Insurance: Mathematics and Economics 71 (2016) 63–78

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

Stochastic loss reserving with dependence: A flexible multivariate Tweedie approach

Benjamin Avanzi [a,b], Greg Taylor [a], Phuong Anh Vu [a,b,*], Bernard Wong [a]

[a] School of Risk and Actuarial Studies, UNSW Australia Business School, UNSW Sydney NSW 2052, Australia

[b] Département de Mathématiques et de Statistique, Université de Montréal, Montréal QC H3T 1J4, Canada

ARTICLE INFO

Article history:
Received March 2016
Received in revised form
July 2016
Accepted 21 August 2016
Available online 29 August 2016

JEL classification:
G22

Keywords:
Stochastic loss reserving
Dependence
Multivariate Tweedie distribution
Common shock
Bayesian estimation

ABSTRACT

Stochastic loss reserving with dependence has received increased attention in the last decade. A number of parametric multivariate approaches have been developed to capture dependence between lines of business within an insurer's portfolio. Motivated by the richness of the Tweedie family of distributions, we propose a multivariate Tweedie approach to capture cell-wise dependence in loss reserving. This approach provides a transparent introduction of dependence through a common shock structure. In addition, it also has a number of ideal properties, including marginal flexibility, transparency, and tractability including moments that can be obtained in closed form. Theoretical results are illustrated using both simulated data sets and a real data set from a property-casualty insurer in the US.

1. Introduction

In a non-life insurance company, there is typically a delay between the occurrence of an insured event and the actual payment of its related claims. This delay can be driven by various reasons, including delays in reporting claims, investigation of claim validity, and legal proceedings. In order to sustain financial stability and meet regulatory requirements, it is essential for insurers to have sufficient reserves to respond to the outstanding claims. These loss reserves are one of the largest liabilities on the balance sheet of an insurer (see for example, Alai and Wüthrich, 2009; Shi, 2014; Zhang et al., 2012), which further emphasises the importance of having an adequate estimation of outstanding liabilities. This is further discussed and motivated in the excellent general reference by Wüthrich and Merz (2008).

When a company has more than one line of business, one approach for loss reserving is to simply add the reserves of each individual business line. However, this approach is only accurate in the perfectly positive dependence case, and as a result, an insurer is unable to enjoy diversification benefits (De Jong, 2012; Shi et al., 2012). This subsequently motivates the development of stochastic loss reserving with non-perfect dependence; see for example, Abdallah et al. (2015), Merz and Wüthrich (2009a), Merz et al. (2013) and Zhang et al. (2012).

One of the main streams of the literature on stochastic loss reserving with dependence focuses on cell-wise dependence between business lines (Merz et al., 2013). This refers to the dependence between claims coming from the same accident and development periods in different run-off triangles. Early modelling developments with this type of dependence include multivariate additive models by Hess et al. (2006); Merz and Wüthrich (2009b) and Schmidt (2006). Copulas have also been used in models with cell-wise dependence, for example, Shi and Frees (2011) and Zhang and Dukic (2013). A later stream of the literature has also incorporated an additional source of dependence arising from calendar year effects, see, for example, Bühlmann and Moriconi (2015), De Jong (2012), Salzmann and Wüthrich (2012) and Shi (2014).

Methodologies used in loss reserving models with dependence can be classified into two main groups, (i) parametric models, and (ii) non-parametric models (Shi et al., 2012). Parametric models utilise distributional families, while the later ones do not. The most popular parametric approach in the literature is to use copulas. Shi

* Corresponding author at: School of Risk and Actuarial Studies, UNSW Australia Business School, UNSW Sydney NSW 2052, Australia.

E-mail addresses: b.avanzi@unsw.edu.au (B. Avanzi), gregory.taylor@unsw.edu.au (G. Taylor), p.vu@unsw.edu.au (P.A. Vu), bernard.wong@unsw.edu.au (B. Wong).

http://dx.doi.org/10.1016/j.insmatheco.2016.08.006
0167-6687/

and Frees (2011) proposed a flexible copula model with marginal generalised linear models (GLM). De Jong (2012) used a Gaussian copula approach to capture the dependence between lines. Zhang and Dukic (2013) developed a Bayesian copula framework with flexible marginal modelling for cell-wise dependence. Abdallah et al. (2015) used a hierarchical Archimedean copula structure to capture calendar year dependence between lines. The copula approach has the great benefit of having flexibility because it allows marginal densities and joint dependence to be modelled separately. Besides copulas, there have also been developments utilising a multivariate distribution for a specific marginal density, particularly, log-normal distributions. Examples include the multivariate log-normal frameworks for incremental claims in Shi et al. (2012) and for log-link ratios in Merz et al. (2013).

The search for multivariate modelling approaches drew our attention to a multivariate Tweedie distribution for margins from the Tweedie family of distributions. The Tweedie family is a major subclass of the exponential dispersion family (EDF) consisting of symmetric and non-symmetric, light-tailed and heavy-tailed distributions (Alai et al., 2016; Jørgensen, 1997). This class and its members are frequently used in loss reserving, see for example, Alai and Wüthrich (2009), Boucher and Davidov (2011), England and Verrall (2002), Peters et al. (2009), Renshaw and Verrall (1998), Taylor (2009, 2015), Wüthrich (2003) and Zhang et al. (2012). Furthermore, it is a generalisation of the plain vanilla Chain Ladder Poisson model. A recapitulation of some properties of the univariate Tweedie family of distributions is provided in Appendix A.

In this work, we focus on a multivariate Tweedie distribution developed by Furman and Landsman (2010). The Tweedie family is a broad class of commonly used distributions, and the multivariate distribution is developed through a common shock approach. These allow a multivariate Tweedie framework to have many advantages. Namely, the advantages of our model include:

- Dependence is introduced with the help of a easily identifiable, explicit common shock dependence structure, which is easily generalised to more than two dimensions.
- While all dimensions must have same parameter p, this can be anywhere in $(-\infty, 0] \cap [1, \infty)$ (rather than fixed at 1 for a Poisson dispersion or 2 for a gamma dispersion, for instance). The flexibility of Tweedie dispersions has been established in Alai et al. (2016), Alai and Wüthrich (2009), Furman and Landsman (2010) and Jørgensen (1997). Additionally, zero data points do not present any issue.
- The distributions of (multivariate) margins belong to the same Tweedie family. Furthermore, moments and cumulants that can be obtained analytically, and cumulants of the sum (of reserves) can be given in closed form.

The paper is organised as follows. In Section 2, we introduce a multivariate Tweedie framework. Appropriate parametrisation is considered and appealing properties of this framework are also discussed. An analysis of moments and cumulants from the framework is performed in Section 3. Bayesian inference is developed in Section 4 for model fitting and forecasting. The choices of prior distributions, estimation procedure and Markov Chain Monte Carlo (MCMC) method are also discussed. Section 5 applies the theoretical framework to simulated data sets to assess the accuracy of the estimation procedure. An illustration on real data is then provided in Section 6. Section 7 concludes the paper with some remarks about the model and its applications.

2. A multivariate Tweedie dependence approach

2.1. Notation

We consider a portfolio of N sub-portfolios that represent N lines of business. Notation $X_{i,j}^{(n)}$ represents the total incremental claim that corresponds to the accident period i, $i \in \{1, \ldots, I\}$, and development period j, $j \in \{0, \ldots, J\}$, in the nth business line. These claims are hence made in calendar period $t = i + j$, $t \in \{1, \ldots, I\}$.

Instead of directly modelling the losses, one often standardises incremental claims using exposure variables to obtain consistency across different accident years and lines of business. Common exposure variables, denoted by $\omega_{i,j}^{(n)}$, can, for example, be the number of policies, the amount of premiums written, or the total amount insured. Standardised incremental claims are denoted by

$$Y_{i,j}^{(n)} = \frac{X_{i,j}^{(n)}}{\omega_{i,j}^{(n)}}. \tag{2.1}$$

The set of all claims observations up to the estimation date is represented by

$$Y^U = \left\{ Y_{i,j}^{(n)};\ 1 \le i \le I,\ 0 \le j \le I - i, 1 \le n \le N \right\}, \tag{2.2}$$

and the set of all outstanding claims that are to be predicted is denoted by

$$Y^L = \left\{ Y_{i,j}^{(n)};\ 1 < i \le I,\ I - i + 1 \le j \le J, 1 \le n \le N \right\}. \tag{2.3}$$

The primary goal of a loss reserving model is to use historical claims information Y^U to predict the amount of outstanding claims Y^L.

2.2. Model construction

In this section we will develop a multivariate Tweedie framework for claims from multiple lines of business. Following the literature stream that models cell-wise dependence between lines of business, a multivariate Tweedie distribution is used to capture dependence between cell-wise claims. Standardised cell-wise claims from the ith accident period and jth development period across all lines of business are first collected into a vector

$$\boldsymbol{Y_{i,j}} = \begin{pmatrix} Y_{i,j}^{(1)} \\ Y_{i,j}^{(2)} \\ \vdots \\ Y_{i,j}^{(N)} \end{pmatrix}. \tag{2.4}$$

Each element of the above vector is assumed to be a sum of two components

$$Y_{i,j}^{(n)} = \frac{\theta}{\theta_{i,j}^{(n)}} W_{i,j} + Z_{i,j}^{(n)}, \tag{2.5}$$

where $W_{i,j}$ is referred to as the "common shock" and $Z_{i,j}^{(n)}$ is referred to as the "idiosyncratic effect". These two components are assumed to be independent and have additive Tweedie distributions

$$W_{i,j} \sim \text{Tweedie}_p^*(\theta, \lambda), \tag{2.6}$$

$$Z_{i,j}^{(n)} \sim \text{Tweedie}_p^*(\theta_{i,j}^{(n)}, \lambda_{i,j}^{(n)}). \tag{2.7}$$

The definition of additive Tweedie distributions as well as other properties of the Tweedie family of distributions are provided in Appendix A.