Operations Research Letters 42 (2014) 273–277

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Copula-based dependence between frequency and class in car insurance with excess zeros

Xiaobing Zhao [a,*], Xian Zhou [b,*]

[a] School of Mathematics and Statistics, Zhejiang University of Finance and Economics, Hangzhou, Zhejiang Province, China

[b] Department of Applied Finance and Actuarial Studies, Macquarie University, NSW, Australia

ARTICLE INFO

Article history:
Received 23 September 2013
Received in revised form
3 April 2014
Accepted 3 April 2014
Available online 13 April 2014

Keywords:
Car insurance
Bonus-malus
Dependent
Zero-inflated
Copula function

ABSTRACT

A bonus-malus system calculates the premiums for car insurance based on the previous claim experience (class). In this paper, we propose a model that allows dependence between the claim frequency and the class occupied by the insured using a copula function. It also takes into account zero-excess phenomenon. The maximum likelihood method is employed to estimate the model parameters. A small simulation is performed to illustrate the proposed model and method.

1. Introduction

In many European and Asian countries, as well as in North-American states or provinces, insurers use experience rating to relate premium amounts, which is calculated by multiplying the reference (base) premium with a percentage attached to the level (class) occupied by the policyholder, to individual past claim experience in automobile insurance. Such systems penalize insured drivers responsible for one or more accidents by premium surcharges (or malus) and reward claim-free policyholders by awarding them discounts (or bonuses), which are called bonus-malus systems (BMS). One of the main tasks of the insurance company is to design a proper tariff structure or an optimal BMS that will fairly distribute the burden of claims among policyholders.

In most commercial bonus-malus systems, the knowledge of the current level and the number of claims during the current period suffices to determine the level in the next period. Thus the future (the level for year $t+1$) depends only on the present (the level for year t and the number of accidents reported during year t) and not on the past. This is closely related to the memory-less property of the Markov chains. In the analysis of a BMS, it is commonly assumed that the claim frequency of an individual policyholder remains unaltered. This assumption ensures constant transition probabilities and enables analysts to model the system using a time-homogeneous Markov chain. This route has been studied for a long time by many authors, such as Norberg [18], Borgan et al. [2], Gilde and Sundt [11] and Lemaire [15]. In reality, however, it is known that the claim frequency of a driver may fluctuate from time to time for various reasons, so that the transition probability changes with it as well. An intuitive approach is to interpret past claims as a factor that changes the mean of the distribution [10,22].

In the existing literature, the claims of the policyholders between contracts are generally assumed to be independent, but correlation between the claims of the same individual is permitted. A type of dependence is serial correlation between previous and current claim numbers. A nice review on this can be found in [4], in which the time dependence of claims during successive periods are discussed separately by the random effects model with a Poisson or negative binomial distribution, the Poisson regression model with past experience modeled as covariates, the INAR(1) model, the common shock model and the copula model.

The future claim number, however, can also be impacted by the current class occupied by the policyholder, which can modify accident proneness. For instance, a policyholder in a high-risk class may modify the perception of danger behind the wheel and lower the risk to report another claim in the future in order to reduce the penalty in the premium. In another case, policyholders in a low-risk class may rarely get accidents due to a great deal

* Corresponding authors.
E-mail addresses: maxbzhao@hotmail.com (X. Zhao), xian.zhou@mq.edu.au (X. Zhou).

http://dx.doi.org/10.1016/j.orl.2014.04.001
0167-6377/

of experience and proficient skill. Insurance rating schemes also provide incentives to careful driving and thus induce negative contagion. Nevertheless, policyholders who reported claims in the past generally tend to produce claims in the future, leading to positive contagion, which is referred to as *epidemiology* [6]. Thus it is of interest to consider the dependence between the future claim number and the current class occupied by the insured.

The dependence between the claim frequency and the class occupied by the insured has been mentioned by some authors. Pitrebois et al. [20] assumed that the distribution of the number of claims is related to the risk classes possessed in multi-event BM scales. Denuit [8] raised the dependence between the bonus class and annual expected claim frequency, but noted that it does not mean bonus classes to be functions of the frequency parameter, but only that their distribution depends on the class. However, they did not discuss extensively the dependence mentioned above.

In this paper, we propose to model the dependence between the current bonus class occupied by the policyholder and the claim numbers in this insurance period by a bivariate copula function. This model can also accommodate excess zeros of the claims as well. The ordinary maximum likelihood is applied to estimate the model parameters, including the copula parameter.

In Section 2 next, we specify our proposed model. Parameter estimation is discussed in Section 3. Section 4 reports some simulation results to assess the proposed model, followed by concluding remarks in Section 5.

2. Model specification

2.1. Marginal distribution

During the observation periods, there are n policies in the portfolio that is divided into s bonus classes, each is observed during T_i periods. For each observation, the responses consist of bonus class $C_{i,t}$ and $N_{i,t}$, which represent the bonus class occupied by insured i at the beginning of period t and the number of claims reported by insured i during tth period, respectively, $i = 1, 2, \dots, n$; $t = 1, 2, \dots, T_i$. The premium scale is $\mathbf{b} = (b_1, b_2, \dots, b_s)^\top$, where b_k is the premium level in class $C_k (k = 1, 2, \dots, s)$. Let $d_{i,t}$ denote the length of this period (the risk exposure). As mentioned by Purcaru and Denuit [21], $d_{i,t} = 1$ is usual, but there are a variety of situations where $d_{i,t} \neq 1$. For example, it may happen that a new period of observation starts when some policy characteristics are modified, such as a change of postcode when the insured moved, or the insured buying a new car, etc.

For each policy, there is certain *priori* information at the beginning of each insurance period. As common in actuarial studies, the starting point for modeling the number of reported claim $N_{i,t}$ for insured i during period t is a Poisson distribution:

$$\Pr(N_{i,t} = n_{i,t}) = \frac{\lambda_{i,t}^{n_{i,t}}}{n_{i,t}!} \exp(-\lambda_{i,t}) \tag{2.1}$$

with mean claim frequency $\lambda_{i,t} = d_{i,t} \exp(x_{i,t}^\top \alpha + \beta)$, where β is the intercept and α is a regression parameter for explanatory variables (covariates) $x_{i,t}$.

As an example, consider the experience rating systems defined in [8]. There are six bonus-malus classes, of which level 5 is the starting class. A higher class number indicates a higher premium. For policyholder i, the class $C_{i,t+1}$ in year $t+1$ is a function of class $C_{i,t}$ in year t. The discount per claim-free year is one class: if no claims have been reported by policyholder i, then he moves one class down: $C_{i,t+1} = \max\{1, C_{i,t} - 1\}$. The penalty per claim is by two levels: if a number of claims, $n_{i,t} > 0$, have been reported during year t, then the policyholder moves $2n_{i,t}$ classes up: $C_{i,t+1} = \min\{6, C_{i,t} + 2n_{i,t}\}$. Therefore we can compute, for example, $\Pr(C_{i,2} = 6 | C_{i,1} = 5) = \Pr(N_{i,1} > 1 | C_{i,1} = 5)$, $\Pr(C_{i,2} = 4 | C_{i,1} = 5) = \Pr(N_{i,1} = 0 | C_{i,1} = 5)$, and so on. Similarly we can calculate the probability for $C_{i,k+1}$ given $C_{i,k}$ at any period t, for $k = 1, 2, \dots, s-1$ and $t = 1, 2, \dots, T_i$. Hence the transition rule, denoted by $T = (t_{mn})_{6\times 6}$, is given by

$$T = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} \{0\} & - & \{1\} & - & \{2\} & \geq \{3\} \\ \{0\} & - & - & \{1\} & - & \geq \{2\} \\ - & \{0\} & - & - & \{1\} & \geq \{2\} \\ - & - & \{0\} & - & - & \geq \{1\} \\ - & - & - & \{0\} & - & \geq \{1\} \\ - & - & - & - & \{0\} & \geq \{1\} \end{pmatrix} \end{array},$$

where t_{mn} denotes the number of claims if the policy is transferred from class m to class n $(m, n = 1, \dots, 6)$.

Then the distribution of class $C_{i,t}$, which represents the class occupied by insured i during period t, is defined as follows:

$$\Pr(C_{i,t} = 1, C_{i,t} = 2, \dots, C_{i,t} = 6) = (0, \dots, 0, 1, 0) \cdot \prod_{j=1}^{t-1} P_{i,j}, \tag{2.2}$$

for $t = 2, \dots, T_i$, where

$$P_{i,j} = \begin{pmatrix} p_{i,j,0} & 0 & p_{i,j,1} & 0 & p_{i,j,2} & q_{i,j,3} \\ p_{i,j,0} & 0 & 0 & p_{i,j,1} & 0 & q_{i,j,2} \\ 0 & p_{i,j,0} & 0 & 0 & p_{i,j,1} & q_{i,j,2} \\ 0 & 0 & p_{i,j,0} & 0 & 0 & q_{i,j,1} \\ 0 & 0 & 0 & p_{i,j,0} & 0 & q_{i,j,1} \\ 0 & 0 & 0 & 0 & p_{i,j,0} & q_{i,j,1} \end{pmatrix} \tag{2.3}$$

with $p_{i,j,k} = \Pr(N_{i,j} = k)$ for $k = 0, 1, 2$ and $q_{i,j,k} = 1 - \Pr(N_{i,j} = k-1)$ for $k = 1, 2, 3$.

2.2. Modeling dependence

Since bonus classes $C_{i,t}$ have the memoryless property, and hence follow a Markov chain, the contribution of $(C_{i,1}, C_{i,2}, \dots, C_{i,T_i})$ for insured i to the likelihood is

$$\Pr(C_{i,1}, C_{i,2}, \dots, C_{i,T_i}) = \prod_{t=1}^{T_i} \Pr(C_{i,t+1} = c_{i,t+1} | C_{i,t} = c_{i,t}) \Pr(C_{i,1} = c_{i,1}), \tag{2.4}$$

where

$$\Pr(C_{i,t+1} = c_{i,t+1} | C_{i,t} = c_{i,t}) = \Pr(N_{i,t} = n_{i,t} | C_{i,t} = c_{i,t}). \tag{2.5}$$

In the literature for bonus-malus system, the probability in the right hand side of (2.5) is often assumed to be independent of the current class occupied by insured i, that is, $\Pr(N_{i,t} = n_{i,t} | C_{i,t} = c_{i,t}) = \Pr(N_{i,t} = n_{i,t})$. This is however hardly justifiable in practice. In this paper, taking into account the effect of the current bonus class on possible claims in this period, we propose a new model to fit the dependence between $N_{i,t}$ and $C_{i,t}$. Pitrebois et al. [20] also considered the distribution of the number of claims related to the risk class in multi-event bonus-malus scales, but did not further investigate it extensively.

In our setting, $N_{i,t}$ and $C_{i,t}$ are assumed to be dependent. The joint probability of $N_{i,t}$ and $C_{i,t}$ can be expressed by their joint cumulative distribution function (cdf) as follows:

$$\begin{aligned} \Pr(N_{i,t} = n_{i,t}, C_{i,t} = c_{i,t}) &= \Pr(N_{i,t} \leq n_{i,t}, C_{i,t} \leq c_{i,t}) - \Pr(N_{i,t} \leq n_{i,t} - 1, C_{i,t} \leq c_{i,t}) \\ &\quad - \Pr(N_{i,t} \leq n_{i,t}, C_{i,t} \leq c_{i,t} - 1) \\ &\quad + \Pr(N_{i,t} \leq n_{i,t} - 1, C_{i,t} \leq c_{i,t} - 1). \end{aligned} \tag{2.6}$$