{\beta}_{42} \ge \boldsymbol{\beta}_{43} \ge \boldsymbol{\beta}_{44} \ge \boldsymbol{\beta}_{45} \ge \boldsymbol{\beta}_{46} \ge \boldsymbol{\beta}_{47}.$$

monotonicity for bonus-malus classes

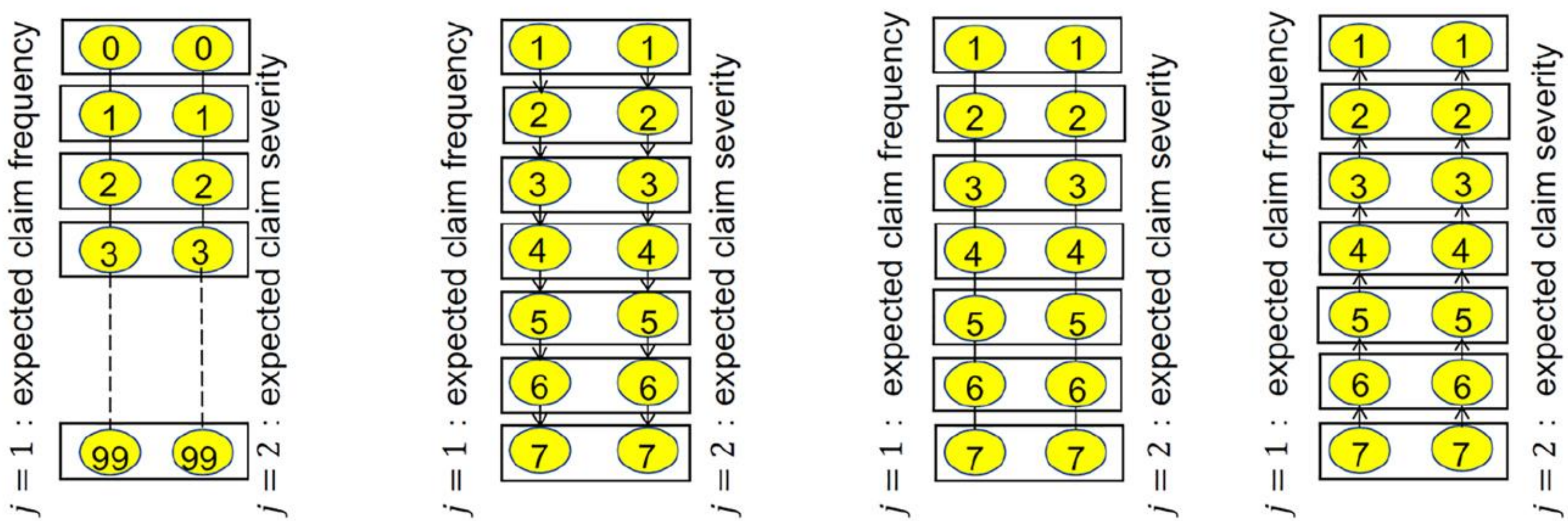
Adjacency graphs

Factor : $i = 1$ — Owner's age : $k = 0,\ldots,99$

$i = 2$ — EV-rate class : $k = 1,\ldots,7$

$i = 3$ — City-size class : $k = 1,\ldots,7$

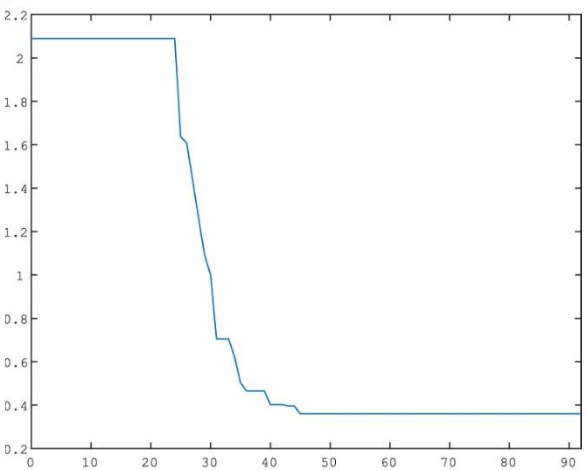
$i = 4$ — Bonus-malus : $k = 1,\ldots,7$



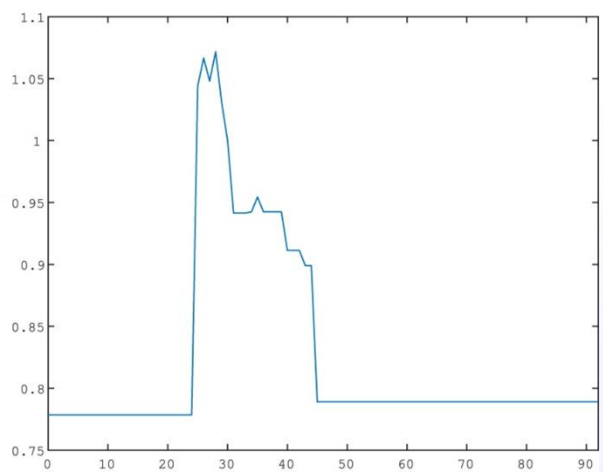
Example: Swedish motorcycle insurance

- Owner’s age are integrated into 14 groups; young ages 0–24, widely ranged older ages 45–99, and eight of them around 30 consist of single ages.
- Expected claim frequency decreases monotonically along owner’s age whereas expected claim severity has its peak in late 20s.

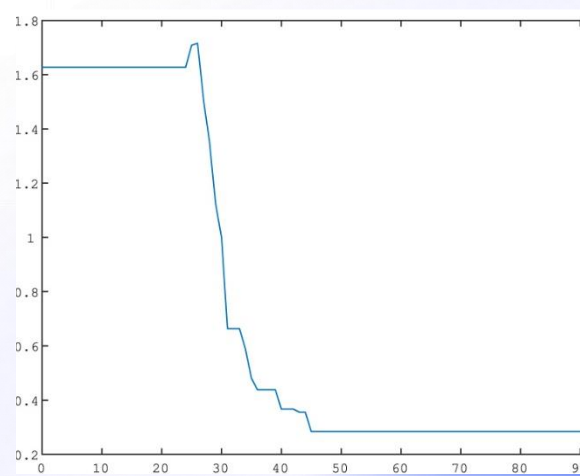
Relative expected claim frequency × Relative expected claim severity = Relative expected total claim cost (pure premium)



Age



Age



Age

Example: Swedish motorcycle insurance

- Though EV-rate classes are integrated into three groups, they have no difference in expected claim severity (ordinal constrains may have worked).
- City-size classes 4-7 are united to one small risk group.
- Bonus-malus classes are all united to one group, i.e. no significant difference in total claim cost is found among them.

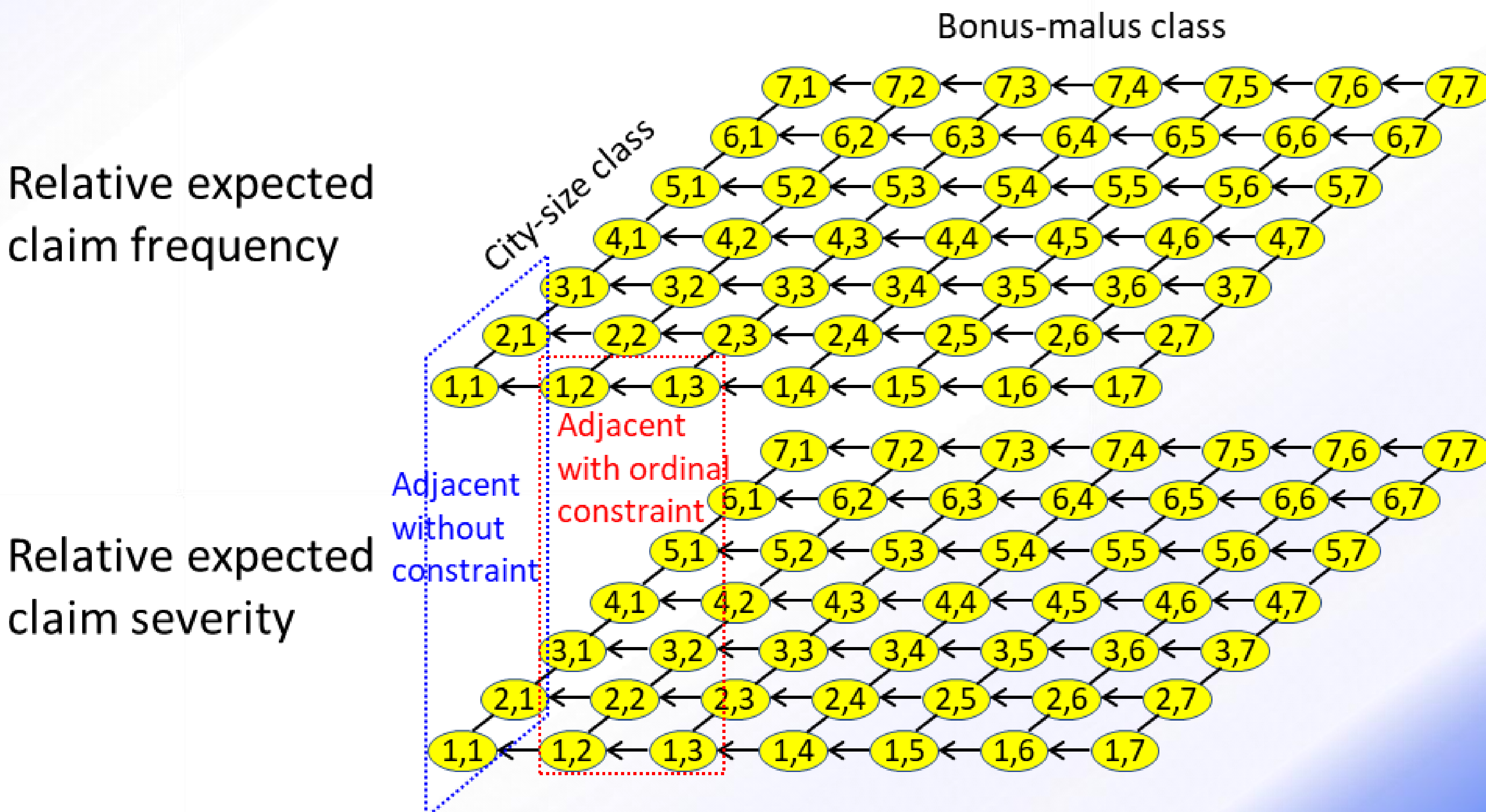
EV-rate class	Relative expected claim frequency	Relative expected claim severity	Relative expected total claim cost
1–4	1.000	1.000	1.000
5	1.313	1.000	1.313
6,7	2.023	1.000	2.023

City-size class	Relative expected claim frequency	Relative expected claim severity	Relative expected total claim cost
1	4.151	1.552	6.443
2	2.539	1.493	3.791
3	1.522	1.147	1.747
4–7	1.000	1.000	1.000

Bonus–malus class	Relative expected claim frequency	Relative expected claim severity	Relative expected total claim cost
1–7	1.000	1.000	1.000

Example: Swedish motorcycle insurance

- Finally, we consider interaction between city-size classes and bonus-malus classes with a two-dimensional adjacency graph.
- Ordinal constraints are imposed to adjacent bonus-malus classes whereas no constraints to adjacent city-size classes.



Example: Swedish motorcycle insurance

In contrast to the previous analysis, differences in expected total claim cost are found between bonus-malus classes (especially in city-size class 1).

In city-size class 3 and 5-7, expected claim severity drops at the highest bonus-malus class.

Relative expected claim frequency

		Bonus–malus class						
		1	**2**	**3**	**4**	**5**	**6**	**7**
City-size class	1	5.741	4.657	4.578	4.578	3.917	3.917	3.917
	2	2.618	2.618	2.618	2.618	2.618	2.618	2.618
	3	1.589	1.589	1.589	1.589	1.589	1.589	1.589
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	0.980
	6	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	7	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Relative expected claim severity

		Bonus–malus class						
		1	**2**	**3**	**4**	**5**	**6**	**7**
City-size class	1	1.747	1.747	1.717	1.717	1.557	1.344	1.344
	2	1.556	1.556	1.556	1.556	1.556	1.556	1.556
	3	1.410	1.410	1.410	1.205	1.205	1.205	0.789
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	0.744
	6	1.008	1.000	1.000	1.000	1.000	1.000	0.650
	7	1.000	1.000	1.000	1.000	1.000	1.000	0.650

Relative expected total claim cost

		Bonus–malus class						
		1	**2**	**3**	**4**	**5**	**6**	**7**
City-size class	1	10.027	8.133	7.860	7.860	6.100	5.264	5.264
	2	4.072	4.072	4.072	4.072	4.072	4.072	4.072
	3	2.241	2.241	2.241	1.915	1.915	1.915	1.254
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	0.729
	6	1.008	1.000	1.000	1.000	1.000	1.000	0.650
	7	1.000	1.000	1.000	1.000	1.000	1.000	0.650

Summary

- We proposed a sparse regularisation method for insurance ratemaking to automatically clustering rating classes of risk factors into groups with the same risk levels.
- We further enhanced the proposed method by adding ordinal constraints on categories in risk factors.
- Our method is available for interaction of risk factors as well.
- We demonstrated our method in the analysis of claim data from Japanese compulsory automobile liability insurance and Swedish motorcycle insurance.

Thank you for your attention!