Effective *a Posteriori* Ratemaking with Large Insurance Portfolios via Surrogate Modeling

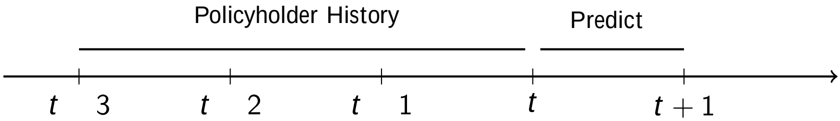
Sebastián Calcetero*, Andrei Badescu, Sheldon Lin
*sebastian.calcetero@mail.utoronto.ca
https://actsci.utstat.utoronto.ca/

Department of Statistical Sciences
University of Toronto
Toronto, Ontario

IDS Conference - June 15, 2023

The Bayesian framework for a posteriori ratemaking

- A insurance company wants to use the attributes of the policyholder and the observed risk experience for those who renew their contracts to set the renewal price of a contract.



- The upgrading process is achieved by means of a Bayesian model:

$$\mathbf{Y_n} = Y_1, \ldots, Y_n | \Theta = \theta \quad {}_{iid} \, f(y|\theta)$$

$$\Theta \quad P(\theta)$$

- Use the Bayes rule to obtain a posterior predictive distribution and so the upgraded premiums:

$$\Pi(Y_{n+1}|Y_1, \ldots, Y_n)$$

The problem with the Bayesian framework

- In reality, insurance data is complex (multi-modal, heavy tailed, and from multiple correlated coverages. Data-driven models are required.
- General Bayesian models will not provide an analytical expression for the credibility premiums. The premiums have to be obtained via some numerical scheme, e.g. MCMC.
- The scheme must be applied at the individual level, thus it is computationally expensive and prohibitive for large portfolios!
- The rate-making becomes a black-box as there are no closed expressions linking the original premium and the claim history

Our Proposed Approach: Surrogate modeling

- We use surrogate modeling to reduce the computational burden for the calculation of credibility premium via:

$$\hat{G}(\mathbf{Y}_n, n, \mathbf{x}) \quad \Pi(Y_{n+1} j \mathbf{Y}_n)$$

  where $\hat{G}$ is fitted with a reduced number of policyholders.
- A surrogate model is a simplified approximation of the output of the numerical scheme while being less computationally intensive.
- It provides an analytical expression for premium calculation, so interpretations and sensitivity analysis, become an easy task.
- Once fitted, the ratemaking becomes tractable for large portfolios (direct evaluation), and the ratemaking transparent (closed forms).

How to find $\hat{G}(\mathbf{Y}_n, n, \mathbf{x}) \quad \Pi(Y_{n+1}|\mathbf{Y}_n)$?

There is a "recipe" for surrogate model, but **the devil is in the details ....**

1. Select a sub-portfolio of representative policyholders to use for fitting. **We use a population sampling method.**
2. Compute credibility premiums using numerical schemes only on such sub-portfolio. **We use importance sampling simulation.**
3. Estimate a flexible surrogate function $\hat{G}$ using a interpolation method. **We introduce a tailor-made summary statistic (dimension reduction tool), the Credibility Index, to handle the curse of dimensionality**.
4. Assess the accuracy of the fitted function and use it to extrapolate to the rest of the portfolio to obtain the premiums.

Real example in auto insurance

- European insurance company with coverage of physical damage (PD), $Y^{(1)}$, and body injury (BI), $Y^{(2)}$, with 184,484 clients over 10 years.
- Consider the exponential premium principle with surcharge of 5%,

$$\Pi(Y_{n+1} j \mathbf{Y_n}) = \frac{1}{0.05} \log E \left(e^{0.05(Y_{n+1}^{(1)} + Y_{n+1}^{(2)})} j \mathbf{Y_n} \right).$$

- The underlying pricing model is a bivariate negative binomial GLMM with a shared random effect that has an inverse Gaussian prior.

$$Y_j = \begin{pmatrix} Y_j^{(1)} \\ Y_j^{(2)} \end{pmatrix} \quad {}_{iid}\, f(y j \Theta, X\beta) = \begin{pmatrix} \mathrm{NegBinom}(\mu^{(1)}\Theta, r^{(1)}) \\ \mathrm{NegBinom}(\mu^{(2)}\Theta, r^{(2)}) \end{pmatrix}$$

$$\log \mu^{(d)} = \beta_0 + \beta_1^{(d)}\mathrm{CarAge} + \beta_2^{(d)}\mathrm{PolicyholderAge} + \beta_3^{(d)}\mathrm{EnginePower} + \mathrm{Others}$$

$$\Theta \quad P(\theta) = \mathrm{InvGauss}(1, \sigma^2)$$

Real example in auto insurance: The setup

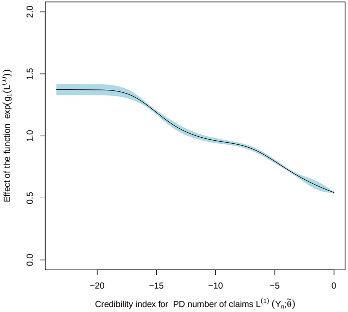
- Only 5% of the policies are selected using the cube method of population sampling,
- Premiums are determined via a importance sampling scheme with 50,000 simulations.
- We use an additive structure for the surrogate function:

$$\hat{G}(L(\mathbf{Y}_n)) = \Pi(Y_{n+1}) \exp\Big(g_1(L^{(1)}(\mathbf{Y}_n)) + g_2(L^{(2)}(\mathbf{Y}_n))\Big)$$

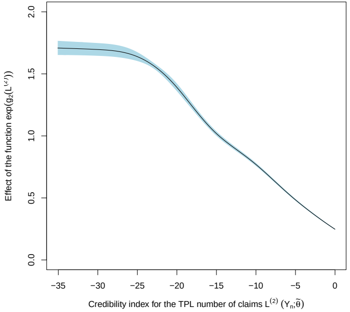
- Where $L^{(j)}(\mathbf{Y}_n)$ is the credibility index, a risk score of the claim history of a policyholder.
- The functions g_1 and g_2 are fitted using a B-Splines representation via the generalized additive model implementation. The response is the simulated premiums, and the inputs are the current premiums and the claim experience.

Real example in auto insurance: Fitted surrogate function

$$\Pi(Y_{n+1} j \mathbf{Y_n}) \quad \Pi(Y_{n+1}) \quad \exp\Big(g_1(L^{(1)}(\mathbf{Y}_n))\Big) \quad \exp\Big(g_2(L^{(2)}(\mathbf{Y}_n))\Big)$$



PD

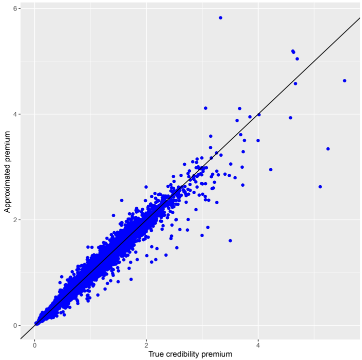


BI

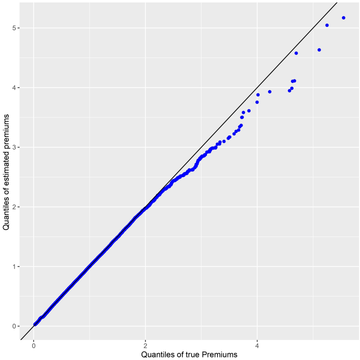
The lower the credibility index, the larger the credibility premium. BI claims can impact the premium more than PD claims.

Real example in auto insurance: Goodness of fit

The coefficient of determination (R2) of fitted premiums vs true premiums, out of the sample is $R^2 = 99\%$



Dispersion plot



Q-Q plot

Figure: Comparison of premiums from the surrogate vs via the simulation.

Real example in auto insurance: Computational cost

Process	Total Portfolio	Sample
Sampling	-	32.94
Simulation for the premium	919,008.00	50,976.00
Fitting surrogate function	-	2,480.22
Extrapolation	-	3.39
Total Time	919,008.00 ($\approx$ 255 h)	53,492.55 ($\approx$ 15 h)

Table: Comparison of the CPU Time (in seconds)

The surrogate model takes around 17 times less time than the whole process!

The overall methodology via surrogate modeling provides desirable results in an inexpensive fashion.

Thank you for listening!

Paper available on Arxiv & SSRN

Calcetero Vanegas, S. and Badescu, A. L. and Lin, X. S. E ective a Posteriori Ratemaking with Large Insurance Portfolios via Surrogate Modeling (November 12, 2022). Available at SSRN: https://ssrn.com/abstract=4275353

Visit our research group website

actsci.utstat.utoronto.ca