Insurance: Mathematics and Economics 56 (2014) 28–37

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

Conditional least squares and copulae in claims reserving for a single line of business

Michal Pešta [a,*], Ostap Okhrin [b]

[a] Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, CZ-18675 Prague, Czech Republic

[b] Humboldt University of Berlin, School of Business and Economics, Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. – Center for Applied Statistics and Economics, Spandauer Strasse 1, D-10178 Berlin, Germany

HIGHLIGHTS

- A time series model for the conditional mean and variance of the claims.
- The time series innovations for the consecutive claims not to be independent.
- Conditional least squares to estimate model parameters and consistency proved.
- A copula approach for modeling the dependence structure.
- A semiparametric bootstrap to estimate the distribution of the reserves.

ARTICLE INFO

Article history:
Received June 2013
Received in revised form
January 2014
Accepted 25 February 2014

JEL classification:
C13
C32
C33
C53
G22

MSC:
60G10
60G25
60J20
62H10
62H20
62J02
62P05

Subject Category and Insurance Branch Category:
IM10
IM11
IM20
IM40

Keywords:
Claims reserving
Reserve distribution

ABSTRACT

One of the main goals in non-life insurance is to estimate the claims reserve distribution. A generalized time series model, that allows for modeling the conditional mean and variance of the claim amounts, is proposed for the claims development. On contrary to the classical stochastic reserving techniques, the number of model parameters does not depend on the number of development periods, which leads to a more precise forecasting.

Moreover, the time series innovations for the consecutive claims are not considered to be independent anymore. Conditional least squares are used to estimate model parameters and consistency of these estimates is proved. The copula approach is used for modeling the dependence structure, which improves the precision of the reserve distribution estimate as well.

Real data examples are provided as an illustration of the potential benefits of the presented approach.

* Corresponding author. Tel.: +420 221 913 400; fax: +420 222 323 316.
E-mail addresses: pesta@karlin.mff.cuni.cz, michal.pesta@mff.cuni.cz (M. Pešta), ostap.okhrin@wiwi.hu-berlin.de (O. Okhrin).

http://dx.doi.org/10.1016/j.insmatheco.2014.02.007
0167-6687/

Dependency modeling
Copula
Conditional least squares

1. Introduction

Claims reserving is one of the most important issues in general insurance. A large number of various methods has been invented, see England and Verrall (2002) or Wüthrich and Merz (2008) for an overview.

The main aim of this paper is to deal with serious issues in contemporary reserving techniques, which are quite often set aside, but cause serious problems in the actuarial estimation and prediction. Such pitfalls are assumptions of independent claims, independent stochastic errors (or residuals) in the corresponding claims reserving model, and considering large number of parameters often depending on the number of observations.

The majority of the classical approaches are based on the assumption that the claim amounts in different years are independent. However, this assumption can sometimes be unrealistic or at least questionable. It has been pointed out that methods, enabling *modeling the dependencies*, are needed, cf. Antonio and Beirlant (2007) or Hudecová and Pešta (2013). Mentioned papers suggest the generalized linear mixed models (GLMMs) or generalized estimating equations (GEEs) to handle the possible dependence among the incremental claims in successive development years. These approaches extend the classical GLM and are frequently used in panel (longitudinal) data analyses. In this paper, we present another possible attitude, namely the *conditional mean–variance* model with a *copula* function.

On the one hand, time series model by Buchwalder et al. (2006) nicely and simply allow to model conditional mean and variance of the claim amounts. On the other hand, that model possesses two disadvantages, which are common for a huge majority of the reserving methods: infinite number of parameters (i.e., depending on the number of observation) and independent errors. Generally, large number of parameters decreases the precision of estimation, because of not sufficient amount of data for the estimation. Furthermore, the classical statistical inference is not valid anymore when the number of parameters depends on the number of observation. To overcome such difficulties, we consider a generalized time series model with *a finite number of parameters not depending on the number of development periods* and, additionally, the *model errors* belonging to the same accident period are *not independent*.

Moreover, all the currently used bootstrap methods in claims reserving require independent residuals in order to estimate the distribution of the reserve and, consequently, calculate some distributional quantities, e.g., VaR at 99.5%. Assumption of independent residuals can be quite unrealistic in the claims reserving setup. Hence, an alternative and more suitable resampling method needs to be proposed in order to sensibly estimate the reserves distribution.

Copulae have already been utilized in the claims reserving to model dependencies between different lines of business, e.g., Shi and Frees (2011). On the contrary, it has to be emphasized that in our approach, only one line of business is taken into account. Copulae are therefore used to model dependencies *within claims* corresponding to that *single line of business*. For sure, our approach can be generalized for several lines of business in the way that a second level of dependence (for instance, modeled again by the copulae) is introduced between the claim amounts from different lines of business.

The structure of this paper is as follows: The claims reserving notation is summarized in Section 2. In Section 3, a generalized time series model for the conditional mean and variance of claim amounts is introduced. Section 4 elaborates copula approach for dependence modeling within the generalized time series model for claims triangles. Section 5 covers estimation techniques for the parameters of the generalized time series model and copula as well. Consistency of the estimates is derived. Section 6 concerns prediction of the actuarial claims reserves and, furthermore, estimation of their distribution. Finally, all the presented methods and approaches are applied on real data in Section 7 in order to show their performance and outstanding benefits. Proofs of the theorems are listed in the Appendix.

2. Claims reserving notation

We introduce the classical claims reserving notation and terminology. Outstanding loss liabilities are structured in so-called claims development triangles, see Table 1. Let us denote $Y_{i,j}$ all the claim amounts up to development year $j \in \{1, \dots, n\}$ with accident year $i \in \{1, \dots, n\}$. Therefore, $Y_{i,j}$ stands for the *cumulative claims* in accident year i after j development periods. The current year is n, which corresponds to the most recent accident year and development period as well. Hence, $Y_{i,j}$ is a random variable of which we have an observation if $i + j < n + 1$ (a run-off triangle). Thus, our data history consists of right-angled isosceles triangle $\{Y_{i,j}\}$, where $i = 1, \dots, n$ and $j = 1, \dots, n + 1 - i$. The diagonal elements $Y_{i,j}$, where $i + j$ is constant, correspond to the claim amounts in *accounting year* $i + j$.

The aim is to predict the ultimate claims amount $Y_{i,n}$ and the outstanding *claims reserve* $R_i^{(n)} = Y_{i,n} - Y_{i,n+1-i}$ for all $i = 2, \dots, n$. Additional to that, *estimation of the whole distribution of the reserves* is needed in order to provide important distributional quantities for the *Solvency II purposes*, e.g., quantiles for the value at risk calculation.

3. Conditional mean and variance model

Run-off triangles are comprised of observations which are ordered in time. It is therefore natural to suspect the observations to be dependent. On one hand, the most natural approach is to assume that the observations of a common accident year are dependent. On the other hand, observations of different accident years are supposed to be independent. This assumption is similar to those of the Mack's chain ladder model, cf. Mack (1993).

$\mathcal{F}_{i,j}$ denotes the information set generated by trapezoid $\{Y_{k,l} : l \leq j, k \leq i + 1 - j\}$, i.e., $\mathcal{F}_{i,j} = \sigma(Y_{k,l} : l \leq j, k \leq i + 1 - j)$ is a filtration corresponding to the smallest σ-algebra containing historical claims with at most j development periods paid in accounting period i or earlier. This notation allows for a zero or even negative index in filtration despite the fact that the claims corresponding to a zero or negative development of accident years are not observed.

Let us define a *nonlinear generalized semiparametric regression* type of model. It can be considered as a generalization of the model proposed by Buchwalder et al. (2006). The first level of generalization is in the mean and variance structure, which was inspired by Patton (2012). The second level of generalization regarding the dependence structure will be introduced in Section 4.