Insurance: Mathematics and Economics 53 (2013) 355–365

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

Bayesian analysis of loss reserving using dynamic models with generalized beta distribution

A.X.D. Dong *, J.S.K. Chan

School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia

HIGHLIGHTS

- We model long tail loss reserving data using a generalized beta distribution.
- The models contain state space, mixture and threshold effects for irregular claims.
- They provide more accurate reserves than conventional models.
- They facilitate the classification of loss payment into different risk groups.
- They provide companies greater insight to distinguish claims at an earlier stage.

ARTICLE INFO

Article history:
Received May 2011
Received in revised form
June 2013
Accepted 1 July 2013

Keywords:
Generalized beta distribution
Bayesian analysis
State space model
Threshold model
Mixture distribution
Loss reserve

ABSTRACT

A Bayesian approach is presented in order to model long tail loss reserving data using the generalized beta distribution of the second kind (GB2) with dynamic mean functions and mixture model representation. The proposed GB2 distribution provides a flexible probability density function, which nests various distributions with light and heavy tails, to facilitate accurate loss reserving in insurance applications. Extending the mean functions to include the state space and threshold models provides a dynamic approach to allow for irregular claims behaviors and legislative change which may occur during the claims settlement period. The mixture of GB2 distributions is proposed as a mean of modeling the unobserved heterogeneity which arises from the incidence of very large claims in the loss reserving data. It is shown through both simulation study and forecasting that model parameters are estimated with high accuracy.

1. Introduction

In order to ensure an insurance company's financial security, it is necessary to estimate future claims liabilities. Reserving for the amount of future claims payments involves a large degree of uncertainty, especially for long tail class business where tail behaviors can be largely different. Hence it can be difficult to estimate the loss reserve precisely. Traditionally, conventional distributions such as the lognormal and gamma are used to model severity (Taylor, 2000). In making these distributional assumptions, researchers may underestimate the risk inherited in the long tail which is affected by large claim liabilities because these distributions do not possess flexible tails to describe the features of large claims. Failure to estimate the large claim liabilities adequately can cause financial instability of the company and eventually lead to insolvency. In order to improve modeling accuracy and reliability, sophisticated loss models have been derived with different distributional assumptions.

A wide choice of distributions including the generalized-t (GT) (Chan et al., 2008), Pareto (Zehnwirth, 1994), the Stable family (Paulson and Faris, 1985), the Pearson family (Aiuppa, 1988), the log-gamma and lognormal (Ramlau-Hansen, 1988) and the lognormal and Burr 12 (Cummins et al., 1999), have been used in loss reserving. While the GT distribution is flexible and nests several important families of distributions including the Student-t, uniform, both leptokurtic and platykurtic, and exponential power, it requires log-transformation for the loss data and the resulting log-linear model is more sensitive to low values than large values. As the residuals are negatively skewed when the data contain low claims, Chan et al. (2008) suggested skewed heavy-tailed distributions such as the skewed-t distribution. This paper remedies the drawback of GT distribution and proposes the flexible generalized beta distribution of the second kind (GB2) to model severity distribution. The GB2 family provides flexible tail estimates, and therefore can be used to model heterogeneous loss reserve data without doing repetitive distribution testings. It includes both heavy-tailed and light-tailed severity distributions, such as The

* Corresponding author.
E-mail addresses: alice.dong10@gmail.com, med_alicedong@hotmail.com (A.X.D. Dong), jchan@maths.usyd.edu.au (J.S.K. Chan).

0167-6687/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.insmatheco.2013.07.001

Gamma, Weibull, Pareto, Burr12, lognormal and the Pearson family, hence providing convenient functional forms to model insurance claims (Cummins et al., 1990, 1999, 2007).

In the traditional methodology of estimating severity distribution, loss data is summarized by accident and development periods, thereby adopting a single aggregated loss distribution (Mack, 1991; Chan et al., 2008). However, the question arises as to whether estimation should be carried out on aggregate loss, or individual loss, where individual losses are observed but the interest is focused on the sum. Estimating loss reserve using aggregated data reduces the impact of potential outliers (Chan et al., 2008); however, in the process of aggregating data, individual information and variability are lost. Recently, more analyses are performed on individual claims (Taylor and McGuire, 2004). Cummins et al. (2007) considered subgroups of claims by accident and development years, and he applied separate GB2 distribution with a constant mean to each cell of the runoff triangle. This model allows greater flexibility than fitting a single severity distribution across cells. However, as a separate model is fitted to each cell, the model does not consider covariate effects, and hence the trend movement of claims across development years cannot be obtained in a statistically efficient manner. In order to capture the effect of individual characteristics, accident and development years, a single model with dynamic distributional parameters should be fitted to the entire set of individual data. Indeed, data of these two types, namely aggregated or individual, which possess very different characteristics should be handled separately. This paper proposes two types of loss reserve models applied specifically to aggregated and individual claims.

In analyzing aggregate loss with long tail lines, it is very likely that practical issues arising in reality, such as legislative changes during the long lag period of claims exposure, will affect claim payments. Failure to account for these factors will result in severe bias in loss reserve. In order to cater for these irregular claims behaviors, we extend the mean of the GB2 distribution to adopt some dynamic models including the ANOVA, state space and threshold models. Threshold models, first introduced in Tong (1978), can be considered when the behavior predicted by the model differs in some important ways, for example, a shift in the mean and/or distributional parameters when a switching variable, such as the accident year exceeds certain thresholds, thus offering a dynamic modeling mechanism of risk factors across a threshold. They are characterized by easy interpretation and consequently a large number of applications to real phenomena can be found. Among some of these applications, Li and Lam (1995) and Ling (1999) investigated the asymmetry and the volatility of financial markets; Montgomery et al. (1998), Koop and Potter (1999) used these structures to model unemployment rate. Chan et al. (2008) applied the threshold model for loss reserve to allow for a change in the development year effect.

Many dynamic models can usefully be written in a state-space form. The properties of state-space model can be found in Hamilton (1994). Verrall (1989) proposed a state space representation of the chain ladder linear model, De Jong and Penzer (2004) present the ARIMA model in state space form, and Chan et al. (2008) used the state space form to model loss reserve data. While the ANOVA model enables the effects of accident years and lag years to act separately on the total loss, the state space model is a dynamic modeling approach which allows parameters to evolve in a flexible time-recursive manner, and thereby allowing an interaction effect between accident and development years. It provides a flexible and unified framework to specific and often complicated circumstances (De Jong and Zehnwirth, 1983).

To analyze individual loss data, controlling for unobserved heterogeneity is an important issue. McDonald and Butler (1987) demonstrated how mixture distributions can be applied to model heterogeneous data. We propose the mixture presentation for GB2 distribution, which allows the dynamic mean and shape parameters to vary across subgroups of claims. The estimated group membership for each observation enables classification of claims into different risk groups. Such information is useful for the managers of insurance companies to derive separate strategies for handling different subgroups of claims with varying risk characteristics.

For model implementation, the use of Bayesian ideas and techniques for loss reserving, dates back to 2000 when Verrall (2000) utilized the Bayesian approach to forecast outstanding claims payments in the lower runoff triangle. The benefit of using Bayesian procedure, lies in the adoption of available prior information and the provision of a complete predictive distribution for the required reserves (de Alba, 2002). Different Bayesian loss reserve models have been proposed for different types of claims data. Zhang et al. (2012) proposed a Bayesian nonlinear hierarchical model with growth curves to model the loss development process, using data from individual companies forming various cohorts of claims. This model allows pooling of information from multiple companies to perform cross-company analyses. Ntzoufras and Dellaportas (2002) investigated various models for outstanding claims problems using a Bayesian approach via Markov chain Monte Carlo (MCMC) sampling strategy and showed that the computational flexibility of a Bayesian approach facilitated the implementation of complex models, such as the state space and threshold models. We adopt the Bayesian approach and propose alternative hierarchical forms of the models through the scales mixtures representation of the GB2 distribution. This approach substantially simplifies the Gibbs sampler without a heavy computational cost.

Although GB2 distribution provides flexible tails for modeling loss data, its usage is still very limited, particularly in analyzing aggregate loss under the Bayesian framework. To enable accurate loss reserving, the objectives of this paper are two-fold: to derive unique modeling strategies to analyze two types of loss data, the aggregate loss in the runoff triangle and the individual loss, which possess very different features and characteristics, using the flexible GB2 distribution with dynamic mean functions and to implement the proposed models using the Bayesian approach.

The rest of this paper is organized as follows: In Section 2, we introduce the GB2 distribution, with its properties and its relationship with other distributions. Section 3 describes the Bayesian approach and how it is incorporated in our models. We present our empirical study, with different modeling strategies to address the practical issues arising in aggregated and individual loss reserving data, in Sections 4 and 5. Section 6 provides some final remarks and the conclusion from this study.

2. The generalized beta distribution

Loss data often exhibits heavy-tailed behavior, particularly for long tail business class. The generalized beta distribution of the second kind (GB2) has attractive features for modeling loss reserve data, as it nests a number of important distributions as its special cases. The GB2 distribution has four parameters, which allows it to be expressed in various flexible densities. The density function is specified as follows:

$$f(y; a, b, p, q) = \frac{|a| y^{ap-1}}{b^{ap} B(p, q)(1 + (y/b)^a)^{p+q}}, \tag{1}$$

for $y > 0$ where b is a scale parameter and a, p and q are the shape parameters such that b, p and $q > 0$ and $a \neq 0$. The beta function $B(p, q)$ is defined by:

$$B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)},$$

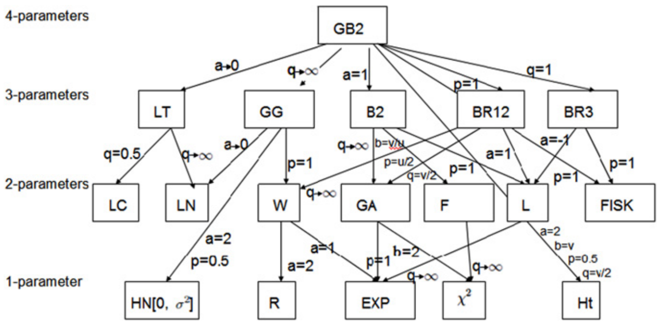

Fig. 1. GB2 distribution tree.

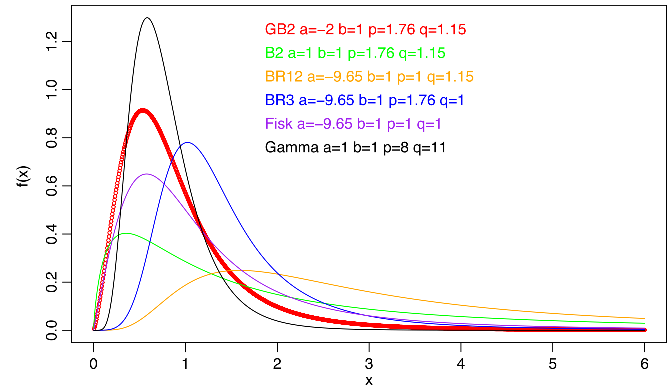

Fig. 2(a). Probability density function across subgroups of distributions in the GB2 family.

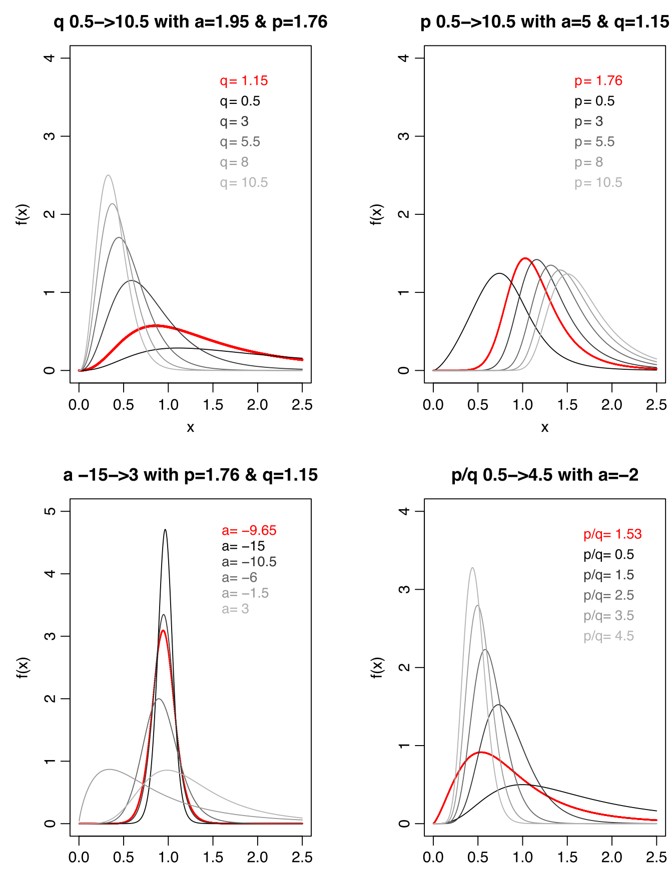

Fig. 2(b). Probability density function across shape parameters in the GB2 family. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and $\Gamma(\cdot)$ denotes the gamma function. Generally, the relative values of p and q determine the skewness of the distribution and negative values of a yield inverse distribution (Cummins et al., 1990). The moments for the GB2 distribution are expressed as follows:

$$E(Y^h) = \frac{b^h B(p + h/a, q - h/a)}{B(p, q)}. \tag{2}$$

In particular, the mean of the distribution is

$$E(Y) = \frac{bB(p + 1/a, q - 1/a)}{B(p, q)} \tag{3}$$

and the mean and variance of the GB2 distribution exist if and only if $-p < 1/a < q$ and $-p < 2/a < q$ respectively. The density function in (1) can also be expressed as scale mixtures of generalized gamma (GG) distribution as follows:

$$f(y|a, b, p, q) = \int_0^\infty f_{GG}(y|p, \lambda, a) f_{GG}(\lambda|q, b, a)\, d\lambda \tag{4}$$

where λ is the mixing parameter and the density function for the GG distribution is

$$f_{GG}(y|\alpha, \beta, \delta) = \frac{\delta[(\beta y)^\delta]^\alpha}{y\Gamma(\alpha)} \exp[-(\beta y)^\delta]$$

with moments

$$E(Y^k) = \frac{\Gamma(\alpha + k/\delta)}{\beta^k \Gamma(\alpha)}.$$

Parameters of the GB2 distribution can be tuned to obtain different special cases and thereby reduce the model complexity. It includes both Pearson and non-Pearson families of distributions. The Pearson distribution first published by Karl Pearson in 1895, is a family of continuous probability distributions, including four types of distributions (numbered I through IV) in addition to the normal distribution. The relationship of GB2 distribution with other distributions is summarized in Fig. 1 by the distribution tree in Fig. 1 of Cummins et al. (1990) adopting the part under the GB2 distribution.

Clearly, the GB2 distribution is more general than any other distributions at a lower hierarchy of the distribution tree. Fig. 1 shows that the special cases of GB2 distribution include the 3-parameter distributions of log-t (LT), generalized gamma (GG), beta of type 2 (B2), Burr types 3 and 12 (BR3 and BR12), the 2-parameter distributions of log-Cauchy (LC), lognormal (LN), Weibull (W), gamma (GA), variance ratio (F), Lomax or shifted Pareto (L), Fisk or loglogistic (Fisk) and the 1-parameter distributions of half normal (HN), Rayleigh (R), exponential (EXP), Chi-square (χ^2) and half-t (Ht) distributions.

Fig. 2(a) graphs the probability density functions for some of these special cases whereas Fig. 2(b) demonstrates how the density function of the GB2 distribution changes when one or two of the four parameters vary while the others are fixed. The density curves for all special cases are standardized by setting $b = 1$. The density curves for the GB2 distribution and its parameters in Fig. 2(b) are highlighted in red. It is clearly shown that as q or p/q decreases, the tail of the distribution becomes heavier. Larger values of a produce heavily right-skewed distributions with thicker tails whereas more negative values of a yield density functions with sharper peaks and longer, fatter tails. By suitably changing the value of a, the GB2 distribution can be expressed in diversified forms ranging from symmetric to heavily right-skewed distributions.

Accumulative Payment	Development Quarter																							
Accident Quarter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Exposure
Dec-02	0.1	0.7	2.0	4.1	7.1	11.3	17.2	22.8	27.7	34.4	40.6	54.9	67.9	77.7	85.6	94.5	102.5	106.5	115.9	119.3	122.6	124.8	130.2	2.6
Mar-03	0.1	0.7	1.6	2.5	3.8	5.5	8.6	11.2	13.9	19.8	25.9	33.3	41.2	49.7	58.8	68.1	73.5	80.1	88.5	93.8	95.2	97.7		2.6
Jun-03	0.1	0.7	1.4	2.2	3.3	5.0	7.6	10.6	14.9	20.9	27.9	33.7	40.8	51.0	62.0	67.6	75.7	81.9	86.7	88.0	95.3			2.6
Sep-03	0.1	0.7	1.5	2.6	3.9	5.9	8.0	11.6	16.0	24.0	29.8	35.2	46.2	55.8	62.2	72.7	80.1	86.3	91.7	95.8				2.6
Dec-03	0.0	0.6	1.5	2.5	3.5	4.6	7.1	10.8	16.1	20.6	28.0	34.0	46.1	51.4	61.2	69.2	75.9	80.3	88.5					2.7
Mar-04	0.1	0.6	1.6	2.5	3.4	5.1	7.4	11.6	16.6	24.7	32.1	40.8	48.8	58.6	67.3	79.1	89.1	94.7						2.7
Jun-04	0.1	0.6	1.4	2.1	3.1	5.0	7.5	11.6	17.4	23.0	30.2	36.4	46.4	52.9	64.4	71.2	76.6							2.7
Sep-04	0.1	0.6	1.3	2.2	3.3	5.4	8.6	13.4	19.2	26.7	34.7	48.9	58.3	73.2	81.7	93.3								2.8
Dec-04	0.1	0.6	1.4	2.2	3.3	5.1	9.2	15.1	21.6	27.5	39.9	49.2	63.1	69.2	79.2									2.8
Mar-05	0.0	0.5	1.5	2.3	3.4	4.9	9.0	15.7	21.2	31.8	39.7	47.9	56.6	65.3										2.8
Jun-05	0.1	0.7	1.5	2.4	3.7	6.0	11.3	15.7	23.1	31.4	40.5	47.9	63.1											2.9
Sep-05	0.1	0.7	1.8	2.7	5.0	7.9	11.7	19.9	28.8	38.3	45.2	53.8												2.9
Dec-05	0.1	0.8	1.8	3.0	4.5	7.6	13.3	20.4	29.1	36.9	46.8													2.9
Mar-06	0.0	0.5	1.1	2.0	2.9	6.1	10.8	18.0	24.8	31.8														3.0
Jun-06	0.1	0.6	1.4	2.1	4.1	8.5	15.9	22.4	30.1															3.0
Sep-06	0.0	0.6	1.3	2.2	4.0	10.2	17.1	24.8																3.0
Dec-06	0.1	0.5	1.5	3.0	5.0	9.4	17.0																	3.1
Mar-07	0.0	0.7	1.7	2.7	4.3	9.9																		3.1
Jun-07	0.1	0.6	1.4	2.3	4.3																			3.2
Sep-07	0.1	0.8	1.7	2.8																				3.2
Dec-07	0.1	0.6	1.3																					3.2
Mar-08	0.1	0.7																						3.3
Jun-08	0.1																							3.3

Fig. 3. QLD CTP cumulative payment data.

3. Bayesian methodology

The inclusion of elaborate models, such as the state space, threshold and mixture models into the mean of the four-parameters' GB2 distribution, complicates the likelihood function and its optimization considerably. The Bayesian approach converts the optimization problem in the likelihood approach to a sampling problem, and by making use of the hierarchical structure of the model and MCMC techniques, it lessens the complexity of model implementation for complicated models. In the case of nonstandard posterior distributions, MCMC techniques (Smith and Roberts, 1993; Gilks et al., 1996) with Gibbs sampling and Metropolis Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) produce samples from the intractable posterior distributions of all unknown parameters. Moreover, the prior probability distributions in Bayesian inference provide a powerful mechanism for incorporating information from previous studies, and for controlling confounding. Even in the situation where there is no agreement on the prior information, we can use non-informative or reference priors. Inference under this circumstance is so called objective Bayesian inference (Berger, 1985).

Furthermore, the emergence of WinBUGS, a user friendly software for Bayesian inference using MCMC techniques, allows non-experts to perform Bayesian analysis of complex statistical models. In this paper, all models are implemented via Bayesian approach using WinBUGS and the codes for all models are available upon request. For each model in the empirical study (Sections 4 and 5), a single Markov chain is run for 40,000 iterations, discarding the initial 10,000 iterations as the burn-in period and sampling every 30th iteration to mimic a random sample of size 1000 from the joint posterior distribution for posterior inference. Parameter estimates are given by the posterior means or medians. The autocorrelation functions and history plots are carefully checked to ensure that the posterior samples have converged and are independent. Computation time depends on the complexity of the model and power of the computer and it is around 4.5 h using a Core 2 Duo 2 GHz PC for fitting the threshold state space models (Section 4.2.3) in the empirical study.

4. Study of aggregated loss data

Aggregated loss triangles have been widely used to estimate insurance liability. Although sophisticated models have been developed to project the expected payments in the lower runoff triangle, the flexibility of the models to allow for extreme claims and legislative changes during the study period is often uncertain. In this study, we explore the use of the flexible GB2 distribution with four mean functions to allow for some extreme and irregular aggregated losses.

4.1. The data

The data we analyze is the amount of cumulative payments for all the compulsory third party (CTP) policies in Queensland (QLD) as at June 2008. CTP insurance policy covers risk that would be referred to as Auto Bodily Injury in the US and Motor Bodily Injury in the UK. In order to remove the influence of inflation for reserving purposes, we utilize the average weekly earning index from the Australian Bureau of Statistics (ABS) to inflate all the values to December 2008 dollars. Hence, the data used in this analysis represents the inflated cumulative payment for QLD CTP portfolio. The data are in the units of millions summarized by accident and development quarters covering periods from December 2002 to June 2008. It contains 276 observations over 23 accident quarters as reported in Fig. 3.

Since CTP insurance policy covers accidental bodily injury or death of third parties as a result of road traffic accidents, the pay out period is typically long. People who are severely injured in a motor vehicle accident require long term medical treatment and rehabilitation, resulting in substantially high losses. In certain circumstances, if the case goes to court, a legal procedure can be extraordinarily long, and furthermore legal costs can be very high. A substantial part of claims cost is from larger CTP claims which take longer to settle. Hence we propose the GB2 distribution with a flexible tail to model the extreme aggregate loss.

4.2. Mean model

In order to capture the irregular claims behaviors, we apply the following four mean models: ANOVA, state space, threshold and state space threshold. For each of the models, we use the accident and development quarter as our covariates, and offset by the number of polices in force (n_i) in each accident quarter.

4.2.1. ANOVA model

Let Y_{ij} denote the aggregated total claims payment made in accident quarter i, settled in lag quarter j and n_i denote the total number of polices in force in accident quarter i. We apply the two factor ANOVA model (Model 1) as follows:

$$Y_{ij} \sim GB2(a, b_{ij}, p, q), \tag{5}$$

$$b_{ij} = \frac{E(Y_{ij})B(p, q)}{B(p + 1/a, q - 1/a)} \tag{6}$$

$$\log(E(Y_{ij})) = \mu_{ij} + \ln(n_i), \tag{7}$$

$$\mu_{ij} = \mu_0 + \alpha_i + \beta_j, \tag{8}$$

where the parameters α_i and β_j which denote accident quarter and development quarter effects respectively satisfy the following constraints:

$$\alpha_1 = \beta_1 = 0. \tag{9}$$

The following diffuse priors:

$$\mu \sim N(0, 100), \qquad \alpha_i \sim N(0, 100), \qquad \beta_j \sim N(0, 100), \tag{10}$$

$$a \sim N(0, 100), \qquad p \sim Ga(0.001, 0.001), \\ q \sim Ga(0.001, 0.001) \tag{11}$$

are assigned to the model parameters where $N(\mu, \sigma^2)$ represents the normal distribution with mean μ and variance σ^2; $Ga(r, u)$ represents the Gamma distribution with mean r/u and variance r/u^2. For parameters without restricted ranges, we assign normal distributions with zero mean and large variance as there is no prior information on their values. For shape parameters with a positive range, they are assigned Gamma distributions with unit mean and large variance to reduce the detrimental effect of estimation risk. In general, this set of priors applies to subsequent analyses.

4.2.2. State space model

The idea behind a state-space representation of a linear model is to capture the dynamics of an observed $(n \times 1)$ vector $\boldsymbol{y}_t$ in terms of a possibly unobserved $(r \times 1)$ vector $\boldsymbol{\xi}_t$ known as the state vector for the model. The dynamics of the state vector are taken to be vector:

$$\boldsymbol{\xi}_{t+m} = \boldsymbol{F}^{\boldsymbol{m}}\boldsymbol{\xi}_t + \boldsymbol{F}^{\boldsymbol{m-1}}\boldsymbol{\upsilon}_{t+1} + \boldsymbol{F}^{\boldsymbol{m-2}}\boldsymbol{\upsilon}_{t+2} + \cdots + \boldsymbol{F}^{\boldsymbol{1}}\boldsymbol{\upsilon}_{\boldsymbol{t+m-1}} \\ + \boldsymbol{F}^{\boldsymbol{0}}\boldsymbol{\upsilon}_{\boldsymbol{t+m}} \quad \text{for } m = 1, 2, \ldots, \tag{12}$$

where $\boldsymbol{F}$ denotes an $(r \times r)$ matrix and the $(r \times 1)$ vector $\boldsymbol{\upsilon_t}$ is taken to be an i.i.d. random vector with zero mean, $\boldsymbol{F^m}$ denotes the matrix $\boldsymbol{F}$ multiplied by itself m times. Hence the optimal m-period-ahead forecast is seen to be

$$E(\boldsymbol{\xi}_{t+m}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t+1}\ldots) = \boldsymbol{F}^m\boldsymbol{\xi}_t. \tag{13}$$

Future values of the state vector depend on $(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots)$ only through the current value $\boldsymbol{\xi}_t$. This framework avoids the need for the detailed tailoring of calculations or mode of analysis and hence is suitable in a wide variety of circumstances.

In particular, the state space (SS) model has the flexibility to allow parameters to develop in an auto-regressive process. When the SS model is applied to describe both the accident (α) and j-th development quarter (β_j) effects with $m = 1$ and $r = 1$ in (12), the resultant model (Model 4) is given by (5) to (7) and the following:

$$\mu_{ij} = \mu_0 + \alpha_i + \beta_{ij}, \tag{14}$$
$$\alpha_i = \alpha_{i-1} + \upsilon_i,$$
$$\beta_{ij} = \beta_{i-1,j} + \nu_i,$$

where the interaction between accident and development quarters is incorporated by β_{ij} giving different β_{ij} for different accident and development quarters and α_i and β_{ij} satisfy the following constraints:

$$\alpha_1 = \beta_{1j} = 0. \tag{15}$$

Diffuse priors in (10) and (11) are assigned to the model and further

$$\upsilon_i \sim N(0, 100), \qquad \nu_i \sim N(0, 100).$$

4.2.3. Threshold model

The threshold (T) model allows the mean function to vary before and after a threshold variable, such as time T. The model is very useful in accounting for events, such as legislation changes and catastrophes in reality. CTP policies normally possess long pay out periods, allowing claims to be exposed to more legislation changes or other unexpected events. In the aggregated loss data, there is a major legislative change: the Civil Liability Act (CLA) 2003 during the claims accident period. Under this new Act, injuries are assigned a point value between 1 and 100 where zero relates to an injury not severe enough to justify any award of general damages and one hundred is an injury of the gravest conceivable kind (www.maic.qld.gov.au).

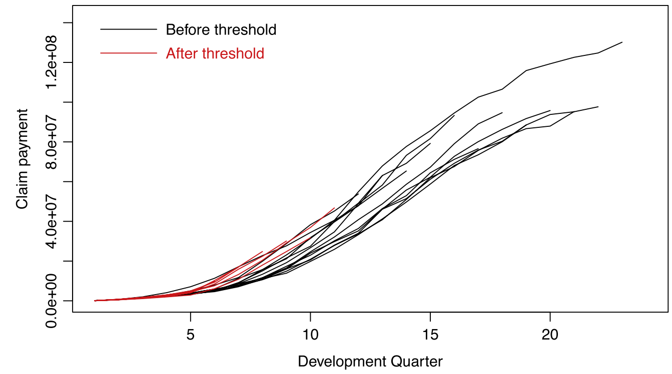

Fig. 4. QLD CTP cumulative claims payment by development quarter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The main implication is that the mean of total claims payment might shift after the new Act took effect. Fig. 4 plots the time series of aggregate loss across development quarters for each accident quarter. The time series before and after December 2005 $(T = 13)$ are marked in black and red lines respectively. The red lines show sharper increases than the black lines showing a lag effect of CLA 2003. We have tuned the model with T ranging from 12 to 14 and found $T = 13$ provides the best model fit. This can be possibly explained by the fact that the rate of finalization of claims change considerably across $T = 13$. A further effect is that the claim frequency steadily decreases with increasing accident quarter. This results from the gradual elimination of lower-severity claims and hence the claim sizes displays rates of increase in excess of inflation.

Furthermore, the data set displays a burst of heavy superimposed inflation, which is claims inflation in excess of normal economic inflation, over the period from early 2006 to early 2008, and quite separate from the effect addressed in the preceding paragraph. The state space threshold model addresses this data pattern by allowing a flexible development year parameter (β_{ij} rather than β_j). Moreover, the threshold is estimated to be $T = 13$ which is at the beginning of 2006. Our model precisely detected this feature of the data. Adopting the same mean function throughout all accident quarters might fail to allow for the model shift after $T = 13$ caused by the lag effect of CLA 2003. Therefore, we introduce the state space threshold model to analyze this data. The adaptive nature of the state space threshold model will cause it to recognize the escalation of claims sizes eventually. However, there are others ways of constructing the model to recognize them immediately, such as adding time related terms in the model. This model will be explored in the future.

The threshold model can be used in conjunction with different kinds of mean models including the ANOVA (Model 2) and state space (Model 5) models. The state space threshold model is expressed as below:

$$Y_{ij} \sim GB2(a, b_{ij}, p, q),$$

$$(a, p, q) = \begin{cases} (a_1, p_1, q_1) & \text{for } i \le T, \\ (a_2, p_2, q_2) & \text{for } i > T, \end{cases}$$

$$\mu_{ij} = \begin{cases} \mu_{1ij} = \upsilon_{10} + \alpha_{1i} + \beta_{1ij} & \text{for } i \le T, \\ \mu_{2ij} = \upsilon_{20} + \alpha_{2,i-12} + \beta_{2,i-12,j} & \text{for } i > T, \end{cases} \tag{16}$$

$$\alpha_{ki} = \alpha_{k,i-1} + \upsilon_{ki},$$
$$\beta_{kij} = \beta_{k,i-1,j} + \nu_{ki}$$

where b_{ij} and $E(Y_{ij})$ are given by (6) and (7), $\upsilon_{ki} \sim N(0, \sigma^2_{\upsilon_k})$, $\nu_{ki} \sim N(0, \sigma^2_{\nu_k})$ and $\alpha_{k,i-1}$ and $\beta_{k,i-1,j}$ satisfy the constraints:

$$\alpha_{11} = \beta_{11,j} = \alpha_{21} = \beta_{21,j} = 0.$$

Table 1
Estimated GB2 distribution parameters.

Models	a	b[a]	p	q
M1 GB2 ANOVA	−8.67	19.3	0.95	21.80
M2 GB2 T ($i < 13$)	−9.073	31.0	1.55	3.24
($i \geq 13$)	−7.59	6.0	1.49	12.42
M3 GB2 T ($i < 13$)	−9.08	31.1	1.54	2.61
($i \geq 13$)	–	6.8	–	–
M4 GB2 SS	−8.68	17.9	1.17	39.38
M5 GB2 SS T ($i < 13$)	−4.95	24.1	1.95	9.12
($i \geq 13$)	−4.80	5.9	1.84	7.42
M6 GB2 SS T ($i < 13$)	−4.96	24.1	1.97	9.20
($i \geq 13$)	–	5.2	–	–

[a] In unit of millions.

Alternatively, simplified ANOVA threshold (Model 3) and state space threshold (Model 6) models could be considered by setting the shape parameters of the GB2 distribution to be consistent across the threshold ($T = 13$); that is,

$$a_1 = a_2, \qquad p_1 = p_2, \qquad q_1 = q_2, \tag{17}$$

if they are similar across T for Models 2 and/or 5. Note that an alternative hierarchical form for Models 1 to 6 using the scales mixtures representation in (4) is

$$Y_{ij} \sim GG(p, \lambda_{ij}, a),$$
$$\lambda_{ij} \sim GG(q, b_{ij}, a) \tag{18}$$

where b_{ij} and $E(Y_{ij})$ are given by (6) and (7), μ_{ij} for the ANOVA, state space and threshold state space models are given by (8), (14) and (16) respectively.

4.3. Numerical result

We start with fitting the traditional ANOVA model with three choices of distributions, Gamma, GT and GB2 distributions to the data. We have also used the chain ladder method, which is a widely recognized method of loss reserving for benchmarking.

The former two distributions are chosen because the Gamma distribution has been widely used by actuaries and the GT family does not nest within the GB2 family. Moreover models with the GB2 distribution and different mean functions are also attempted. Parameter estimates for Models 1 to 6 are reported in Table 1. Note that b_{ij} in (6) varies across accident quarter i and development quarter j for all models. The reported b is an average of b_{ij} over all quarters for Models 1 and 4. For the remaining models, the two b reported in Table 1 are averaged over accident quarters $T < 13$ and $T \geq 13$ respectively. From Table 1, the estimated values of a are always negative, which indicates inverse distributions; the estimate of p from Model 1 is close to 1, which implies the Burr type 12 error distribution. The GB2 distribution is the most suitable distribution for the remaining models. In other words, any conventional distributions within the GB2 family are less suitable and hence yield a less accurate prediction in loss reserving.

To evaluate these models, two criteria, model-fit and prediction accuracy, are considered. To assess the model-fit, two criteria: the R percentage and deviance information criteria (DIC) are adopted. The R percentage is the mean of predicted over actual loss less one, which is a popular measure to quantify the difference between actual and predicted values whereas DIC originated by Spiegelhalter et al. (2002) is a Bayesian analogue of Akaike's Information Criterion (AIC) which is commonly used in Bayesian analysis. DIC consists of a measure of model fit which is the posterior mean deviance, and a measure of model complexity which is an estimate of the effective number of parameters. It has a competitive advantage over the traditional AIC as it is not only limited to nested

Table 2
Model selection for aggregated data.

Models	DIC	R (%)
M0a GT ANOVA	16,142	−4.58
M0b Gamma ANOVA	8,586	−7.10
M0c Chain Ladder	–	−1.87
M1 GB2 T	8,521	−2.98
M2 GB2 ANOVA	8,566	−2.04
M3 GB2 SS T	8,651	−0.58
M4 GB2 T[a]	8,504	−2.55
M5 GB2 SS	8,093	−2.48
M6 GB2 SS T[a]	8,566	0.54

[a] Represents models with different p and q values before and after T.

models. The DIC is given by

$$DIC = -\frac{4}{M}\sum_{m=1}^{M}\sum_{i=1}^{23}\sum_{j=1}^{24-i}\ln\left[f(y_{ij}|\boldsymbol{\theta}^{(m)})\right] + 2\sum_{i=1}^{23}\sum_{j=1}^{24-i}\ln\left[f(y_{ij}|\bar{\boldsymbol{\theta}})\right] \tag{19}$$

where $\boldsymbol{\theta}^{(m)}$ denotes the vector of parameter estimates in the m-th iteration of the posterior sample $M = 1000$, $\bar{\boldsymbol{\theta}}$ denotes the posterior mean of $\boldsymbol{\theta}^{(m)}$ and $f(y_{ij}|\boldsymbol{\theta})$ represents the observed likelihood in (1) for each observation where b is given by (6) and μ_{ij} in $E(Y_{ij})$ is given by (8), (14) and (16) for the three types of mean models.

Predictive performance is assessed by comparing the predicted aggregated loss with the actual loss in the last diagonal of the triangle ($i = 1, \ldots, 23$ and $i + j = 24$). The predicted aggregated loss $\hat{y}_{ij} = E(y_{ij})$ is calculated using (7) where μ_{ij} is given by (8), (14) and (16) for the three types of mean models. To project the cumulative loss in the lower triangle, the parameter estimates $\beta_{2,i-12,j}$ where both accident and development quarters are beyond the threshold $T = 13$ are unavailable and they are estimated by $\beta_{1j}\,\frac{\beta_{2,j-1}}{\beta_{1,j-1}}$ for the threshold ANOVA model and $\beta_{1,i-12,j}$ for the threshold state space model. Then the total of aggregate loss (as at June 2008), the observed y_{ij} and predicted $\hat{y}_{ij}$ are calculated as

$$y_o = \sum_{i=1,j=24-i}^{23} y_{ij} \quad \text{and} \quad y_p = \sum_{i=1,j=24-i}^{23} \hat{y}_{ij}$$

and the ratio defined as $R = \frac{y_p}{y_o} - 1$ is reported in Table 2 together with y_p and DIC.

Model with the smallest DIC and R percentage is preferred. R percentage measures the prediction accuracy. The GB2 state space threshold model (M6) with the lowest R percentage demonstrates the best predictive power as shown in Table 2. The model fit is quantified by DIC. The GB2 state space model (M5) provides the best model fit with the lowest DIC. It is not surprising that Model 0a and 0b perform less favorably according to DIC because the GT family requires transformation of the data and Gamma is a special case of the GB2 family. From a distribution perspective, the GB2 (M2) model outperforms the Gamma (M0a) and GT (M0b) models given the same ANOVA mean function; from a mean function perspective, the threshold (M4) and state space models (M5) outperform the ANOVA (M2) given the same GB2 distribution.

In Fig. 5, we apply Model 6 to project the claims payment flight path. The predicted cumulative loss $\hat{y}_{ij}$ in the lower triangle are plotted in dotted black lines when $i < T$ and dotted red lines when $i \geq T$ whereas the observed y_{ij} in the upper triangle are plotted in solid lines. The figure demonstrates a general upward trend at a fast rate until the 15-th development quarter, a slow rate thereafter, and a gradual level off. It also shows a distinct pattern before and after the threshold T: the later pattern shows a sharper upward trend at an earlier development quarter.

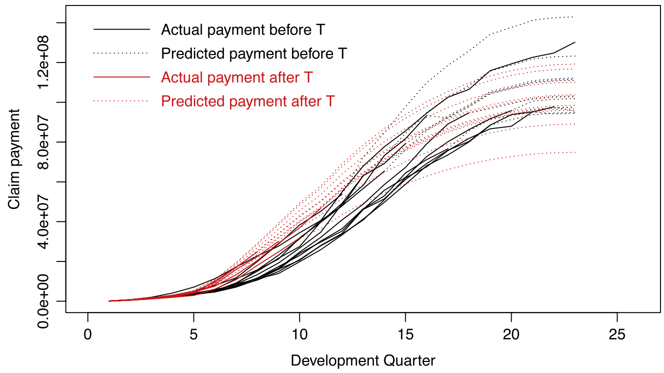

Fig. 5. QLD claims payment projection by development quarter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6 presents three triangular heat maps to visualize the ratio of fitted to actual loss by accident and development quarters in the upper triangle for three models. Generally speaking, green represents good fit; yellow color indicates overprediction whereas blue color reveals underprediction. There are more green cells in the first graph than the second one, which implies that the GB2 ANOVA model provides a better fit than the Gamma ANOVA model given the same ANOVA mean function; with more blue cells in the second graph, the Gamma model shows predominantly an underprediction. Comparing the first and third graphs, the GB2 state space threshold model (M6) with more green cells clearly demonstrates the best fit to the data, which is in conjunction with the results in Table 2. Some common patterns are observed in the three heat maps. For example, the underprediction in the first 4 to 8 development quarters of the accident quarter 11 (Jun-05) exists in all three graphs, and also the underprediction from the 6 to 8 development quarters of accident quarter 1 (Dec-02) in the first and second graphs, but became much less in the third graph. One possible reason is that the Civil Liability Act 2003 applied to accident dates on and after the first of December 2002. So the majority of accident quarter 1 is subject to a legislative regime different from that applying to all subsequent accident quarters. The state space threshold model represented by the third graph offers different parameters for each development and accident quarter, thereby recognizing these patterns well. Furthermore, the threshold models assume a threshold at accident quarter 13 (Dec-05), which coincides with a number of significant changes in claims experience occurring around this period. It is generally agreed by actuaries for Queensland CTP insurers and for the Queensland regulator that the relative incidence of lower severity claims increased, rates of claim finalization changed in response, and that claim sizes were affected by very high superimposed inflation between payment quarters Jun-06 to Mar-08. The threshold models, to some extent, recognize these changes.

4.4. Simulation study

In this simulation study, we evaluate the performance of the basic GB2 ANOVA model. The performance of other GB2 models will be similar due to the same distributional assumption. The GB2 ANOVA model has 48 parameters. We use the estimated parameter values to simulate $N = 200$ data sets; each contains $n = 276$ observations, which is the same as the size of the QLD CTP payment data. Table 3 reports the mean, absolute percentage bias (APB), and standard deviation (SD) of the parameter estimates over $N = 200$ replications. APB is expressed as the absolute value of the difference between estimated and true value as a proportion of the true value. The results show that the parameters involved in the mean function including μ, α_i and β_j achieve a high level of accuracy whereas the shape parameters are estimated to a moderate level of accuracy. It is well known that shape parameters are often more difficult to estimate because distributions are less sensitive to some shape parameters. However, as the main focus of the analysis is on the projection of loss reserve based on the expected mean value of each cell in the loss triangle, the shape parameter estimates have a minimal effect on the projection. In summary, the model performance is satisfactory, and therefore the parameter estimates and forecasts in the empirical study of aggregated data are reliable.

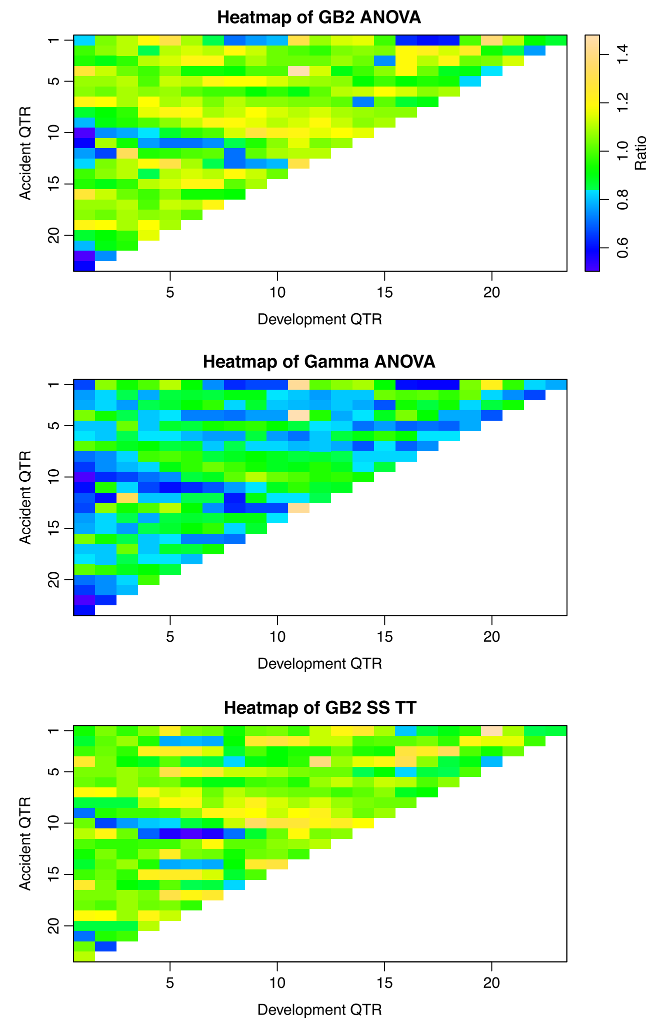

Fig. 6. Triangular actual vs. expected heat map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Simulation results.

Parameters	a	p	q	μ_0	α_1	β_1
True values	−8.667	0.946	21.800	2.009	0.315	−5.895
Estimated values	−9.105	1.108	23.017	2.067	0.317	−5.892
APB	0.051	0.171	0.056	0.029	0.005	0.000
SD	1.507	0.371	4.485	0.019	0.031	0.034

5. Study of individual loss data

Model for individual loss data has been increasingly adopted to analyze the effect of individual characteristics because it helps to identify the underlying drivers of claim cost. It provides not only a link between changes in the claims processes and reserving, but also an understanding and quantification of the drivers of a claim. Moreover, loss reserve models for individual claims provide individual estimates of future claim costs arising from existing claims. These individual predictions form the basis of reinsurance premium calculations; it is as important as total reserve estimation.

5.1. The data

In this study, the data that we analyze refers to the workers' compensation (WC) journey claims which have been re-directed to CTP insurers in QLD as of June 2008. It consists of 2516 individual claims. In order to predict the full claim cost, the data include only the finalized claims and are inflated to December 2008. Three key variables are selected for the WC claims predictors. Their definitions and levels are given as follows:

Finalization delay (F_i): The number of months taken for the claims to be finalized.

Role of claimants: The role of claimants in an accident (Levels: Driver (Dri; $x_{1i} = 1, x_{2i} = 0$), motorcycle rider (Mcr; $x_{1i} = 0$, $x_{2i} = 1$), (Passenger; $x_{1i} = x_{2i} = 0$)). For each claim i, exactly one of the x_{ki} takes a nit value and the remainder are equal to zero.

Treatment Indicator: The length of treatment the claimant received (Levels: Short term ($\leq$6 weeks; $x_{3i} = 1$), long term ($>$6 weeks; $x_{3i} = 0$))

The Ninety, Ninety-five and ninety-nine percentiles of the data are 13,657, 46,424 and 325,322 dollars respectively, showing a dramatic increase from the ninety-fifth to ninety-ninth percentiles. A preliminary data analysis reveals that the maximum value of the data is 85 times the mean. Moreover, the mean, standard error, skewness and kurtosis for the data are 3839, 13 955, 14.09 and 258.38, respectively and the claims above the 95th percentile account for 50.4% of the total claims cost. All these features present strong evidence that the WC data are heavily tailed and exhibit considerable heterogeneities. Modeling the tail behavior is sure to have a high impact on the accuracy of loss reserve.

5.2. The mixture model

In general insurance practice, small and large CTP claims present distinct risk characteristics based on managers' claims handling experience. Consequently, CTP claims managers usually handle small and large claims separately. We adopt an alternative modeling approach in which a single model is applied to the whole data set with the characteristics of both groups. From a modeling perspective, this allows the share of information, resulting in a more efficient model.

To allow for such heterogeneity in the WC data, we assume there are two underlying subgroups in the mixture of GB2 model. It captures group-specific characteristics arising from the low and high level loss payments in the WC data. Besides, the model facilitates classification of loss payment into different risk groups, thereby providing insurance companies with a greater insight so as to distinguish claims at an earlier stage.

Suppose that there are two risk groups and each claim has a probability, $\pi_k \geq 0$ of coming from group k, $k = 1, 2$ and $\pi_2 = 1 - \pi_1$. We define the unobserved group-k membership indicator $I_{ki} = 1$ if a claim i belongs to group k and $I_{ki} = 0$ otherwise and it follows the Bernoulli distribution with probability π_k. If a claim Y_i belongs to group k,

$$Y_i \sim GB2(a_k, b_{ik}, p_k, q_k), \tag{20}$$

$$b_{ik} = \frac{E(Y_i)B(p_k, q_k)}{B(p_k + 1/a_k, q_k - 1/a_k)}, \tag{21}$$

$$\log(E(Y_i)) = \mu_k + \alpha_k F_i + \beta_{1k} x_{1i} + \beta_{2k} x_{2i} + \beta_{3k} x_{3i}. \tag{22}$$

Then Y_i is said to arise from a finite mixture model with probability density function $f(y)$:

$$f(y) = \pi_1 * f_1(y) + (1 - \pi_1) * f_2(y),$$

where $f_k(y)$ is the probability density function (1) for component k with parameters $\boldsymbol{\theta}_k = (a_k, p_k, q_k, \mu_k, \alpha_k, \beta_{1k}, \beta_{2k}, \beta_{3k})^T$ and together with the missing observation I_{ki}, the whole vector of parameters $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \pi_1, \boldsymbol{I}_1)$ where $\boldsymbol{I}_1 = (I_{11}, \ldots, I_{1n})$, π_1 is the weight. Note that $\boldsymbol{\Theta}$ contains two types of parameters, the model parameters such as p_k and β_{jk} and the missing group membership I_{1i}. Both of them are estimated by drawing samples from their conditional posterior distributions.

To derive posterior distributions for the model parameters, the following diffuse priors are assigned to the above modeling parameters. The weight π_k is assigned a non-informative Uniform prior on the range from 0 to 1, as it represents the weights between the two components of the mixture model. The shape parameters p_k and q_k are assigned Gamma distributions with different sets of hyper-parameters, to allow for more flexible tail behavior which is described by the ratio between p_k and q_k. The mean and variance of q_k are set to be larger than that of p_k as q_k is more sensitive and tends to take greater values from experience.

$$\mu_k \sim N(0, 100), \qquad \alpha_k \sim N(0, 100),$$
$$\beta_{jk} \sim N(0, 100), \qquad \pi_k \sim U(0, 1),$$
$$a_k \sim N(0, 100), \qquad p_k \sim Ga(0.001, 0.001),$$
$$q_k \sim Ga(0.01, 0.0001)$$

where $U(a, b)$ denotes the uniform distribution with support a to b.

One issue we encountered in implementing the mixture of GB2 distribution via the Markov chain Monte Carlo (MCMC) Bayesian approach is the label switching problem. It is primarily caused by the likelihood of a mixture model being invariant to permutations of the labels. The permutation can change many times across MCMC iterations making it difficult to infer component-specific parameters of the model. Lee et al. (2008) solved the label switching problem by imposing identifiability constraints on the parameters in a normal mixture model. We adopt the same idea by constraining the intercept of the first group to be smaller than that of the second group; that is $\mu_1 < \mu_2$. In other words, μ_2 is sampled from the range of (μ_1, ∞). This constraint ensures the vector of parameter estimates corresponds to its unique claims group in each MCMC iteration. The adoption of this identifiability constraint has substantially stabilized the parameter estimates in the simulation process resulting in more reliable measures of component specific effects.

5.3. Numerical result

We start with Gamma and GB2 error distributions without mixture effects. We then consider the proposed mixture model. The parameter estimates in the mean function for all three models are reported in Table 4. The direction of effects in the mean function across models is consistent for the four variables. Longer finalization delay and length of treatment lead to higher claim cost; the costs associated with drivers are less than those of passengers which in turn is less than those of motorcycle riders. Classification of loss payments into the two distinct risk groups can be performed using I_{ik}: loss payment i is assigned to risk group k if $\bar{I}_{ik} > 0.5$. For the mixture GB2 model, the first group consists of lower claims with a mean of 1621.5 and accounting for 60% of claims whereas the second group contains mainly large claims with a mean of 8405.0. From checking the credible intervals for the differences in parameter estimates across the two groups, we found that α and β_3 are significantly different.

Table 4
Parameter estimates for the mean function.

Models	Intercept μ	Fin. del. α	Role: Dri. β_1	Role: Mcr β_2	Tre: Sho. β_3	Prob. π
Gamma	7.520	0.040	−0.404	0.587	−0.145	–
GB2	9.273	0.015	−0.162	0.198	−0.131	–
Mixture GB2 (lower)	7.342	0.004	−0.089	0.036	−0.034	0.592
Mixture GB2 (higher)	8.165	0.045	−0.357	0.309	−0.425	0.408

Table 5
Parameter estimates and model-fit measures for individual loss data.

Models	a	b	p	q	*RMSE*	*DIC*
Gamma	6.99[a]	2.00	–	–	16,454	47,832
GB2	−2.04	502.26	0.52	1.72	16,351	44,152
Mixture GB2 (lower)	−2.67	127.97	0.56	44.22	15,634	56,367
Mixture GB2 (higher)	−1.29	4594.94	1.35	1.16	–	–

In the Gamma model, a and b denote the scale and shape parameters respectively.
[a] The value of a should be multiplied by 10,000.

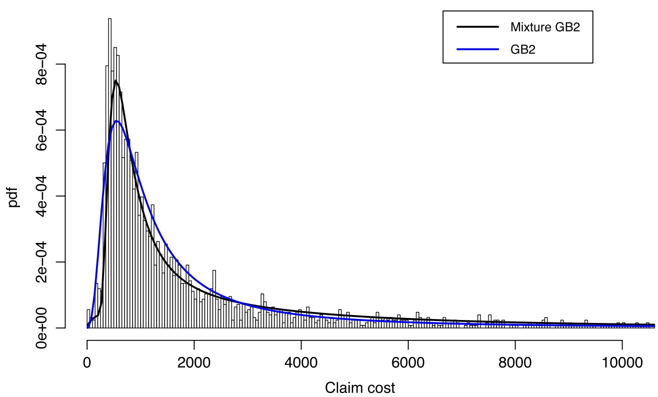

Fig. 7. Density of mixture GB2 and GB2 models.

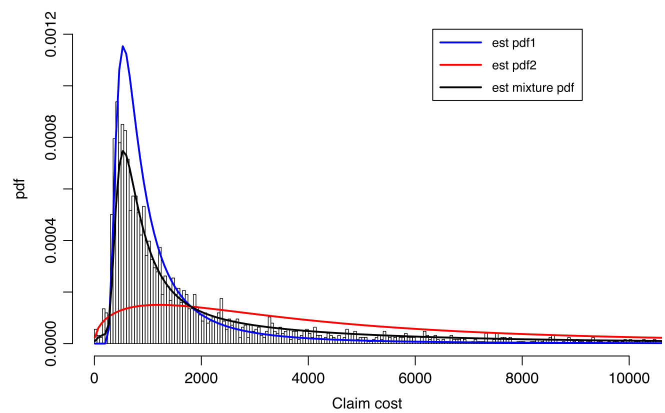

Fig. 8. Density of mixture GB2 distribution and its components. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To compare these models, we evaluate the model-fit using *DIC*. Celeux et al. (2002) advanced the DIC for mixture model as follows:

$$DIC = -\frac{4}{M}\sum_{m=1}^{M}\sum_{i=1}^{n}\sum_{k=1}^{G} I_{ki}^{(m)} \ln\left[\pi_k^{(m)} f(y_i|\boldsymbol{\theta}_k^{(m)})\right] + 2\sum_{i=1}^{n}\sum_{k=1}^{G}\bar{I}_{ik}\ln\left[\overline{\pi}_k f(y_i|\overline{\boldsymbol{\theta}}_k)\right] \quad (23)$$

where $\boldsymbol{\theta}_k^{(m)}$ denotes the vector of model parameter estimates in the m-th iteration of the posterior sample for group k, G is the number of groups, $I_{ik}^{(m)}$ denotes the estimate of I_{ik} in the m-th iteration, $\bar{\boldsymbol{\theta}}_k$ and $\bar{I}_{ik}$ denotes the posterior mean of $\boldsymbol{\theta}_k^{(m)}$ and $I_{ik}^{(m)}$ respectively. $f(y_i|\boldsymbol{\theta}_k)$ represents the observed likelihood in group k where $E(Y_i)$ in b_{ik} is given by (22). Parameter estimates for the three models and measures of model-fit including RMSE and DIC are listed in Table 5. The results show that *DIC* favors the GB2 model whereas root mean square error *RMSE* indicates the mixture GB2 model provides the best fit. The mixture GB2 model provides a less satisfactory *DIC* because all group membership indicators I_{1i} are considered as parameters, thereby substantially inflating the number of parameters for the mixture GB2 model as compared to the GB2 model without mixture effects. The model fit component of *DIC* for the mixture GB2 model is −46,781, which is in fact less than −44,143 for the GB2 model ($G = 1$). Moreover, Fig. 7 clearly shows that the mixture GB2 model outperforms the model using a single GB2 distributions in modeling the peak for the small claims and the tail for the large claims.

In the mixture GB2 model, the parameter estimate of q in the lower level group with a large value reveals that the distribution approaches generalized gamma whereas that of the higher level group (q close to 1) tends to Burr type 3. In all cases, a possesses negative values which imply inverse distributions. Note that the values of b are averages over all observations for the first two models in Table 5. In the mixture case, the two b are further weighted by the estimates of the group membership indicators $\bar{I}_{ki}$. Estimates of all shape parameters (a, p and q) are found to be significantly different between risk groups as revealed by the credible intervals. Since most of the parameter estimates are significantly different across the two groups, there is a strong evidence that the two group mixture model is most suitable for the WC data. Fig. 8 provides a graphical illustration that the density curve of the mixture of GB2 distributions closely envelops the empirical data histogram. The lower level group (in blue) accounts for the peak of the density whereas the higher level group (in red) has a thicker tail than that of the lower level group to accommodate extremely large claims. The two groups compensate for the deficiency of each other in the combined mixture model (in black).

As the mixture model enables identification of claims groups using $\bar{I}_{ki}$, it provides an insight for an insurance company to manage claims. Fig. 9 plots the indicator estimates $\bar{I}_{1i}$ for the lower level claim group. The red line indicates the cut off between lower and higher level claims groups. We can see that as claims cost increases, the proportion of claims classified to the lower claims group decreases.

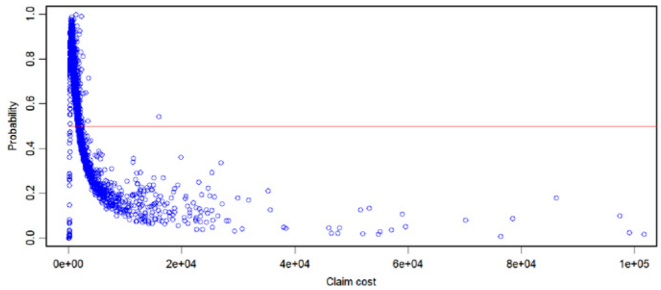

Fig. 9. Membership of the lower level claims group.

6. Conclusions

In this paper, we explore dynamic modeling for the long tail loss reserving data. The flexible modeling approach presented in Section 4 allows for the distributional parameters and means to evolve with respect the actual circumstances of the data. We demonstrate the threshold and state space mean models under the GB2 distribution tailoring to the CTP loss reserving data under the influence of legislation changes. The GB2 distribution is shown to encompass many useful heavy and light tail distributions, and hence provides sufficient flexibility to address the tail where a substantial amount of losses are above the 95th percentile. Although Cummins et al. (2007) applied GB2 distribution to the claims data, they fitted separate distributions to the claims in each cell of the payout triangle to allow for the change in risk over time and across lags. The parameters estimated using this approach are only based on the data in one cell, and hence resulting in less reliable estimates because of the limited data size and ignorance of dependency between cells. Our approach significantly improves their model by utilizing all the data in the payout triangle, and at the same time allowing for any unexpected change across accident or development years. The simulation study in Section 4.4 confirms that our parameter estimates are very reliable and achieve a high level of accuracy.

The second part of this paper (Section 5) presents the GB2 distribution under the mixture framework for the individual loss data tailoring to its heterogeneous features. The mixture framework allows for any parameters to vary across different risk groups. In real insurance practice, the mixture model has a major advantage over traditional models as it models unobserved heterogeneity in the data and enables classification of claims into different risk groups. The resulting groups provide an insight to the insurance companies for better claim management. On the other hand, it allows actuaries to focus on the high risk group for the accurate projection of large claims cost, which has a substantial impact on expected severities under reinsurance contracts and recoveries calculations. To facilitate the implementation of the proposed models, we present the Bayesian hierarchy using WinBUGS. Moreover, we introduce the alternative hierarchical forms of the models using the scales mixtures representation for the GB2 distribution to simplify the Gibbs sampler in the MCMC simulation.

Although our proposed model is flexible enough to model claims in many real situations, there are certainly ways for further improvements. In particular, we can explore the possibilities of allowing the shape of the distribution to change across each risk combination by adopting more flexible forms of the distributional parameters. It can be achieved by equating functions of individual claim characteristics, accident and development quarters etc., not only to the mean, but also to the distributional parameters. This is equivalent to building one model for each risk combination, but with the benefit of incorporating all data. Yang et al. (2011) consider a multivariate GB2 model to capture non-elliptical and asymmetric dependencies among claim portfolios. Extending our proposed approaches to multivariate modeling of claims will surely increase the applicability of the models considerably. Although the full extent of the dynamic models is still to be evaluated, our results do show promise.

Acknowledgment

The authors wish to acknowledge the useful comments of Dr. Gareth Peters from University College London.

References

Aiuppa, T.A., 1988. Evaluation of Pearson curves as an approximation of the maximum probable annual aggregate loss. Journal of Risk and Insurance 55, 425–441.
Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, second ed. Springer-Verlag, New York.
Celeux, G., Forbes, F., Robert, C.P., Titterington, D.M., 2002. Deviance information criteria for missing data models. The Review of Economics and Statistics 69 (2), 232–240.
Chan, J.S.K., Choy, S.T.B., Makov, U.E., 2008. Robust Bayesian analysis of loss reserves data using the generalized-*t* distribution. Astin Bulletin 38 (1), 207–230.
Cummins, J.D., Dionne, G., McDonald, J.B., Pritchett, B.M., 1990. Application of the GB2 family of distributions in modelling insurance loss processes. Insurance: Mathematics and Economics (9), 257–272.
Cummins, J.D., Lewis, C.M., Philips, R.D., 1999. Pricing excess of loss reinsurance contracts against catastrophic loss. In: Froot, Kenneth (Ed.), The Financing of Catastrophe Risk. University of Chicago Press, Chicago.
Cummins, J.D., McDonald, J.B., Craig, M., 2007. Risk loss distributions and modelling the loss reserve pay-out tail. Review of Applied Economics 3 (1–2), 1–23.
de Alba, E., 2002. Bayesian estimation of outstanding claim reserves. North American Actuarial Journal 6 (4), 1–20.
De Jong, P., Penzer, J., 2004. The ARMA model in state space form. Statistics and Probability Letters 70, 119–125.
De Jong, P., Zehnwirth, B., 1983. Claims reserving state space models and Kalman filter. The Journal of the Institute of Actuaries (110), 157–181.
Gilks, W.R., Richardson, S., Spiegelhalter, D. (Eds.), 1996. Practical Markov Chain Monte Carlo. Chapman-Hall, New York.
Hamilton, J.D., 1994. Chapter 50 State-space models. Handbook of Econometrics 4, 3039–3080.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 57, 97–109.
Koop, G., Potter, S.M., 1999. Dynamic asymmetries in US unemployment. Journal of Business and Economic Statistics 17, 298–313.
Lee, K., Marin, J., Mengersen, K., Robert, C., 2008. Bayesian inference on mixtures of distributions. Working Paper.
Li, W.K., Lam, K., 1995. Modelling the asymmetry in stock returns by a threshold ARCH model. Journal of the Royal Statistical Society: Series B 44, 333–341.
Ling, S., 1999. On the probabilistic properties of a double threshold ARMA conditional heteroskedasticity model. Journal of Applied Probability 36, 688–705.
Mack, T., 1991. A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. Astin Bulletin 21, 93–109.
McDonald, J.B., Butler, R.J., 1987. Some generalized mixture distributions with an application to unemployment duration. The Review of Economics and Statistics 69 (2), 232–240.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., 1953. Equations of State calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1091.
Montgomery, A.L., Zarnowitz, V., Tsay, R.S., Tiao, G.C., 1998. Forecasting the US unemployment rate. Journal of the American Statistical Association 93, 478–493.
Ntzoufras, I., Dellaportas, P., 2002. Bayesian modeling of outstanding liabilities incorporating claim count uncertainty. North American Actuarial Journal 6 (1), 113–128.
Paulson, A.S., Faris, N.J., 1985. A practical approach to measuring the distribution of total annual claims. In: Cumins, J.D. (Ed.), Strategic Planning and Modeling in Property-Liability Insurance. Kluwer Academic Publishers, Norwell, MA.
Ramlau-Hansen, H., 1988. A solvency study in non-life insurance. Part 1. Analysis of fire, windstorm, and glass claims. Scandinavian Actuarial Journal 3–34.
Smith, A.F.M., Roberts, G.O., 1993. Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. Journal of the Royal Statistical Society: Series B 55, 3–23.
Spiegelhalter, D., Best, N.G., Carlin, B.P., Van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B 64, 583–616 (with discussion).
Taylor, G.C., 2000. Loss Reserving: An Actuarial Perspective. Kluwer Academic Publishers, Boston.
Taylor, G.C., McGuire, G., 2004. Loss reserving with GLMs: a case study. In: Casualty Actuarial Society 2004 Discussion Paper Program, pp. 327–392.
Tong, H., 1978. On a threshold models. In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing. Sijthoff and Noordhoff, Amsterdam.

Verrall, R.J., 1989. A state space representation of the chain ladder linear model. Journal of the Institute of Actuaries 116, 589–609.
Verrall, R.J., 2000. An investigation into stochastic claims reserving models and the chain-ladder technique. Insurance: Mathematics and Economics 26 (1), 91–99.
Yang, X., Frees, E.W., Zhang, Z., 2011. A generalized beta copula with applications in modeling multivariate long-tailed data. Insurance: Mathematics and Economics 49 (2), 265–284.
Zehnwirth, B., 1994. Probabilistic development factor models with applications to loss reserve variability, prediction intervals, and risk-based capital. In: Casualty Actuarial Society Forum, 1994 Spring Forum. pp. 447–605.
Zhang, Y., Dukic, V., Guszcza, J., 2012. A Bayesian nonlinear model for forecasting insurance loss payments. Journal of the Royal Statistical Society A 175, 1–20.

Further reading

Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. In: Applications of Mathematics, Springer, Heidelberg.
Spiegelhalter, D., Thomas, A., Best, N., 2000. Bayesian inference using Gibs sampling for Windows version (WinBugs), software for Bayesian analysis using MCMC method and Gibbs sampler. http://www.mrc-bsu.cam.ac.uk/bugs/.