QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE
RESEARCH CENTRE

UTS

THINK.CHANGE.DO

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 196 **May 2007**

Robust Bayesian analysis of loss reserves data using the generalized-*t* distribution

Jennifer Chan, Boris Choy and Udi Makov

ISSN 1441-8010 www.qfrc.uts.edu.au

Robust Bayesian analysis of loss reserves data using the generalized-*t* distribution

By

Jennifer S.K. Chan
School of Mathematics and Statistics
The University of Sydney, Australia
Email: jchan@maths.usyd.edu.au

S.T. Boris Choy
Department of Mathematical Sciences
University of Technology Sydney, Australia
Email: boris.choy@uts.edu.au

and

Udi E. Makov
Department of Statistics, University of Haifa, Israel
Email: makov@stat.haifa.ac.il

Abstract: This paper presents a Bayesian approach using Markov chain Monte Carlo methods and the generalized-*t* (GT) distribution to predict loss reserves for the insurance companies. Existing models and methods cannot cope with irregular and extreme claims and hence do not offer an accurate prediction of loss reserves. To develop a more robust model for irregular claims, this paper extends the conventional normal error distribution to the GT distribution which nests several heavy-tailed distributions including the Student-*t* and exponential power distributions. It is shown that the GT distribution can be expressed as a scale mixture of uniforms (SMU) distribution which facilitates model implementation and detection of outliers by using mixing parameters. Different models for the mean function, including the log-ANOVA, log-ANCOVA, state space and threshold models, are adopted to analyze real loss reserves data. Finally, the best model is selected according to the deviance information criterion (DIC).

Keywords: Bayesian approach, state space model, threshold model, scale mixtures of uniform distribution, deviance information criterion.

1. Introduction

An insurance policy is a promise of an insurance company to pay claims to the insureds if some defined events (death, accident, injury, etc.) occur. However in many cases, claims originating in a particular year are often not settled in that year, but with a time delay of years or perhaps decades. Therefore, the insurance company must have the necessary loss reserves to pay these outstanding claims and settlement costs incurred. With many uncertainties in the time lags inherently involved in the claims settlement process, the estimation procedure of the required loss reserves is extremely complicated. Since loss reserves generally represent by far the largest liability and the greatest source of financial uncertainty in an insurance company, accurate prediction of the loss reserves is of great importance. Failure to estimate the loss reserves accurately may result in large profit losses and hence weaken the financial stability of the company which may ultimately drive it into insolvency.

Denote $Y_{i,j}$, $i, j = 1, \ldots, n$, the value of claims paid by an insurance company for accidents that occurred in policy year i and were settled after j-1 years (lag year j). The observed claim amounts $Y_{i,j}$, $i = 1, \ldots, n$, $j \le n - i + 1$ over a period of n policy years can be presented by a run-off triangle; see Table 1. The total number of observed claims in the upper triangle is $S_U = \frac{n(n+1)}{2}$ and the number of unobserved claims to be predicted in the lower triangle is $S_L = \frac{n(n-1)}{2}$. The aim is to predict the unobserved values in the lower triangle using an appropriate statistical model.

Table 1. Run-off triangle for claim data.

		Lag year *j*							
		1	*2*	*3*	.	.	.	*n - 1*	*n*
Policy year *i*	*1*	$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	.	.	.	$Y_{1,n-1}$	$Y_{1,n}$
	2	$Y_{2,1}$	$Y_{2,2}$	.	.	.	.	$Y_{2,n-1}$	
	3	$Y_{3,1}$	.	.	.	.	.		
	.	.	.	.	.	.			
	.	.	.	.	.				
	.	.	.	.					
	n - 1	$Y_{n-1,1}$	$Y_{n-1,2}$						
	N	$Y_{n,1}$							

The most popular method for prediction is the chain-ladder method (Renshaw, 1989) which uses loss ratio estimate and loss development factors. However it is increasingly apparent that this method lacks some measures of variability. Over the years, stochastic models with hierarchical structure have been developed to overcome such shortcomings. For example, Verrall (1991, 1996) and Renshaw and Verrall (1998) adopt the ANOVA-type and ANCOVA-type models. The interaction between the policy year and lag year prompted the development of the state space models which allow dynamic evolution of parameters in a time-recursive way. See Verrall (1989, 1994). In addition, Hazan and Makov (2001) propose the threshold models to allow for structural changes in the trend of outstanding claims over time. Analyses of real data suggest that the dynamic nature of the state space model and threshold model improves the prediction.

It is well known that the normal error distribution falls short of allowing for irregular and extreme claims and hence contaminates the estimation procedure and leads to poor prediction. To allow for irregular claims, error distributions that possess flexible tails are recommended. The Student-*t* distribution has been widely used by many researchers for this purpose while Choy and Chan (2003) use the exponential power (EP) family of distributions. In this paper, we shall adopt the generalized-*t* (GT) distribution (McDonald and Newey, 1988) for the errors. The GT distribution is symmetric and is governed by two shape parameters. By suitably choosing these shape parameter values, one can easily show that the normal, Cauchy, Student-*t*, EP and Laplace distributions are members of the GT family.

For a long time the Bayesian approach had very limited applications in statistical inference because the posterior functionals that involve high-dimensional integration cannot be obtained analytically. Gelfand and Smith (1990) develop the simulated-based Markov chain Monte Carlo (MCMC) techniques that transform the integration problem into a sampling problem. The idea is to construct a Markov chain whose limiting distribution is the intractable joint posterior distribution. The simulated realizations from the Markov chain mimic a random sample from the joint posterior distribution and the posterior functionals can be estimated from these realizations. Amongst the MCMC algorithms, Gibbs sampling (Geman and Geman, 1984), Metropolis-Hastings (Metropolis *et al.*, 1953 and Hastings, 1970) and Reversible Jump MCMC (Green, 1995) are very common. Bayesian statistical inferences can be easily done using WinBUGS (Spiegelhalter *et al.*, 2004), an easy-to-use software for MCMC algorithms. Readers who are interested in or unfamiliar to WinBUGS may find the educational materials on the official WinBUGS website very useful. See **www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml**. They are referred to Gilks *et al.* (1998) and Gelman *et al.* (2004) for MCMC applications in many statistical problems. For actuarial and insurance applications, see, for example, Makov (2001), De Alba (2002), Ntzoufras and Dellaportas (2002), Ntzoufras *et al*. (2005), Scollnik (1998, 2001, 2002) and Verrall (2004).

To simplify the computational procedure for the Bayesian calculation and to speed up the Gibbs sampling algorithm, the GT error distribution is expressed into a scale mixtures representation. The GT distribution can be expressed as a scale mixture of uniforms (SMU) distribution as all of its constituting families of distributions, the EP and Student-*t* distributions, can be expressed as SMU distributions. This paper makes the first attempt to adopt the GT error distribution via the SMU form for model implementation and to protect inferences from possible outliers that can be detected using the mixing parameters of the SMU representation.

The rest of this paper is structured as follows. Section 2 introduces the data that demonstrate the proposed loss reserve models. Section 3 reviews the models with normal error distribution. Section 4 outlines the GT distribution and its scale mixtures representations. Section 5 implements the models using the Bayesian approach and proposes the DIC for choosing the best model. Section 6 reports and compares the results. Finally, discussion of results and further research direction is outlined in Section 7.

2. The data

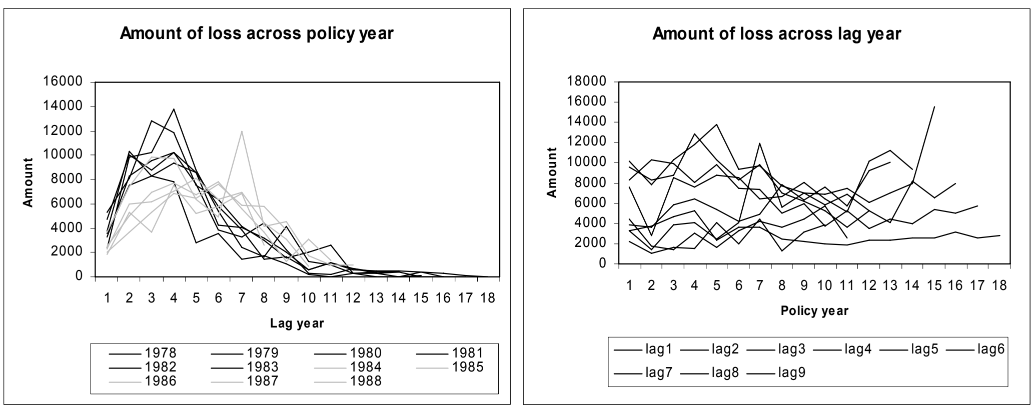
To demonstrate the models proposed in Sections 2 and 3, a loss reserve data set as shown in Table 2 is analyzed. The data are the amount of claims paid to the insureds of an insurance company during the period of 1978 to 1995 over $n = 18$ years. The upper triangle has $N = 171$ observations and the 153 observations in lower triangle are not yet observed. For mathematical convenience, two zero claim amounts are replaced by 0.01. Some general trends are obvious. Figure 1(a) shows that given a policy year, the amount of claims paid follows an increasing trend in the first 4 to 6 lag years and then a decreasing trend thereafter. On the other hand, Figure 1(b) shows no obvious trends for each lag year.

This data set contains some extreme outliers which are underlined in Table 2. For example, extremely large claims (in italic), amount to 11920 and 15546 dollars, were made in the 7-th lag year of policy year 1984 and in the 4-th lag year of policy year 1992 respectively. These outliers distort the general trends in the data and inflate the standard errors of the model parameters. Other outliers have much smaller amount of claims than their neighboring claims. These claims may lead to underestimates of loss reserves and hence lower the solvency and increase the risk of bankruptcy for an insurance company. Hence for robustness consideration, heavy-tailed error distributions are adopted to accommodate these irregular claims.

Table 2. The amount of claims paid to the insureds of an insurance during 1978 to 1995.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1978	3323	8332	9572	10172	7631	3855	3252	4433	2188	333	199	692	311	0.01	405	293	76	14
1979	3785	10342	8330	7849	2839	3577	1404	1721	1065	156	35	259	250	420	6	1	0.01	
1980	4677	9989	8746	10228	8572	5787	3855	1445	1612	626	1172	589	438	473	370	31		
1981	5288	8089	12839	11829	7560	6383	4118	3016	1575	1985	2645	266	38	45	115			
1982	2294	9869	10242	13808	8775	5419	2424	1597	4149	1296	917	295	428	359				
1983	3600	7514	8247	9327	8584	4245	4096	3216	2014	593	1188	691	368					
1984	3642	7394	9838	9733	6377	4884	11920	4188	4492	1760	944	921						
1985	2463	5033	6980	7722	6702	7834	5579	3622	1300	3069	1370							
1986	2267	5959	6175	7051	8102	6339	6978	4396	3107	903								
1987	2009	3700	5298	6885	6477	7570	5855	5751	3871									
1988	1860	5282	3640	7538	5157	5766	6862	2572										
1989	2331	3517	5310	6066	10149	9265	5262											
1990	2314	4487	4112	7000	11163	10057												
1991	2607	3952	8228	7895	9317													
1992	2595	5403	6579	15546														
1993	3155	4974	7961															
1994	2626	5704																
1995	2827																	

Figure 1 (a) Claim amounts across policy years. (b) Claim amounts across lag years.



3. Modeling the mean function

The popular log-linear model was investigated by Renshaw (1989), Verrall (1991, 1996) and Renshaw and Verrall (1998) in actuarial context. Depending on the formulation of policy year and lag year effects, we consider the log-ANOVA, log-ANCOVA and state space models for the mean function. For the state space model, the interaction effects between the policy-year and the lag-year are assumed. Each of these models is then allowed to switch to the model with a new set of parameters after certain thresholds. All models are implemented using WinBUGS with vague and non-informative prior distributions assigned to the model parameters.

3.1 ANOVA models

Within the Bayesian hierarchical modeling framework, the log-adjusted claim amount μ_{ij}, written in year i and paid with a delay j-1 years, adopts a two-way ANOVA model with normal error distribution where μ_{ij} is the mean function, α_i is the policy year effects and β_i is the lag year effects as below:

$$\log(Y_{ij}) = \mu_{ij} + \varepsilon_{ij} \tag{1}$$

$$\mu_{ij} = \mu + \alpha_i + \beta_j \tag{2}$$

$$\varepsilon_{ij} \sim N(0,\sigma^2) \quad (3)$$

for $i = 1, \ldots, n,\ j \le n - i + 1$ subject to the constraints $\sum_{i=1}^{n} \alpha_i = \sum_{j=1}^{n} \beta_j = 0.$ To complete the Bayesian framework, we adopt the following priors for the model parameters:

$$\mu \sim N(0,\sigma_\mu^2),\ \alpha_i \sim N(0,\sigma_\alpha^2),\ \beta_j \sim N(0,\sigma_\beta^2),\ \sigma^2 \sim IG(a,b)$$

where $IG(a,b)$ is the inverse gamma distribution with density

$$f(\sigma^2) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left(-\frac{b}{\sigma^2}\right).$$

Diffuse priors can be obtained by setting the hyperparameters $\sigma_\mu^2 = \sigma_\alpha^2 = \sigma_\beta^2 = \infty$ and $a = b = 0$, i.e. inflating the variances of the prior distributions to reflect the lack of prior information (Ntzoufras and Dellaportas, 2002). In this case, the joint prior density is

$$p(\mu,\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2) \propto \frac{1}{\sigma^2}$$

where $\boldsymbol{\alpha} = (\alpha_1,\ldots.\alpha_n)$ and $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_n)$.

3.2 ANCOVA model

In the ANCOVA model, the effects of policy year i and lag year j are linear. The three different combinations of effects are given below.

1. Linear effect of policy year and categorical effect of lag year,
2. Categorical effect of policy year and linear effect of lag year and
3. Linear effect of both policy year and lag year.

Preliminary result of the analysis reveals that the first combination provides the best fit to the data and hence is chosen for all subsequent analyses of the ANCOVA model. Analysis of the model follows from the log-ANOVA model from (1) to (3) except that the mean function becomes

$$\mu_{ij} = \mu + \alpha \cdot i + \beta_j \quad (4)$$

for $i = 1, \ldots, n,\ j \le n - i + 1$ subject to the constraints $\sum_{j=1}^{n} \beta_j = 0.$ The diffuse priors are chosen to be

$$p(\mu,\alpha,\boldsymbol{\beta},\sigma^2) \propto \frac{1}{\sigma^2}$$

where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_n)$.

3.3 State space model

To account for the interaction between the policy year and lag year in the mean function, the state space model (Ntzoufras and Dellaportas, 2002, De Jong and Zehnwirth, 1983 and Verrall, 1991, 1994) is considered. In the model, parameters are allowed to evolve in a time-recursive pattern: α_i depends on α_{i-1} and an error term h_i while β_{ij} depends on $\beta_{i-1,j}$ and an error term v_i. The model again follows from log-ANOVA model from (1) to (3) except that the mean function becomes

$$\mu_{ij} = \mu + \alpha_i + \beta_{ij} \tag{5}$$

for $i = 1, \ldots, n, \ j \le n - i + 1$ where the recursive associations are

$$\alpha_i = \alpha_{i-1} + h_i, \qquad h_i \sim N(0, \sigma_h^2), \qquad i = 2,3,\ldots,n \tag{6}$$

$$\beta_{ij} = \beta_{i-1,j} + v_i, \qquad v_i \sim N(0, \sigma_v^2), \qquad i,j = 2,3,\ldots,n \tag{7}$$

subject to the constraints $\alpha_1 = 0$ and $\beta_{i1} = 0$ for $i = 1,\ldots,n$. The priors for σ_h^2 and σ_v^2 follow $\sigma_h^2 \sim IG(a_h, b_h)$ and $\sigma_v^2 \sim IG(a_v, b_v)$ respectively. Diffuse prior distributions for the parameters are chosen to be

$$p(\mu, \beta_1, \sigma^2, \sigma_h^2, \sigma_v^2) \propto \frac{1}{\sigma^2 \sigma_h^2 \sigma_v^2}$$

where $\beta_1 = (\beta_{12}, \ldots, \beta_{1n})$.

3.4 Threshold model

Hazan and Makov (2001) suggested a switching regression model in which the mean functions before and after a threshold *T* along the axis of policy year are assigned the same model but with a different set of parameters. The reasoning behind is that some events such as a financial crisis or a change in the insurance regulation may take place on or before certain threshold year *T*. These events may change the effects of some factors on the claim amounts and hence a new set of parameters is adopted to reveal such changes. The threshold models based on different mean function structures are given below:

Threshold log-ANOVA model: $\mu_{ij} = \mu_1 + \alpha_{1i} + \beta_{1j}$ for $i < T$

$\mu_{ij} = \mu_2 + \alpha_{2i} + \beta_{2j}$ for $i \ge T$, $j \le n - T + 1$

Threshold log-ANCOVA model: $\mu_{ij} = \mu_1 + \alpha_1 \times i + \beta_{1j}$ for $i < T$

$\mu_{ij} = \mu_2 + \alpha_2 \times i + \beta_{2j}$ for $i \ge T$, $j \le n - T + 1$

Threshold state space model: $\mu_{ij} = \mu_1 + \alpha_{1i} + \beta_{1ij}$ for $i < T$ (8)

$\mu_{ij} = \mu_2 + \alpha_{2i} + \beta_{2ij}$ for $i \ge T$, $j \le n - T + 1$ (9)

The main difference in these threshold models is related to the set of β parameters which change after the threshold policy year *T*. On the contrary, the criterion that the α parameters are different before and after the threshold also holds for the simple models. Prior distributions for the parameters in the three models remain the same as those in Sections 3.1 to 3.3. The threshold *T* can be selected based on some model selection measures such as the Akaike information criterion (*AIC*) (Akaike, 1974), the Bayesian information criterion (*BIC*) (Schwarz, 1978), the Deviance information criterion (*DIC*) (Gelman *et al.*, 2004) and the posterior expected utility *U* (Walker and Gutiérrez-Peña, 1999). We will use the *DIC* in this paper because it is particularly useful in Bayesian model selection problems where the posterior distributions of the model parameters have been obtained by MCMC simulation. Threshold models with threshold lag-years, for example,

Threshold log-ANOVA model: $\mu_{ij} = \mu_1 + \alpha_{1i} + \beta_{1j}$ for $j < T$,

$\mu_{ij} = \mu_2 + \alpha_{2i} + \beta_{2j}$ for $j \ge T$, $i \le n - T + 1$, are also possible. However they are not considered further in this paper as the loss reserves data in the numerical illustration show no clear trends of claim amounts across policy years for each given lag-year (see Figure 1b).

4. Error distributions

Past experience has shown that irregular claims which often present in loss reserve data may seriously distort the parameter estimates and hence affect the accuracy of the prediction. The traditional log-linear models with normal errors fall short of allowing for the irregularities in the claim amounts. Heavy-tailed distributions such as the Cauchy, Student-*t*, Laplace and EP distributions are adopted for the errors in order to robustify the statistical inferences. However searching across these error distributions for the most suitable distribution, though workable, is time-consuming. The GT distribution, nesting all these competing heavy-tailed distributions as well as lighter-tailed alternatives including the popular normal and uniform distributions, provides a favorable choice for the error distribution (Butler *et al.*, 1990).

4.1 Generalized-*t* distributions

Proposed by McDonald and Newey (1988), the GT distribution is symmetric and unimodal and has a probability density function (PDF)

$$GT(x\,|\,\mu,\sigma,p,q) = \frac{p}{2q^{\frac{1}{p}}\sigma\,\mathrm{B}\left(\frac{1}{p},q\right)\left(1+\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^{p}\right)^{q+\frac{1}{p}}} \qquad (10)$$

where $\mu \in \Re$ is a location parameter, $\sigma > 0$ is a scale parameter, $p > 0$ and $q > 0$ are two shape parameters and $B(\cdot)$ is the beta function. When $\mu = 0$ and $\sigma = 1$, the r^{th} moment of the GT distribution is given by

$$E[X^r] = \frac{q^{\frac{r}{p}}\,\Gamma\left(\frac{r+1}{p}\right)\Gamma\left(q-\frac{r}{p}\right)}{\Gamma\left(\frac{1}{p}\right)\Gamma(q)} \qquad \text{provided that } r < pq.$$

In particular, the variance is given by

$$V[X] = \frac{q^{\frac{2}{p}}\,\Gamma\left(\frac{3}{p}\right)\Gamma\left(q-\frac{2}{p}\right)}{\Gamma\left(\frac{1}{p}\right)\Gamma(q)} \qquad \text{provided that } pq > 2.$$

The tail behavior and other characteristics are controlled by *p* and *q*. Large values of *p* and *q* signify distributions with thinner tails than the normal distribution whereas small values of *p* and *q* signify distributions with thicker tails. Hence the GT distribution can accommodate both leptokurtic and platykurtic distributions.

The GT family includes several important distributions: the normal $(p = 2, q \to \infty)$, the Cauchy $(p = 2,\ q = 1/2)$, the Student-*t* with $2q$ degrees of freedom $(p = 2)$, the Laplace $(p = 1, q \to \infty)$, the EP with shape parameter p $(q \to \infty)$, and the uniform $(p \to \infty)$ distributions. When $p = 2$, the PDF in (10) reduces to

$$t(x \mid \mu,\tau,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\tau\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{(x-\mu)^2}{\tau^2\nu}\right)^{\frac{\nu+1}{2}}}$$

which is the PDF of the Student-*t* distribution with degrees of freedom $\nu = 2q$ and scale parameter $\tau = \sigma/\sqrt{2}$. Moreover, (10) converges to the PDF of the EP distribution with shape parameter p

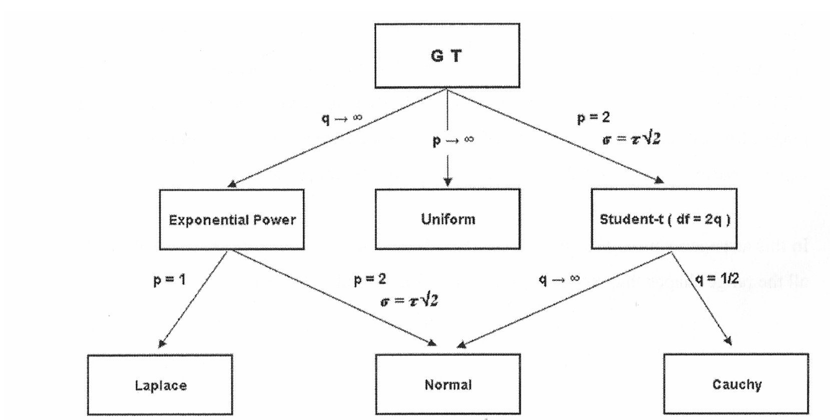
$$EP\,(x \mid \mu,\sigma,p) = \frac{\exp\left(-\left|\frac{x-\mu}{\sigma}\right|^p\right)}{2\sigma\,\Gamma\left(1+\frac{1}{p}\right)}$$

when $q \to \infty$. Lastly, (10) also converges to the PDF of the uniform distribution

$$U(x \mid \mu,\sigma) = \frac{1}{2\sigma} I(x \in (\mu-\sigma,\mu+\sigma))$$

when $p \to \infty$. In addition, when $p \le 1$, the GT distribution is cuspidate. Figure 2 summarizes the relationship amongst the well-known distributions within the GT family. Figure 9 (a)-(e) in Appendix B gives the PDFs for different values of p and q when $\mu = 0$ and $\sigma = 1$.

Figure 2: The GT distribution family tree.



4.2 The Generalized gamma distributions

The generalized gamma (GG) distribution with parameters $\alpha,\beta,\gamma > 0$, denoted by $GG(\alpha,\beta,\gamma)$, has a PDF given by

$$GG(x \mid \alpha,\beta,\gamma) = \frac{\gamma\beta^{\alpha\gamma}}{\Gamma(\alpha)} x^{\alpha\gamma-1} \exp\left(-\beta^{\gamma} x^{\gamma}\right)$$

where β is a scale parameter and α and γ are shape parameters. The s^{th} moment of the distribution is

$$E[X^s] = \frac{\Gamma\left(\alpha + s\gamma^{-1}\right)}{\beta^s \Gamma(\alpha)}.$$

The GG distribution includes the Weibull ($\alpha = 1$), gamma ($\gamma = 1$), exponential ($\alpha = \gamma = 1$) and lognormal ($\alpha = 0$) distributions as special cases. In particular, when $\gamma = 1$, the GG distribution reduces to the gamma $Ga(\alpha, \beta)$ distribution having the PDF

$$Ga(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} \exp(-\beta x),$$

the mean $E[X] = \dfrac{\alpha}{\beta}$ and the variance $V[X] = \dfrac{\alpha}{\beta^2}$. For inferential procedures for the GG distribution, see Hager and Bain (1970). See also McDonald and Butler (1987) for the applications in finance.

4.3 The scale mixtures representation of the GT distribution

Recently, Arslan and Genc (2003) showed that the GT distribution can be expressed into a scale mixtures of Box and Tiao distribution (Box and Tiao, 1962), or the scale mixtures of exponential power (SMEP) distribution because the Box and Tiao distribution is another name for the EP distribution. Using the SMEP density representation, the PDF in (10) can be rewritten as

$$GT(x \mid \mu, \sigma, p, q) = \int_0^\infty EP\left(x \middle| \mu, q^{1/p} s^{-1/2} \sigma, p\right) GG\left(s \middle| q, 1, \frac{p}{2}\right) ds. \quad (11)$$

The latent variable, s, arises in the above expression is called the mixing variable of the SMEP density representation and has a $GG(q, 1, p/2)$ distribution. Similar to the mixing variables of the scale mixtures of normal (SMN) distributions (Choy and Smith, 1997) and the scale mixtures of uniform (SMU) distributions (Choy and Chan, 2003), the mixing variable s in (11) can be used as a global diagnostic of possible outliers. Small mean value of s associates with possible outlier because the corresponding EP distribution in (11) must inflate its variance to accommodate the outlier. See Choy and Smith (1997) and Choy and Chan (2003) for outlier diagnostics using the mixing variable of the SMN and SMU distributions, respectively.

Many symmetric and non-normal distributions are extremely difficult to handle in statistical inference because of the non-conjugate structure of the models. In Bayesian computation, expressing the PDF of a distribution into a certain kind of scale mixtures form enables efficient simulation-based algorithms. Without using a scale mixtures representation, sampling non-standard full conditionals in Bayesian Gibbs sampling procedure relies on the Metropolis-Hastings algorithm, the slice sampler (Damien *et al.*, 1999) or other simulation algorithms. However, the Metropolis-Hastings algorithm and some other simulation algorithms require expertise in simulation techniques to provide reliable and efficient algorithms. The slice sampler is an easy-to-use algorithm that introduces an auxiliary variable to reduce a non-standard full conditional in a Gibbs sampler to standard full conditionals. However, there is no physical meaning for the auxiliary variable. On the contrary, the use of a scale mixtures density representation for a non-normal distribution results in adding extra latent variables in the model and hopefully produces a set of standard full conditionals for the Gibbs sampler. In addition, these latent variables are used to identify possible outliers. Therefore, the slice sampler and the use of scale mixtures density representation play different roles in Gibbs sampling even though they both use latent variables.

Although the GT distribution can be expressed as a SMEP distribution, the EP distribution is difficult to handle analytically in statistical inference because its density function involves an absolute term. Walker and Gutiérrez-Peña (1999) showed that the EP distribution can be expressed as a SMU distribution and Choy and Chan (2003) successfully adopt this SMU representation of the EP distribution in an insurance application. In this paper, we combine the SMEP for the GT distribution and the SMU for the EP distribution to obtain a SMU representation for the GT distribution. The resulting Gibbs sampler can be shown to have a set of easily sampled standard full conditional distributions.

Let U be a gamma $Ga\left(1+1/p,1\right)$ random variable where the PDF of the $Ga\,(a,b)$ distribution is

$$Ga\,(u\,|\,a,b)=\frac{b^a}{\Gamma(a)}u^{a-1}e^{-bu} \qquad u,a,b>0$$

and the mean and variance are $E[U]=\dfrac{a}{b}$ and $V[U]=\dfrac{a}{b^2}$, respectively, and S be a generalized gamma $GG\left(\dfrac{p}{2},1,q\right)$ random variable. Then, if X is a GT random variable with parameters μ,σ,p and $q,$ we have

$$X\,|\,U=u,S=s\sim Unif\left(\mu-q^{\frac{1}{p}}s^{-\frac{1}{2}}u^{\frac{1}{p}}\sigma,\ \mu+q^{\frac{1}{p}}s^{-\frac{1}{2}}u^{\frac{1}{p}}\sigma\right)$$

$$U\sim Ga\left(1+\frac{1}{p},1\right)$$

and

$$S\sim GG\left(q,1,\frac{p}{2}\right).$$

In other words, the PDF in (10) can be rewritten as

$$\int_0^\infty\int_\phi^\infty Unif\,(x\,|\,\mu-q^{\frac{1}{p}}s^{-\frac{1}{2}}u^{\frac{1}{p}}\sigma,\ \mu+q^{\frac{1}{p}}s^{-\frac{1}{2}}u^{\frac{1}{p}}\sigma)Ga\left(u\,|\,1+\frac{1}{p},1\right)GG\left(s\,|\,\frac{p}{2},1,q\right)du\;ds$$

where

$$\phi\;=\;\frac{1}{q}\left|\frac{x-\mu}{\sigma}\right|^p s^{\frac{p}{2}}.$$

The conditional variance of the above uniform distribution is $(qu)^{2/p}\sigma^2/(3s)$ which inflates with a large u and a small s. To identify potential outliers, we propose to use the ratio

$\psi = u^{1/p} s^{-1/2}$ because it controls the support and hence the variance of the uniform distribution.

Hence the sampling is performed with the observed data x and the mixing parameters u and s which are used for identification of outliers. Although the number of parameters to be sampled becomes larger with the inclusion of mixing parameters, sampling can be conducted more efficiently from standard full conditional distributions. Hence the SMU representation simplifies the computational procedures for Bayesian calculation and speeds up the Gibbs sampling algorithm in WinBUGS as designated.

5. Bayesian analysis

5.1 Model implementation

Bayesian approach is adopted for parameter estimation and is performed using WinBUGS. Command codes for the implementations are given in the Appendix A. Parameters are estimated from samples drawn from the posterior conditional distributions. Due to the complexity of the models, high posterior correlations exist between some parameters. These dependences may slow down the convergence rate in the Gibbs samplers for some parameters. As a result, the number of iterations M should be large enough to ensure that the sample is uncorrelated, large and stationary. We set M = 105,000 and the burn-in period is at least 5,000 iterations. After the burn-in period, parameters are taken from every 50$^{\text{th}}$ iteration to mimic a random sample of size at least K = 1,000 from the intractable joint posterior distribution. Trajectory plots and autocorrelation plots of the simulated values are used to check for the independence and convergence of the sample. The posterior sample means, or medians where appropriate, are reported as parameter estimates.

5.2 Model Selection

To choose the most appropriate Bayesian model for the loss reserves data, the Deviance information criterion (*DIC*) is used. The *DIC* is proposed by Spiegelhalter, *et al*. (2002) as a model selection method for complex hierarchical models in which the number of parameters is not clearly defined. The *DIC* is defined as

$$DIC = E[D(\theta)] + \Delta D(\hat{\theta})$$

where $E[D(\theta)]$ is a measure of the adequacy of model fitting and $\Delta D(\hat{\theta}) = E[D(\theta)] - D(\hat{\theta})$ estimates the effective number of parameters in the model. Based on the posterior sample, we calculate the DIC as

$$DIC = -2\left(\frac{2}{N} \sum_{k=1}^{1000} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \ln f_{ij}(z_{ij} \mid \boldsymbol{\theta}_k) - \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n-i+1} \ln f_{ij}(z_{ij} \mid \hat{\boldsymbol{\theta}}) \right)$$

where $z_{ij} = \ln y_{ij}$ is the log claim size, $f_{ij}(\cdot)$ is the GT density function given by (10) when the mean μ is replaced by μ_{ij} in (2), (4) and (5) for the log-ANOVA, log-ANCOVA and state space models respectively, $\boldsymbol{\theta}_k$ is a vector of all parameters in the k^{th} posterior sample and $\hat{\boldsymbol{\theta}}$ is a vector of posterior means or medians of all parameters. Obviously a model with the smallest DIC is the best model.

6. Results

6.1 Comparison between mean functions and error distributions

Various models with different mean functions and error distributions are fitted to the claim data. For each type of mean function, we set the two shape parameters, p and q, of the GT distribution to take different values, which signify different distributions that are nested within the GT family. In the simulation study, we set p = 50 to approximate the uniform distribution and q = 50 to approximate the EP distribution which includes the normal and Laplace distributions. From (10), these values for p and q approximate the limiting cases of $p \to \infty$ and $q \to \infty$ well. Moreover, more general families of distributions can be obtained by setting either p or q or both to be random. For example, $p = 1$ and q is random, and p is random and q = 0.5, 1 and 2, respectively, are used in the simulation study. Models are ranked according to *DIC*.

Table 3 exhibits the *DIC* values for a wide choice of error distributions for the log-ANOVA, log-ANCOVA and state space models. For the log-ANOVA model, the most appropriate error distribution is the GT distribution with $p = 1$ and q being random. The posterior mean of q is 2.46. For the log-ANCOVA and state space models, the GT error distribution with a random p and $q = 2$ is chosen and the posterior medians of p are 1.20 and 1.12, respectively. These p and q values correspond to the GT distributions that are heavier-tailed than the normal distribution. The *DIC* values for these three models are 336.0, 332.6 and 321.1, respectively and the state space model with a random p and $q = 2$ is preferred amongst the models studied.

Table 3. Goodness-of-fit measures for models with different mean functions and error distributions. Bold types values correspond to fixed parameter values.

Model	**ANOVA**				**ANCOVA**				**State space**			
Dist.	*p*	*q*	DIC	Rank	*p*	*q*	DIC	Rank	*p*	*q*	DIC	Rank
Normal	**2**	**50**	542.4	30	**2**	**50**	529.9	29	**2**	**50**	528.3	28
t	**2**	1.26	357.4	17	**2**	1.24	338.6	12	**2**	1.26	333.6	8
Cauchy	**2**	**0.5**	375.0	22	**2**	**0.5**	340.8	13	**2**	**0.5**	356.7	16
EP	1.24	**50**	430.4	27	1.24	**50**	416.9	25	1.24	**50**	418.3	26
Laplace	**1**	**50**	375.6	23	**1**	**50**	359.0	20	**1**	**5**	350.9	15
Uniform	**50**	**2**	582.1	31	**50**	**2**	617.2	32	**50**	**2**	800.4	33
GT q=0.5	3.75	**0.5**	383.2	24	3.54	**0.5**	357.9	19	3.52	**0.5**	374.0	21
GT q=1	2.01	**1**	357.5	18	1.90	**1**	333.4	6	1.93	**1**	332.3	4
GT q=2	1.13	**2**	345.0	14	1.20	**2**	332.6	5	1.12	**2**	321.1	1
GT p=1	**1**	2.46	336.0	10	**1**	3.09	333.5	7	**1**	2.49	321.9	2
GT	0.93	3.45	337.3	11	1.16	2.26	333.9	9	1.06	2.73	323.4	3

6.2 Comparison between simple and threshold models

To allow for structural changes in the trend of claim amount over time, threshold model that assumes a structural change at policy year T is suggested. For the loss reserve data used throughout this paper, state space model is shown to perform better than the log-ANOVA and log-ANCOVA models across a wide choice of error distributions in general. Therefore, the state space threshold model is used to re-analyze the claim data in this section. The threshold T is fixed within the range (1, n) for the policy year, not too close to either sides of the range. The mean function takes two different sets of parameters – one set before T and the other set on and after T. Table 4 exhibits the posterior medians and posterior standard deviations of p, q and *DIC* of various state space threshold models. Figure 3 displays the change of *DIC* versus

T. The state space model with threshold year T = 7 (year 1984) is shown to be the most appropriate model (DIC = 228.3). The posterior medians of p and q of the GT distribution are $\hat{p}$ = 0.81 and $\hat{q}$ = 2.63, respectively. This threshold model is chosen for the loss reserve prediction.

Table 4. Posterior medians of p and q (standard deviations in parentheses) and DIC for different state space threshold models.

Threshold	$T = 4$	$T = 5$	$T = 6$	$T = 7$	$T = 8$	$T = 9$
p	1.19 (0.16)	0.84 (0.10)	0.79 (0.07)	0.81 (0.16)	0.84 (0.08)	0.79 (0.10)
q	2.12 (0.45)	2.77 (0.48)	3.12 (0.64)	2.63 (0.61)	2.56 (0.47)	2.93 (0.58)
DIC	290.3	269.8	261.6	228.3	240.3	280.2

Figure 3. DIC versus threshold year, T for different state space threshold models.

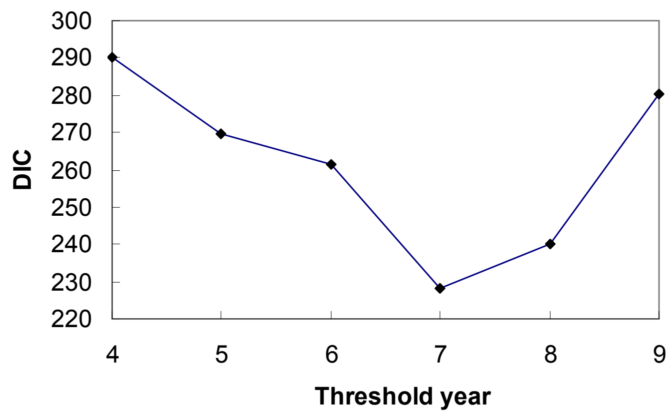
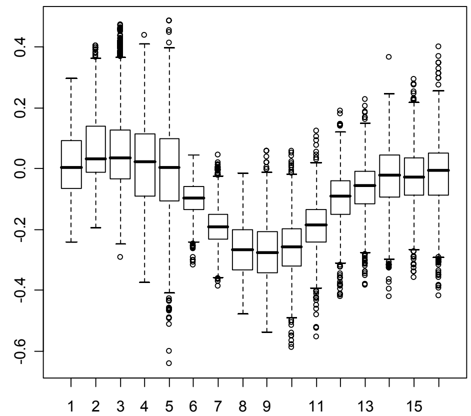


Figure 4 gives the box-plots of all model parameters in the posterior samples for the chosen model except the beta parameters, β_{ij}, which have too many parameters. The alpha and beta parameters in (8) and (9) are given by (6) and (7) respectively before and after the threshold year of 1984. The beta parameters for each policy year form a periodic trend across lag years: they increase up to lag years 4 to 6 and then decrease. Moreover the beta parameters at higher lag years have larger variability due to the insufficiency of observations to estimate the parameters accurately. Box-plots of alpha parameters also show trends of increasing variances across policy years before and after 1984 (the first five box-plots and the remaining box-plots respectively).

Figure 4 (a) Box-plots of alpha parameters, α_i. (b) Box-plots of mu parameters, μ_{ki}.



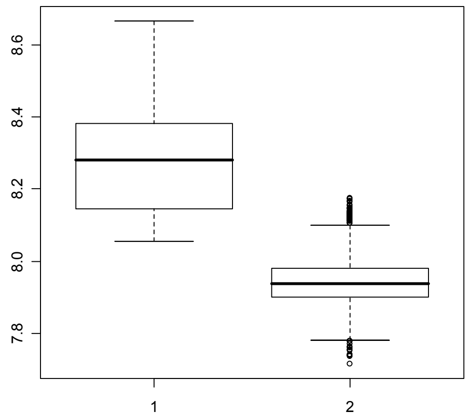
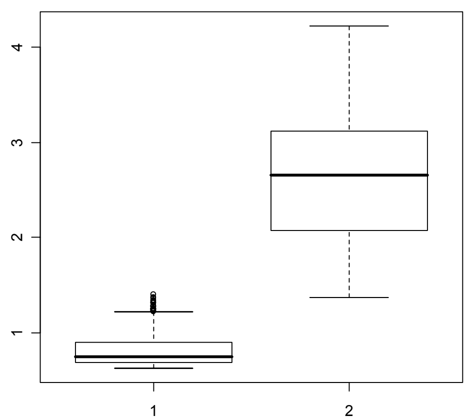


Figure 4 (c) Box-plots of shape parameters, p and q. (d) Box-plots of scale parameter, σ^2.



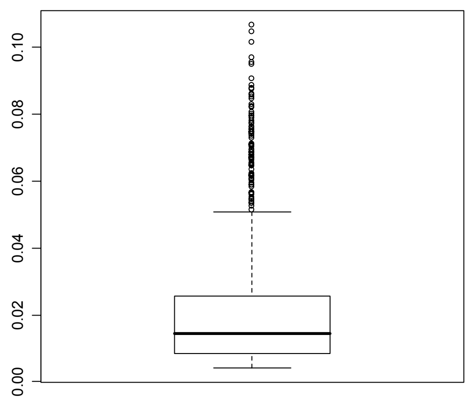


Table 6 reports the predicted claims in the upper triangle. Claims in italic refer to the predicted claims from policies written on and after 1984. Claims underlined are predicted from claims which are extremely low or high (in italic) in values as compared with neighboring claims.

The observed claims (Table 2) in solid line and the predicted claims in dotted line are plotted in Figure 5. The figure shows clearly two different trends of claims before and after 1984 as indicated by the black and grey lines respectively. For claims from policies written before 1984, the predicted claims increase with the lag year till the 4th lag year and drop slowly thereafter. For policies written on and after 1984, the predicted claims start from lower levels, rise to the lower maximums at the 6th lag year and drop more rapidly thereafter. In general, the predicted claims (dotted lines) lie close to the observed claims showing that the chosen model fits the data well.

Table 6. Run-off triangle of predicted and projected levels of claims using the chosen model.

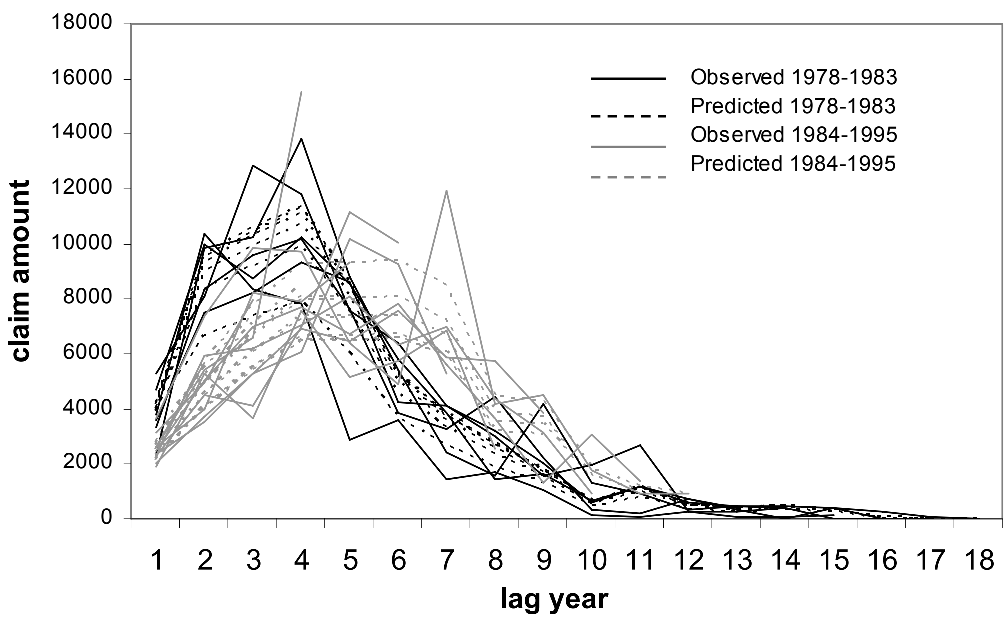
E(Y)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1978	3948	8373	9300	9881	7567	4606	3365	2387	1595	553	1013	443	342	397	332	31	58	14
1979	3964	6630	7364	7824	5991	3647	2665	1890	1263	438	802	351	271	314	263	24	46	11
1980	4082	9165	10179	10815	8282	5042	3684	2613	1746	605	1109	485	374	434	363	34	70	16
1981	4091	9517	10571	11231	8600	5236	3825	2713	1813	628	1152	504	388	451	377	35	70	16
1982	4032	9662	10732	11402	8731	5315	3883	2754	1840	638	1169	511	394	458	383	34	70	16
1983	3959	9388	10427	11079	8483	5164	3773	2676	1788	620	1136	497	383	432	363	32	68	15
1984	2802	5687	7723	9293	9284	9311	8434	4362	4377	1720	1162	927	243	283	237	22	41	10
1985	2542	4976	6757	8131	8122	8147	7379	3816	3829	1505	1016	800	209	242	200	20	34	9
1986	2315	4572	6209	7471	7464	7486	6781	3507	3519	1383	953	733	195	226	186	18	32	8
1987	2144	4148	5633	6777	6770	6791	6151	3181	3192	1238	844	663	174	201	168	16	29	7
1988	2124	4064	5519	6640	6633	6653	6026	3116	1244	1211	834	652	171	197	164	16	27	7
1989	2162	4104	5573	6706	6699	6719	6086	3074	4073	1231	845	646	171	198	164	16	29	7
1990	2328	4528	6148	7398	7390	7413	6641	3405	6203	1371	933	726	191	222	185	18	32	8
1991	2554	5047	6854	8247	8239	8060	7388	3834	7986	1520	1040	812	211	245	204	20	35	9
1992	2647	5269	7155	8609	8561	8475	7742	4016	8244	1588	1090	858	222	254	214	21	37	9
1993	2738	5473	7433	8814	8928	8875	8101	4211	8509	1651	1146	885	231	269	225	22	39	10
1994	2727	5506	7496	8924	8993	8938	8175	4241	7750	1661	1154	893	235	271	224	22	39	10
1995	2782	5546	7524	8938	9054	8997	8127	4279	4037	1687	1163	894	236	272	224	22	38	10

Values in italic are estimated or predicted claims from policies written after 1984.
Values in black and underlined are predicted claims for outliers (predicted claims from large outliers are in italic).
Values in grey and purple in the lower triangle are projected claims.
Values in underlined in the lower triangle are projected claims with standard errors higher than half of the values.
Values in purple are projected claims estimated using (12).

Table 7. Total projected outstanding claims for each policy year and their standard errors.

	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Total	11	86	131	515	928	868	1558	2450	3539	4729	10760	20295	31810	42016	52743	60016	61717
s.e.	69	161	210	229	254	219	11082	10828	8996	8697	8561	10202	12729	12688	12604	14927	17500

Figure 5. Observed and predicted claim amounts using the chosen model, the state space threshold model with a GT error distribution ($T = 7$ and p and q are random).



Projection for the outstanding claims in the lower triangle of Table 2 using the chosen model is reported in Table 6. The values are the posterior medians of the projected claims using the

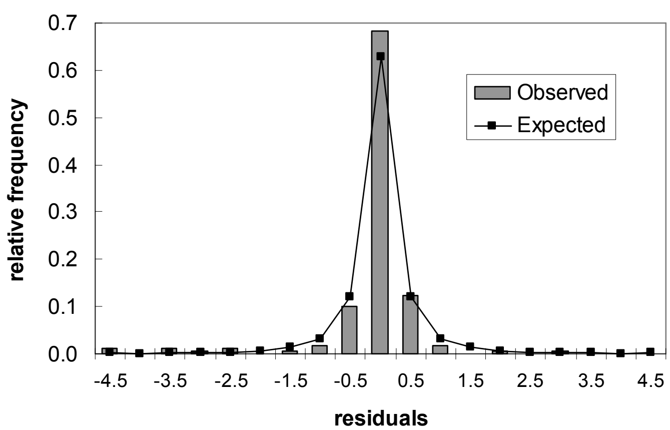
$K = 1{,}000$ sets of parameter estimates. The projected total outstanding claims, as the posterior median of the 1000 sums of projected claims, is 296,159 dollars with an estimated standard error of 123,867 dollars. The projected total outstanding claims for each policy year and their standard errors are reported in Table 7. Note that the projected claims $\hat{Y}_{ij},\ i \ge 7,\ j \ge 13$, written in purple in Table 6, are estimated using

$$\beta_{2,ij} = \beta_{1,i-1,j} + \nu_{2,i}, \quad i \ge 8,\ j \ge 13 \tag{12}$$

and $\beta_{2,7j} = \beta_{1,1j}$, $j \ge 13$ because their lag year effects $\beta_{2,ij}$ are not given by the model. Moreover it should be noted that some projected claims are underlined in Table 6 because their standard errors are more than half of the estimated values. As discussed above, the beta parameters for higher lag year higher variability because fewer observations are available to estimate these parameters accurately.

Figure 6 plots the observed relative frequencies and the expected probabilities using the density function (10) for the residuals $r_{ij} = \ln y_{ij} - \mu_{ij}$ where μ_{ij} is given by (8) and (9). As the observed relative frequencies are very close to the expected probabilities, the chosen model provides a very good fit to the loss reserves data. The mean, median, standard deviation, skewness and kurtosis for the residuals are -0.1655, -0.00715, 1.1306, -5.651 and 46.189 respectively. The residuals are quite leptokurtic and negatively skewed due to the existence of several extremely small outliers.

Figure 6. Observed and predicted relative frequencies for residuals using the chosen model.



6.3 Detection of Outliers

The SMU representation of the GT distribution not only simplifies the model implementation using the Bayesian approach but also allows detection of possible outliers using the parameter $\psi_{ij} = u_{ij}^{1/p} s_{ij}^{-1/2}$. An unusually large ψ_{ij} value indicates that the observed claim amount is a possible outlier. The posterior medians of ψ_{ij} for the chosen model are displayed in Figure 7. They identify observations 14, 29, 33, 34 and 35 (64 and 65 are marginal), labeled across a row from top to bottom in Table 2, to be possible outliers since their $\psi_{ij} > 10$ (ψ_{ij} for observations 64 and 65 are 8.94 and 8.70 respectively). The level of 10 is chosen to identify a moderate number of observations as outliers. Table 7 gives a summary of these outlying observations. These observations, grey in color and underlined in Table 2, correspond to the claim amounts which are substantially lower than their neighboring values. They may lead to an underestimation for the loss reserves and hence lower the solvency of an insurance company and increase its risk of bankruptcy. With ψ_{ij}, effects from these outliers are down-weighted and hence inference is protected.

Figure 7. Posterior medians of the parameters $\psi_{ij} = u_{ij}^{1/p} s_{ij}^{-1/2}$ in the chosen model.

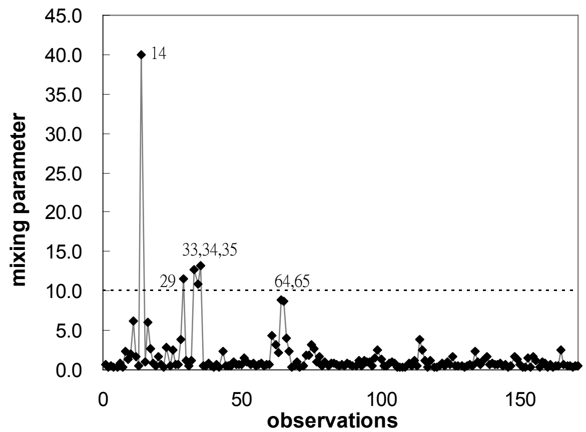


Figure 8 plots the observed verses predicted claims (in black for the chosen model). The plot shows that observations 100 and 165 (grey, italic and underlined in Table 2) are potential outliers as the observed claim amounts are much higher than their predicted values. However, their ψ_{ij} values are 1.3 and 2.5 respectively indicating that ψ_{ij} is less sensitive to large outliers. One possible explanation lies on the log-linear model: the log function is more sensitive to low values than large values. As effects from large outliers are less likely to be down-weighted by the mixing parameters in the SMU representation, loss reserves may be over-estimated. However the problem of over-estimating loss reserves may lower the profit of an insurance company but it may not weaken its solvency. Hence the risk of bankruptcy may not be seriously affected by the over-estimation.

Figure 8. Observed value y_{ij} versus predicted value $E(y_{ij})$ for the GT and chain ladder models.

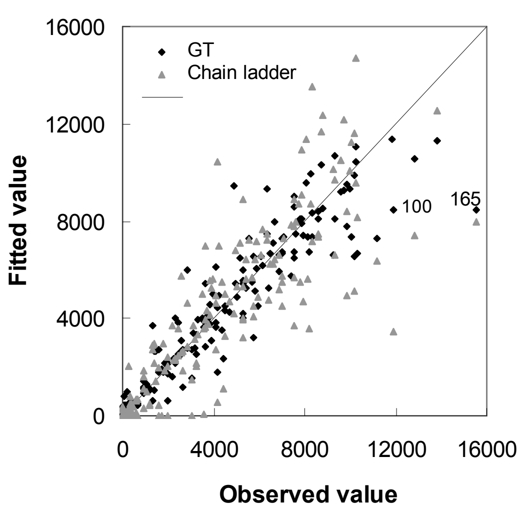


Table 8 gives a summary of the parameters ψ_{ij}, observed claims y_{ij} and expected claims $E(y_{ij})$. The posterior p-values p_{ij} is approximated using the following equation

$$p_{ij} = \begin{cases} 2F(\ln(y_{ij})), & r_{ij} < 0, \\ 2[1 - F(\ln(y_{ij}))], & r_{ij} \geq 0, \end{cases}$$

where $F(\cdot)$ is an approximated cumulative distribution function of the density function in (10) and μ_{ij} is given by (6) to (9). All except observations 100 and 165, written in italic in Table 8, give negative residuals, $r_{ij} = \ln y_{ij} - \mu_{ij}$. The posterior p-values p_{ij} indicate that these extreme observations, except observations 100 and 165, are outliers.

Table 8. Summary of the parameters ψ_{ij}, observed values y_{ij}, fitted values $E(y_{ij})$ and the posterior p-values p_{ij} of the extreme observations.

Outliers	Policy year i	Lag year j	Mix. par. ψ_{ij}	Observed y_{ij}	Fitted $E(y_{ij})$	Post. p-value p_{ij}
14	1978	14	40.05	0.01	392	0.0000
29	1979	11	11.55	35	798	0.0106
33	1979	15	12.69	6	244	0.0080
34	1979	16	10.81	1	29	0.0099
35	1979	17	13.25	0.01	3	0.0000
64	1981	13	8.94	38	391	0.0197
65	1981	14	8.70	45	450	0.0213
100	*1984*	*7*	*1.30*	*11920*	*8493*	*0.1712*
165	*1992*	*4*	*2.50*	*15546*	*8467*	*0.2595*

6.4 Comparison with chain ladder method

The chosen model produces more accurate prediction of claims than the popular chain ladder method which indirectly estimates the incremental loss y_{ij} by projecting the cumulative loss S_{ij}

$$\hat{S}_{ij} = \hat{S}_{i,j-1} \times f_j, \qquad i = 2, \ldots, n, \ \ j = n - i + 2, \ldots, n$$

where $S_{ij} = \sum_{k=1}^{j} y_{ij}$ and $\hat{S}_{i,n+1-i} = S_{i,n+1-i}$ using the projective factors

$$f_j = \frac{\sum_{i=1}^{n+1-j} S_{ij}}{\sum_{i=1}^{n+1-j} S_{i,j-1}}, \ \ j = 2, \ldots, n.$$

The prediction of S_{ij} on higher lag years, which is further away from the latest known claim amount requires the product of more projection factors and hence becomes less reliable as less data are available. We predict S_{ij} in the upper triangle of Table 2 using only one year ahead projective factor so that

$$\hat{S}_{ij} = S_{i,j-1} \times f_j, \qquad 2 \leq j \leq n - i + 1$$

and $\hat{S}_{i1} = S_{i1}$. We project S_{ij} in the lower triangle by

$$\hat{S}_{ij} = S_{i,n-i+1} \times \prod_{k=n-i+2}^{j} f_k, \quad i = 2,\ldots,18, \quad j > n-i+1.$$

The predicted and projected claims are reported in Table 9. Note that there are some negative values. They refer to predicted and projected claims which are very close to zero. The *MSE* for the predicted claims using the chain ladder method is 3,908,069 which is much larger than 1,584,478 for the chosen model. Moreover Figure 8 reveals that the predicted claims using the chosen model (black points) are closer to the observed claims than those using the chain ladder method (grey points). Thus the GT model provides more reliable prediction of loss reserves. Accurate prediction of loss reserves is very important to insurance companies since the levels of loss reserves have a dramatic impact on the profitability and solvency of insurance companies.

Table 9. Run-off triangle of predicted and projected levels of claims using the chain ladder method.

Y	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1978	3323	7160	10531	11618	7967	5582	2164	1086	3700	807	178	-326	489	264	-30	299	222	47
1979	3785	8156	13531	9052	4663	32	2180	-549	1017	-38	34	-372	101	214	399	-77	-54	-22
1980	4677	10078	11688	9548	7505	6450	4164	1446	442	141	470	632	372	389	443	256	-46	15
1981	5288	11394	7437	16182	8890	4769	4647	1505	2005	-61	1837	2048	13	-20	9	-15	52	17
1982	2294	4943	14693	12513	12380	6408	3594	-222	582	2703	1142	331	60	376	327	130	49	16
1983	3600	7757	8682	9725	7330	7008	2702	2067	2433	679	449	688	490	322	219	113	43	14
1984	3642	7848	8409	12191	7485	4186	3439	10430	3284	2954	1603	315	673	307	272	140	53	17
1985	2463	5307	5745	8738	6448	5493	7216	3923	2928	31	2984	883	287	242	214	111	42	13
1986	2267	4885	7680	7307	5599	7232	5494	5469	3729	1840	766	491	291	245	217	112	42	14
1987	2009	4329	3956	6627	6213	5700	7172	4487	5258	2727	937	464	275	232	205	106	40	13
1988	1860	4008	7009	3698	7113	4091	5188	5717	1953	910	803	398	236	199	176	91	34	11
1989	2331	5023	3190	6609	5117	10118	8915	3589	2274	1056	932	462	273	230	204	105	40	13
1990	2314	4986	4962	4514	6385	11242	9712	3150	2395	1112	981	486	288	243	215	111	42	14
1991	2607	5617	3602	10906	6596	8534	4821	3048	2317	1076	949	471	279	235	208	107	40	13
1992	2595	5592	6234	7988	16540	7136	5868	3710	2820	1310	1156	573	339	286	253	131	49	16
1993	3155	6798	4703	10078	7247	5860	4818	3046	2316	1075	949	470	278	235	208	107	40	13
1994	2626	5658	6735	8070	6765	5470	4498	2844	2162	1004	886	439	260	219	194	100	38	12
1995	2827	6092	7162	8640	7243	5856	4815	3044	2314	1075	948	470	278	234	208	107	40	13

7. Conclusion

In this paper, log-linear models with three different forms of mean function are used to model loss reserves. For the data set analyzed, the GT error distribution performs better than the normal and other well-know heavy-tailed error distributions such as Student-*t* and EP distributions. Meanwhile, the time-recursive state space model with threshold in 1984 provides the best fit to the data. Prediction and projection of loss reserves are based on Bayesian approach. Expressing the GT distribution into the uniform scale mixtures form makes Gibbs sampler more efficient and enables outlier diagnostics using the scale mixture mixing parameters. Model selection relies on the Deviance information criterion. Comparing with the popular chain ladder method, the chosen GT model is shown to provide more accurate predicted claims. The chosen model is then used to project outstanding claims of the lower triangle in Table 2.

As the log-linear model is more sensitive to low values than large values, the residuals for the data set analyzed are negatively skewed. To allow for a skewed error distribution, one may consider the skewed heavy-tailed distributions including the skewed-*t* distribution or the scale mixtures of beta distribution. Further investigation into the skewed error distributions will surely improve the modeling of loss reserve data.

Acknowledgement

The authors would like to thank Deborah Street and anonymous referees for very constructive comments, Connie P.Y. Lam for calculating the DICs and Wai-Yin Wan for running some WinBUGS programs. The second author is supported by an ECR grant of the University of Technology Sydney, Australia.

Reference

Akaike, H., *"A new look at the statistical model identification",* IEEE Transactions on Automatic Control, 19, 716–723, 1974.

Arslan, O. and Genc, A.I., *"Robust location and scale estimation based on the univariate generalized t (GT) distribution",* Communications in Statistics: Theory and Methods, 32, 1505-1525, 2003.

Box, G.E.P. and Tiao, G.C., "*A further look at robustness via Bayes' theorem",* Biometrika, 49, 419-432, 1962.

Butler R.J., McDonald B.J., Nelson R.D., and White S.B., *"Robust and Partially Adaptive Estimation of Regression Models"*, MIT Press, 72, 321-327, 1990.

Choy, S.T.B. and Chan, C.M., *"Scale mixtures distributions in insurance applications",* ASTIN Bulletin, 33, 93-104, 2003.

Choy, S.T.B. and Smith, A.F.M., *"Hierarchical models with scale mixtures of normal distribution"*, TEST, 6, 205-221, 1997.

De Alba, E., *"Bayesian estimation of outstanding claim reserves"*, North American Actuarial Journal, October 2002.

De Jong, P. and Zehnwirth, B., "*Claims reserving state space models and Kalman filter"*, The Journal of the Institute of Actuaries, 110, 157-181, 1983.

Damien, P., Wakefield, J. and Walker, S., *"Gibbs sampling for Bayesian non-Conjugate and hierarchical models by using auxiliary variables",* Journal of the Royal Statistical Society, Series B, 61, 331-344, 1999.

Gelfand, A.E. and Smith A.F.M., *"Sampling-based approaches to calculating marginal densities"*, Journal of American Statistical Association, 85, 398-409, 1990.

Geman, S. and Geman, D., *"Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741, 1984.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., *Bayesian Data Analysis***,** 2nd edition, Chapman & Hall/CRC, 182–184, 2004.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., *"Markov chain Monte Carlo in practice",* Boca Raton, Fla.: Chapman and Hall, 1998.

Green, P.J., *"Reversible jump Markoc chain Monte Carlo computation and Bayesian model determination",* Biometrika, 82, 711-732, 1995.

Hager, H.W. and Bain, L.J., *"Inferential procedures for the generalized gamma distribution"*, Journal of the American Statistical Association, 65, 1601-1609, 1970.

Hastings, W.K., "*Monte Carlo sampling methods using Markov Chains and their applications",* Biometrika, 57, 97-10, 1970.

Hazan, A. and Makov, U.E., *"A switching regression model for loss reserves"*, Technical report TR24, 2001.

Makov, U.E., *"Perspective applications of Bayesian methods in actuarial science: a perspective"*, North American Actuarial Journal, 2001.

McDonald, J.B. and Butler, R.J., *"Some generalized mixture distributions with an application to unemployment duration"*, Review of Economics and Statistics, 69, 232-240, 1987.

McDonald, J.B. and Newey, W.K., *"Partially adaptive estimation of regression models via the generalized t distribution"*, Economic Theory, 4, 428-457, 1988.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H., *"Equations of State calculations by fast computing machines"*, Journal of Chemical Physics, 21, 1087-1091, 1953.

Ntzoufras, I. and Dellaportas P., *"Bayesian modeling of outstanding liabilities incorporating claim count uncertainty",* North American Actuarial Journal, 6, 113-128, 2002.

Ntzoufras, I., Katsis, A. and Karlis, D., *"Bayesian assessment of the distribution of insurance claim counts using reversible jump MCMC",* North American Actuarial Journal, 9, 90-108, 2005.

Renshaw, A.E., *"Chain ladder and interactive modeling"*, The Journal of the Institute of Actuaries, 116, 559-587, 1989.

Renshaw, A.E., and Verrall, R.J., *"A stochastic model underlying the chain-ladder technique"*, British Actuarial Journal, 4, 903-926, 1998.

Schwarz, G., *"Estimating the dimension of a model",* Annals of Statistics 6(2), 461-464, 1978.

Scollnik, D.P.M., *"On the analysis of truncated generalized Poisson distribution using a Bayesian method",* ASTIN Bulletin, 28, 135-152, 1998.

Scollnik, D.P.M., *"Actuarial modeling with MCMC and BUGS"*, North American Actuarial Journal, 5, 96-124, 2001.

Scollnik, D.P.M., *"Implementation of four models for outstanding liabilities in WinBUGS: A discussion of a paper by Ntzoufras and Dellaportas"*, North American Actuarial Journal, 6, 128-136, 2002.

Spiegelhalter, D., Thomas, A., Best, N.G. and Lunn, D., *"Bayesian inference using Gibbs sampling for Window version (WinBUGS)"*, Version 1.4.1, MRC Biostatistics Unit, Institute of Public health, Cambridge, UK. (www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml), 2004.

Spiegelhalter, D., Best, N.G., Carlin, B.P. and van der Linde, A., *"Bayesian measures of model complexity and fit"*, Journal of the Royal Statistical Society, Series B, 64, 583-639, 2002.

Verrall, R.J., *"State space representation of the chain ladder linear model"*, The Journal of Institute of Actuaries, 116, 589-610, 1989.

Verrall, R.J., *"Chain ladder and maximum likelihood"*, The Journal of Institute of Actuaries, 118, 489-499, 1991.

Verrall, R.J., *"A method of modeling varying-off evolutions in claims reserving"*, ASTIN Bulletin, 24, 325-332, 1994.

Verrall, R.J., *"Claims reserving and generalized additive models"*, The Journal of Institute of Actuaries, 19, 31-43, 1996.

Verrall, R.J., *"A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving"*, North American Actuarial Journal, 8, 67-89, 2004.

Walker, S. G. and Gutiérrez-Peña, E., *"Robustifying Bayesian procedures"*, In: Bayesian Statistics 6. Oxford, New York, 685-710, 1999.

Appendix A WinBUGS command codes for M1 to M4

M1 log-ANOVA model

```
model {
for( i in 1 : N ) {
z[i] <- log(y[i])
ub[i] <- pow(abs(z[i]-mu[i]),p)*pow(s[i],b)/(q*pow(sigma,p))
u[i] ~ dgamma(a,1)I(ub[i],1000)
s[i] ~ gen.gamma(q,1, b)
z[i] ~ dunif(lower[i],upper[i])
lower[i] <- mu[i]-sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
upper[i] <- mu[i]+sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
mu[i] <- mu0+alpha[row[i]]+beta[col[i]]
}
for (j in 2:18){
alpha[j] ~ dnorm(0, 0.01)
beta[j]~ dnorm(0, 0.01)
}
alpha[1]<- 0-sum(alpha[2:18])
beta[1]<- 0-sum(beta[2:18])
mu0 ~ dnorm(0, 0.001)
tau2 ~ dgamma(0.001,0.001)
sigma2 <- 1/tau2
sigma <- pow(sigma2,0.5)
p ~ dgamma(0.001,0.001)       # when p is random
#p <- 2                       # when p is fixed at 2 say
q ~ dgamma(0.001,0.001)       # when q is random
#q <- 2                       # when q is fixed at 2 say
a <- 1+1/p
b <- p/2
}
```

M2 log-ANCOVA model

```
model {
for( i in 1 : N ) {
z[i] <- log(y[i])
ub[i] <- pow(abs(z[i]-mu[i]),p)*pow(s[i],b)/(q*pow(sigma,p))
u[i] ~ dgamma(a,1)I(ub[i],1000)
s[i] ~ gen.gamma(q,1, b)
z[i] ~ dunif(lower[i],upper[i])
lower[i] <- mu[i]-sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
upper[i] <- mu[i]+sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
mu[i] <- mu0+alpha*row[i]+beta[col[i]]
}
for (j in 2:18){
beta[j]~ dnorm(0, 0.01)
}
beta[1]<- 0-sum(beta[2:18])
alpha ~ dnorm(0, 0.01)
mu0 ~ dnorm(0, 0.001)
tau2 ~ dgamma(0.001,0.001)
sigma2 <- 1/tau2
sigma <- pow(sigma2,0.5)
p ~ dgamma(0.001,0.001)       # when p is random
#p <- 2                       # when p is fixed at 2 say
q ~ dgamma(0.001,0.001)       # when q is random
#q <- 2                       # when q is fixed at 2 say
a <- 1+1/p
b <- p/2
}
```

M3 State space model

```
model {
for( i in 1 : N ) {
z[i] <- log(y[i])
ub[i] <- pow(abs(z[i]-mu[i]),p)*pow(s[i],b)/(q*pow(sigma,p))
u[i] ~ dgamma(a,1)I(ub[i],1000)
s[i] ~ gen.gamma(q,1, b)I(0.01,1000)
z[i] ~ dunif(lower[i],upper[i])
lower[i] <- mu[i]-sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
upper[i] <- mu[i]+sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
mu[i] <- mu0+alpha[row[i]]+beta[row[i],col[i]]
}
for (i in 1:18) {
beta[i,1] <- 0
}
for (j in 2:18) {
beta[1,j] ~ dnorm(0, 0.01)
}
for (i in 2:18) {
for (j in 2:18) {
beta[i,j] <- beta[i-1,j]+v[i]
}
alpha[i] <- alpha[i-1]+h[i]
v[i] ~ dnorm(0, tau2v)
h[i] ~ dnorm(0, tau2h)
}
alpha[1] <- 0
mu0 ~ dnorm(0, 0.001)
tau2 ~ dgamma(0.001,0.001)
tau2v ~ dgamma(0.001,0.001)
tau2h ~ dgamma(0.001,0.001)
sigma2 <- 1/tau2
sigma2v <- 1/tau2v
sigma2h <- 1/tau2h
sigma <- pow(sigma2,0.5)
p ~ dgamma(0.001,0.001)       # when p is random
#p <- 2                       # when p is fixed at 2 say
q ~ dgamma(0.001,0.001)       # when q is random
#q <- 2                       # when q is fixed at 2 say
a <- 1+1/p
b <- p/2
}
```

M4 Threshold SS model

```
model {
# trend 1
for( i in 1 : N1-1 ) {
z[i] <- log(y[i])
ub[i] <- pow(abs(z[i]-mu[i]),p)*pow(s[i],b)/(q*pow(sigma,p))
u[i] ~ dgamma(a,1)I(ub[i],1000)
s[i] ~ gen.gamma(q,1, b)I(0.01,1000)
z[i] ~ dunif(lower[i],upper[i])
lower[i] <- mu[i]-sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
upper[i] <- mu[i]+sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
mu[i] <- mu01+alpha1[row[i]]+beta1[col[i]]
}
for (i in 1:T-1) {
beta1[i,1] <- 0
}
for (j in 2:18) {
beta1[1,j] ~ dnorm(0, 0.01)
beta1r1[j] <- beta1[1,j]
}
for (i in 2:T-1) {
for (j in 2:18) {
beta1[i,j] <- beta1[i-1,j]+v1[i]
}
alpha1[i] <- alpha1[i-1]+h1[i]
v1[i] ~ dnorm(0, tau2v1)
h1[i] ~ dnorm(0, tau2h1)
}
alpha1[1] <- 0
beta1r1[1] <- 0
# trend 2
for( i in N1 : N ) {
z[i] <- log(y[i])
ub[i] <- pow(abs(z[i]-mu[i]),p)*pow(s[i],b)/(q*pow(sigma,p))
u[i] ~ dgamma(a,1)I(ub[i],1000)
s[i] ~ gen.gamma(q,1, b)I(0.01,1000)
z[i] ~ dunif(lower[i],upper[i])
lower[i] <- mu[i]-sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
upper[i] <- mu[i]+sigma*pow(q,1/p)*pow(s[i],-1/2)*pow(u[i],1/p)
mu[i] <- mu02+alpha2[row[i]-T+1]+beta2[col[i]]
}
for (i in 1:18-T+1) {
beta2[i,1] <- 0
}
for (j in 2:18-T+1) {
beta2[1,j] ~ dnorm(0, 0.01)
beta2r1[j] <- beta2[1,j]
}
for (i in 2:18-T+1) {
for (j in 2:18-T+1) {
}
beta2[i,j] <- beta2[i-1,j]+v2[i]
}
alpha2[i] <- alpha2[i-1]+h2[i]
```

```
v2[i] ~ dnorm(0, tau2v2)
h2[i] ~ dnorm(0, tau2h2)
}
alpha2[1] <- 0
Beta2r1[1] <- 0
# end trend 2
mu01 ~ dnorm(0, 0.001)
mu02 ~ dnorm(0, 0.001)
tau2 ~ dgamma(0.001,0.001)
tau2v1 ~ dgamma(0.001,0.001)
tau2h1 ~ dgamma(0.001,0.001)
tau2v2 ~ dgamma(0.001,0.001)
tau2h2 ~ dgamma(0.001,0.001)
sigma2 <- 1/tau2
sigma2v1 <- 1/tau2v1
sigma2h1 <- 1/tau2h1
sigma2v2 <- 1/tau2v2
sigma2h2 <- 1/tau2h2
sigma2v1 <- 1/tau2v1
sigma <- pow(sigma2,0.5)
p ~ dgamma(0.001,0.001)      # when p is random
#p <- 2                      # when p is fixed at 2 say
q ~ dgamma(0.001,0.001)      # when q is random
#q <- 2                      # when q is fixed at 2 say
a <- 1+1/p
b <- p/2
}
```

Remark:

1. Data are entered using the following format:

```
y[] row[] col[]
3323   1  1
8332   1  2
...
2827   18  1
END
```

2. Values for some constants are:

```
list(N=171, T=7, N1=94)    # threshold model
list(N=171)                # simple model
```

For T=5,6,7,8,9, N1=67,81,94,106,117 respectively.

3. Starting values for the parameters are: 0.1 for mu01, mu02, 0.000001 for tau2, tau2h1, tau2v1, tau2h1, tau2v1, 0 for h1, h2, v1 and v2, all 1 for u and all 2 for s) for the threshold SS model.

4. Model parameters for the threshold SS model which are stored for calculation include:

mu01,mu02,v1,h1,v2,h2,alpha1,alpha2,beta1r1,beta2r1,sigma2h1,sigma2v1,sigma2h2,sigma2v2,u,s,p,q,sigma2.
Note that beta1r1 and beta2r1 store the beta in the first row for trend 1 and 2 respectively.
The lengths of alpha1,h1,v1 are T-1, the length of beta1r1 is 18 and alpha1[1]= beta1r1[1]=h1[1]=v1[1]=0.
The lengths of alpha2, beta2,h2,v2 are 18-T+1 and alpha2[1]= beta2r1[1]=h2[1]=v2[1]=0.
β_{ij} can be calculated using (7) for each k, beta1r1, beta2r1, v1 and v2.

Appendix B

Figure 9(a) – (e) Various density functions for the GT distribution.