Insurance: Mathematics and Economics 64 (2015) 232–245

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

An individual loss reserving model with independent reporting and settlement[☆]

Jinlong Huang[a], Chunjuan Qiu[a,*], Xianyi Wu[a], Xian Zhou[b]

[a] Department of Statistics and Actuarial Science, East China Normal University, China

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

ARTICLE INFO

Article history:
Received June 2014
Received in revised form
April 2015
Accepted 24 May 2015
Available online 3 June 2015

Keywords:
Individual data
Loss reserving
IBNR reserve
Maximum likelihood estimator
Asymptotic normality

ABSTRACT

The main purpose of this paper is to assess and demonstrate the advantage of claims reserving models based on individual data in forecasting future liabilities over traditional models on aggregate data both theoretically and numerically. The available information consists of the reporting delays, settlement delays and claim payments. The model settings include Poisson distributed frequency of claims produced by each policy, claims payable at the settlement time, and the amount of payment depending only on its settlement delay. While such settings are applicable to certain but not all practical cases, the principal purpose of the paper is to examine the efficiency of individual data against aggregate data. We refer to *loss reserving* as to estimate the projections of the outstanding liabilities on observed information. The efficiency of the individual loss reserving against classical aggregate loss reservings, namely Chain-Ladder (C-L) and Bornhuetter–Ferguson (B–F), is assessed by comparing the asymptotic variances of the errors in estimating the conditional expectation (projection) of the outstanding liability between individual, C-L and B–F reservings. The research shows a significant increase in the accuracy of loss reserving by using individual data compared with aggregate data.

1. Introduction

Loss reserving is to determine the funds required to cover the liabilities produced by incurred losses or claims. It has always been one of the core issues for general insurance operations.

In the literature, many methods, including Chain-Ladder (C-L) and Bornhuetter–Ferguson (B–F), have been introduced and some of them are widely studied and used in practice. Most of them arose first in the deterministic version (see, e.g., Taylor, 2000; England and Verrall, 2002, 2006; Wüthrich and Merz, 2008, for comprehensive introduction), and then stochastic models were introduced in the late 1980s to measure the variability of reserves. Generally, there are two branches of stochastic models: one based on the aggregate data (or macro-level data) and the other on individual data (or micro-level data, cf. Antonio and Plat, 2014).

The first branch makes certain probabilistic assumptions on the data in a run-off triangle, but not much on the mechanisms supporting those assumptions has been discussed. Remarkable works in this branch include Mack (1993), who underlied the most extensively applied C-L algorithm by making certain assumptions on the first two moments of payments in a run-off triangle, Gogol (1993), who assumed log-normal distributions on both accumulated payments and their conditional distribution given the payments in earlier periods, and Verrall (2000), who used negative binomial distributions, to mention just a few. More models, such as credibility, exact Bayesian and generalized linear models, can be found in, e.g., England and Verrall (2002, 2006), Wüthrich and Merz (2008) and the references therein.

Studies involving individual data have a history of more than 20 years. Representative literature includes Arjas (1989), Norberg (1993, 1999), Jewell (1989, 1990), Antonoi et al. (2006), Larsen (2007) and so on. Arjas (1989) and Norberg (1993, 1999) modeled an insurer's liability development caused by every individual claim via Marked Poisson Processes and presented a probabilistic framework for general computation of claims reserve—conditional mean of outstanding liabilities on updated historical information. Jewell (1989, 1990) fitted the numbers of claims and reporting/settlement delays with fully parametric Bayesian models in continuous and discrete time settings. Larsen (2007) made stochastic reserving via decomposing marked Poisson processes into independent sections

☆ This work was supported by Shanghai Philosophy and Social Science foundation under grant no. 2010BJB004, NSFC under grant no. 71371074 and the 111 Project.

* Corresponding author.
E-mail address: cjqiu@stat.ecnu.edu.cn (C. Qiu).

http://dx.doi.org/10.1016/j.insmatheco.2015.05.010
0167-6687/

Table 1
Intuitive explanation for the calculation of $L_{\mathrm{I}} = L_{\mathrm{I}}^{\mathrm{RBNS}} + L_{\mathrm{I}}^{\mathrm{IBNR}}$.

Accident	Reporting year						(1)	(2)	(1)×(2)	Reserves
year	0	1	2	…	…	J_1	$\mathring{N}_t$	$\mathring{\mu}_t$	$\mathring{N}_t\mathring{\mu}_t$	
1						$N^{\mathrm{RBNS}}_{J_2-J_2+1,J_1}$	$\to \mathring{N}_{J_2}$	$\mathring{\mu}_{J_2}$	$\mathring{N}_{J_2}\mathring{\mu}_{J_2}$	
⋮					$\nearrow$	⋰	⋮	⋮	⋮	
⋮				$\nearrow$	$\nearrow$	⋰	⋮	⋮	⋮	
$I-J_2-1$			$N^{\mathrm{RBNS}}_{I-2-J_2+1,2}$	$\nearrow$	$\nearrow$	⋰	⋮	⋮		
$I-J_2$		$N^{\mathrm{RBNS}}_{I-1-J_2+1,1}$	⋰	$\nearrow$	$\nearrow$	⋰	⋮	⋮	⋮	
$I-J_2+1$	$N^{\mathrm{RBNS}}_{I-J_2+1,0}$	⋰	⋰	$\nearrow$	$\nearrow$	⋰	⋮	⋮	⋮	
⋮	⋰	⋰	⋰	$\nearrow$	$\nearrow$	$N^{\mathrm{RBNS}}_{I-J_1-1+1,J_1}$	$\to \mathring{N}_1$	$\mathring{\mu}_1$	$\mathring{N}_1\mathring{\mu}_1$	$\sum = L_{\mathrm{I}}^{\mathrm{RBNS}}$
⋮	⋰	⋰	⋰	$\nearrow$	$\nearrow$	$\lambda_{J_1}e_{J_2+1}$	$\to \mathrm{E}N^{\mathrm{IBNR}}_{J_2+1}$	$\mathring{\mu}_0$	$\mathring{\mu}_0 e_{J_2+1}\sum_{j=J_1}^{J_1}\lambda_j$	
⋮	⋰	⋰	⋰	$\nearrow$	⋰	⋮	⋮	⋮	⋮	
$I-2$	$N^{\mathrm{RBNS}}_{I-3+1,0}$	$N^{\mathrm{RBNS}}_{I-1-2+1,1}$	$N^{\mathrm{RBNS}}_{I-2-1+1,2}$	$\lambda_3 e_{I-2}$	$\to$	$\lambda_{J_1}e_{I-2}$	$\to \mathrm{E}N^{\mathrm{IBNR}}_{I-2}$	$\mathring{\mu}_0$	$\mathring{\mu}_0 e_{I-2}\sum_{j=3}^{J_1}\lambda_j$	
$I-1$	$N^{\mathrm{RBNS}}_{I-2+1,0}$	$N^{\mathrm{RBNS}}_{I-1-1+1,1}$	$\lambda_2 e_{I-1}$	$\to$	$\to$	$\lambda_{J_1}e_{I-1}$	$\to \mathrm{E}N^{\mathrm{IBNR}}_{I-1}$	$\mathring{\mu}_0$	$\mathring{\mu}_0 e_{I-1}\sum_{j=2}^{J_1}\lambda_j$	
I	$N^{\mathrm{RBNS}}_{I-1+1,0}$	$\lambda_1 e_I$	$\lambda_2 e_I$	$\to$	$\to$	$\lambda_{J_1}e_I$	$\to \mathrm{E}N^{\mathrm{IBNR}}_{I}$	$\mathring{\mu}_0$	$\mathring{\mu}_0 e_I\sum_{j=1}^{J_1}\lambda_j$	$\sum = L_{\mathrm{I}}^{\mathrm{IBNR}}$

Note. Here the purposes of both $\nearrow$ and $\to$ are two-fold: One is to indicate the direction along which the summation is computed and the other shows that it acts as an ellipsis sign.

and splitting the corresponding likelihood function into different factors that can be maximized separately. More recently, following the framework of Norberg (1993, 1999), Antonio and Plat (2014) used the framework developed by Arjas (1989) and Norberg (1993) to analyze a set of real insurance data from a European company by first estimating the unknown parameters and then inserting them into the projection of the liabilities.

This paper discusses a discrete time stochastic model for loss reserving in general insurance based on individual data. It is similar to but extending an early work of Norberg (1986) on IBNR liabilities and estimation of unknown distributional parameters on reporting delays to simultaneously treating IBNR and RBNS liabilities (see Section 2 for their meanings). In contrast to the existing literature, mentioned above or not, our main purpose is to demonstrate the advantage of introducing individual level claims data in increasing the accuracy of predicting an insurer's outstanding liabilities. For this theoretical purpose, we consider a simplified version of the model as follows:

(1) The number of policies in each accident year and exposure of each policy are deterministic quantities.
(2) The number of claims produced by each policy is Poisson distributed and claims development processes are mutually independent among different policies.
(3) The insurer pays a claim at its settlement.
(4) The settlement process of each claim is independent of the reporting process.

Often in practice, the settlement process may depend on the reporting process. For example, accidents with possibly larger losses tend to be reported in shorter time and are more likely to be carefully examined by the insurer (resulting in longer delays), leading to a negative correlation between reporting delay and settlement time. Nevertheless, it is also easy to find real insurance cases with independent settlement and reporting processes. Examples include health insurance, in which an insurer can quickly check out whether or not the insured's claimed disease belongs to the list of its insurance liability so that any reporting delay has little effect on the settlement process; and ship or air transport accident insurance, in which the reporting delay is mainly the length to determine the liability of the accident by the relevant authorities of a country, which has generally little relation to the liability of the accident and hence the settlement process. A most recent case is to find out how the airplane of Malaysia Airlines Flight MH370 disappeared. Moreover, the simplified specification of the model provides theoretical advantages in assessing and demonstrating the benefits of using individual data in a more explicit and convincing way.

The research findings and the structure of this paper are summarized as follows:

(1) After describing the individual-level data and making their distributional assumptions, we establish a link between this individual model and the aggregate run-off triangle model deduced from it, and discuss the probabilistic aspects of certain fundamental statistics in Section 2.
(2) We demonstrate loss reserves based on both individual claims data model and its aggregate run-off triangle model via the projections of the outstanding liabilities on the observations in Section 3. In this section, we also design an algorithm (see Table 1) to compute the individual loss reserve and discuss the accuracy of individual loss reserve against aggregate loss reserve by means of mean squared prediction errors.
(3) The distributional parameters are estimated in Section 4 by means of maximum likelihood method and their asymptotic properties are also analyzed there.
(4) Section 5 addresses the individual and aggregate (C-L and B–F) loss reservings by inserting the estimated parameters into the corresponding loss reserves and derives the asymptotic distributions of the differences using those reservings to predict the individual loss reserve. The asymptotic variances of those differences are then compared numerically, which show significant reduction in errors of individual data model versus the aggregate data model. We further report a simulation study in this section, which shows a significant improvement in the accuracy of prediction for finite sample sizes as well.

The final section concludes the paper with some discussions on possible implications of the findings. Due to a large number of theorems, lemmas and corollaries in this paper, their intricate and lengthy proofs are relegated to Appendix to smooth the flow of the text.