A bivariate mixed Poisson claim count regression model with varying dispersion and shape

Insurance Data Science Conference, City University of London
16/06/2023

Joint work of George Tzougas[1] and Despoina Makariou[2]
[1]Department of Actuarial Mathematics and Statistics, Heriot-Watt University
[2]Institute of Insurance Economics, University of St. Gallen

Outline

1. Introduction
2. Motivation
3. Methodology
4. Contributions
5. Numerical illustration
6. Research extensions

Introduction

Multivariate count data in insurance

- Underwriting of several insurance lines
- Dependence structures for claim counts and size
- Multivariate data in actuarial practice

Motivation

Motivation

- Focus on non-life insurance
- Model various types of claims counts
- Capture dependence structures
- Account for overdispersion and correlation
- Flexibility

Methodology

Methodology

- Develop a MVPGIG regression model with varying dispersion and shape
- Flexible general class of models
- Mixture of independent Poisson distributions
- Single random effect variable distributed according to GIG
- MVPGIG parameters are modelled in terms of covariates
- Focus on the bivariate case

Model framework and assumptions

- Non-life insurance policy
- Insured j, where $j = 1, \ldots, n$
- Multi-peril claim frequencies $\mathcal{K}_{i,j}$, for $i = 1, \ldots, m$ types of coverage
- Assume that given the random variables $Z_j > 0$, $\mathcal{K}_{i,j} = k_{i,j} | Z_j$ per claim type i are distributed according to a Poisson distribution with probability mass function (pmf) given by

$$\mathbb{P}\left(\mathcal{K}_{i,j} = k_{i,j} | z_j\right) = \frac{\exp[-\mu_{i,j} z_j](\mu_{i,j} z_j)^{k_{i,j}}}{k_{i,j}!},$$

for $k_{i,j} = 0, 1, 2, 3, \ldots$, where $\mu_{i,j} > 0$, with mean and variance $\mathbb{E}(\mathcal{K}_{i,j} | Z_j = z_j) = \mu_{i,j} z_j$ and $\mathrm{Var}(\mathcal{K}_{i,j} | Z_j = z_j) = \mu_{i,j} z_j$.

Model framework and assumptions (continued)

- Z_j are random variables from a GIG distribution with probability density function (pdf) given by

$$g(z_j;\sigma_j,\nu_j) = \frac{c_j^{\nu_j}}{2K_{\nu_j}\left(\frac{1}{\sigma_j}\right)} z_j^{\nu_j-1} \exp\left[-\frac{1}{2\sigma_j}\left(c_j z_j + \frac{1}{c_j z_j}\right)\right],$$

for $\sigma_j > 0$ and $-\infty < \nu_j < \infty$, where

$c_j = \frac{K_{\nu_j+1}(\sigma_j^{-1})}{K_{\nu_j}(\sigma_j^{-1})}$ and $K_{\nu_j}(\omega)$ is the modified Bessel function of the third kind of order ν_j and argument ω. This parameterization ensures that the model is identifiable since $\mathbb{E}(Z_j) = 1$.

Here, $\mu_{i,j}$ and $\sigma_j \rightarrow$ observable risk characteristics and $z_j \rightarrow$ non-observable risk characteristics

The MVPGIG model

The unconditional distribution of $\mathcal{K}_{i,j}$ will be a MVPGIG distribution with joint probability mass function (jpmf) given by

$$\mathbb{P}(\mathcal{K}_{i,j} = k_{i,j}) = \frac{\prod_{i=1}^{m} \mu_{i,j}^{k_{i,j}}}{\prod_{i=1}^{m} k_{i,j}!} \frac{c_j^{\nu_j}}{2K_{\nu_j}(\frac{1}{\sigma_j})} \left[(2\sum_{i=1}^{m} \mu_{i,j} + \frac{c_j}{\sigma_j})c\sigma_j\right]^{-\frac{\sum_{i=1}^{m} k_{i,j}+\nu_j}{2}}$$

$$2K_{\sum_{i=1}^{m} k_{i,j}+\nu_j} \left[\sqrt{\frac{1}{c\sigma_j}(2\sum_{i=1}^{m} \mu_{i,j} + \frac{c}{\sigma_j})}\right].$$

The BPGIG model

- Consider the bivariate (m=2) case of MVPGIG
- Assume that the mean, dispersion and shape parameters of the BPGIG are modelled as functions of explanatory variables with parametric linear functional forms

$$\mu_{1,j} = \exp({x_{1,j}}^T \beta_1), \qquad \sigma_j = \exp({x_{3,j}}^T \beta_3),$$
$$\mu_{2,j} = \exp({x_{2,j}}^T \beta_2), \qquad \nu_j = {x_{4,j}}^T \beta_4.$$

- The covariance between $\mathcal{K}_{1,j}$ and $\mathcal{K}_{2,j}$ is given by

$$\text{Cov}\left(\mathcal{K}_{1,j}, \mathcal{K}_{2,j}\right) = \mu_{1,j}\mu_{2,j}\left(c_j^{-2} + \frac{2(\nu_j + 1)}{c_j}\sigma_j - 1\right).$$

Contributions

Contributions

- Risk factors into the mean, dispersion, and shape parameters
- All MVPGIG parameters are modelled in terms of covariates $\rightarrow$ very flexible and versatile family of models
- Modelling of the positive correlation between the two claims types in a stylised manner
- Novel Expectation Maximization algorithm for maximum likelihood estimation of the model parameters $\rightarrow$ additional complexities

Numerical illustration

Numerical illustration

- Sample of MTPL claim frequency data
- Modelling of bodily injury and property damage claims with their associated claim counts using the BPGIG
- Comparison with BNB, BPIG based on Global Deviance, AIC and SBC
- Calculation of the A Posteriori Premiums

The explanatory variables and their description

Variables	Categories		
	C1	C2	C3
City population	$\leq$ 1,000,000	1,000,001-2,000,000	$\geq$ 2,000,001
Number of years that the policyholder has been registered with the insurance company	$<$ 5 years	$>$ 5 years	-
Horsepower of the vehicle	0-1400 cc	1400-1800 cc	$\geq$ 1800 cc

Descriptive statistics for the two responses

K_1		K_2	
statistic	value	statistic	value
Minimum	0	Minimum	0
Median	0	Median	0
Mean	0.0954	Mean	0.0618
Variance	0.1375	Variance	0.0644
Maximum	4	Maximum	3
Kendall's τ: 0.1760			
Spearman's ρ: 0.1777			

Key results - In sample performance

Model	*df*	Global Deviance	AIC	SBC
BNB	18	4388	4424	4542
BPIG	18	4249	4285	4403
BPGIG	19	4223	4261	4386

Key results - Out of sample performance

- Split the data into training and test data at the ratio of 9:1.
- 4149 data points for training and 1037 data points for testing.
- The BPGIG model outperforms the two competing bivariate mixed Poisson models.

Model	Deviance
BNB	490.80
BPIG	475.25
BPGIG	472.35

Research extensions

Research extensions

- Claim size instead of claim count
- Two random effects variables instead of one
  $\rightarrow$ Insurance bundling
  $\rightarrow$ Relax assumption of positive correlation
- Deep neural networks and hybrid models

Thank you!