MODELLING INSURANCE LOSSES USING CONTAMINATED GENERALISED BETA TYPE-II DISTRIBUTION

BY

J.S.K. CHAN, S.T.B. CHOY, U.E. MAKOV AND Z. LANDSMAN

ABSTRACT

The four-parameter distribution family, the generalised beta type-II (GB2), also known as the transformed beta distribution, has been proposed for modelling insurance losses. As special cases, this family nests many distributions with light and heavy tails, including the lognormal, gamma, Weibull, Burr and generalised gamma distributions. This paper extends the GB2 family to the *contaminated* GB2 family, which offers many flexible features, including bimodality and a wide range of skewness and kurtosis. Properties of the contaminated distribution are derived and evaluated in a simulation study and the suitability of the contaminated GB2 distribution for actuarial purposes is demonstrated through two real loss data sets. Analysis of tail quantiles for the data suggests large differences in extreme quantile estimates for different loss distribution assumptions, showing that the selection of appropriate distributions has a significant impact for insurance companies.

KEYWORDS

Loss distribution, generalised beta type-II distribution, contaminated mixture distribution, Bayesian MCMC, Value-at-Risk, tail conditional expectation, EM algorithm.

1. INTRODUCTION

To ensure an insurance company's financial stability, it is necessary to estimate accurately future claims liabilities. Reserving for the amount of future claims payments and evaluating premiums involves a large degree of uncertainty, especially for long tail class business where the claims often exhibit extreme severity with very different tail behaviours. Traditionally, conventional distributions, such as lognormal and gamma, are used to model claim severity (Taylor, 2000). In making these distributional assumptions, researchers may underestimate the risk manifested by the long right tail where large claims occur. This is because

 doi: 10.1017/asb.2017.37

most distributions do not possess flexible tails to describe the features of infrequent large claims.

To capture a wide range of population heterogeneity and tail behaviour, one practical way for practicing actuaries is to conduct analyses over subsets of claims with distinct claim characteristics. However, this approach falls short of providing a full picture of the claim dynamics. Alternatively, sophisticated distributions with flexible shapes and tail behaviours are proposed to model losses. These distributions include the generalised-t (GT) (McDonald and Newey, 1988), Pareto (Embrechts *et al.*, 1997), Stable family (Paulson and Faris, 1985), Pearson family (Aiuppa, 1988), log-gamma and lognormal (Ramlau-Hansen, 1988), Burr 12 (Cummins *et al.*, 1999) and GB2 (Venter, 1983; Kleiber and Kotz, 2003; Cummins *et al.*, 2007) distributions. While distributions on the real support, such as the GT, are flexible and nest important distributions like the exponential power, Student-t and uniform, that are both leptokurtic and platykurtic, they require log-transformation for claim data and the resulting log-GT distribution is negatively skewed and sensitive to small claims (Chan *et al.*, 2008).

There are several ways to deal with these problems. Brazauskas and Kleefeld (2011) propose the folded-t (FT) and log-FT distributions for insurance loss data which avoid the log transformation. In general, the probability density function (pdf) of a folded distribution is given by

$$f(y) = g(y) + g(-y), \tag{1}$$

where $g(\cdot)$ is a pdf on real support. Nadarajah and Bakar (2015) further propose a number of folded distributions including folded Gumbel, folded logistic, folded exponential power and folded GT, etc. Eling (2012) considers the log skew-normal and log skew-t distributions which offer flexible skewness. The GB2 family on a positive support is also popular as it can handle data with either positive or negative skewness. It includes both heavy and light-tailed distributions, such as gamma, Weibull, Pareto, Burr 12, lognormal and the Pearson family as special cases and hence provides convenient functional forms to model insurance claims (Cummins *et al.*, 1990, 1999, 2007; Dong and Chan, 2013). The past decade has seen a wide application of GB2 distributions in sophisticated modelling. Frees and Valdez (2008) and Frees *et al.* (2009) consider a three-component model for the number of claims, claim types and claim sizes, which are assigned negative binomial, multinominal logit and GB2 marginal distributions, respectively. In a more recent paper, Yang *et al.* (2011) adopt a multivariate GB2 copula model for bodily injury claims.

From a risk management perspective, most choices of distribution which are devoted to modelling the tail behaviour ignore the main body of the distribution. In the actuarial context, there are often smaller claims with high frequencies as well as larger claims with low frequencies. To tackle this problem, composite distribution based on two component densities $f_i(y),\ i = 1, 2$ and threshold θ expressed as

$$f(y) = \begin{cases} cf_1(y|\phi_1) & \text{if } 0 \le y \le \theta \\ cf_2(y|\phi_2) & \text{if } \theta \le y \le \infty \end{cases} \tag{2}$$

has been proposed where c is a normalising constant. The parameters (ϕ_1, ϕ_2) are subject to two constraints

$$f_1(\theta) = f_2(\theta) \quad \text{and} \quad f_1'(\theta) = f_2'(\theta) \tag{3}$$

to ensure continuity and differentiability at the threshold. Cooray and Ananda (2005) first derive the lognormal-Pareto (LP) composite distribution, consisting of a truncated lognormal before the threshold and a shifted Pareto thereafter. Scollnik (2007) extends the model to unrestricted mixing weights and generalised Pareto distribution. Then Scollnik and Sun (2012) consider Weibull–Pareto (WP) composite distribution and Nadarajah and Baker (2014) adopt logistic mixing weights for the composite lognormal-Burr distribution.

In contrast, our interest focuses on the underlying heterogenous groups of small and large claims. In this regard, a mixture distribution defined as

$$f(y) = (1-\pi) f_1(y) + \pi f_2(y), \tag{4}$$

where π is the unknown mixture weight and $f_i(y)$ are the mixture component densities, is used to capture these versatile features. To model the small proportion of large claims, $f_2(y)$ can adopt an inflated scale. This mixture distribution can be regarded as a generalisation of the folded distribution in (1) when the weights are equal and $f_2(y) = g(-y)$ corresponds to a shift of location instead of an inflated scale for the second component. Comparing the composite and mixture distributions, the components in composite distributions are defined over restricted ranges, before and after the threshold (Appendix G), whereas the components in the mixture distribution are defined over the entire range. Both distributions have pros and cons in modelling losses. The problems of continuity and differentiability of the pdf at the threshold have greatly lessened the flexibility of composite distributions (Appendix G) and complicated the pdf formulation. On the other hand, mixture distributions have a more appealing model interpretation and enable classification of losses based on latent group membership estimates. Although the derivatives of the log-likelihood function based on (4) is complicated in the classical likelihood approach, the expectation-maximisation (EM) and Bayesian MCMC techniques provide efficient ways to tackle the problem.

To choose a flexible distribution for the mixture components, McDonald and Butler (1987) consider the GB2 distribution and demonstrate how a mixture of GB2 distributions can be applied to model the heterogeneous unemployment duration data. Dong and Chan (2013) propose a two-group GB2 mixture distribution in which the location and shape parameters vary across subgroups of small and large claims. In the spirit of Tukey (1960), Landsman and Makov (2003) introduce the contaminated exponential dispersion loss model (CEDLM) as a mixture of two exponential dispersion distributions with identical canonical parameter but different dispersion parameters. All these mixture distributions can handle both small and extremely large claims.

This paper makes the first attempt to extend the GB2 distribution to contaminated GB2 (CGB2) distribution with two additional parameters which control

the degree of contamination (π) and the contaminating dispersion (k). The benefit of adopting the CGB2 distribution as a loss distribution is that it offers many flexible features including bimodality and a wide range of skewness and kurtosis. Furthermore, being a mixture distribution, the CGB2 distribution classifies large and small claims into latent subgroups with distinct characteristics and allows one to reveal classification through the estimated latent group-membership indicators. The assumption of group-specific scale parameters but consistent shape parameters are natural as experience shows that shape parameters which are related to the skewness and kurtosis of a distribution are often more difficult to estimate and require more sample information.

We formulate some important features of the CGB2 distribution, including the mean, variance, skewness and kurtosis with closed-form expressions, and study their trends across the contaminating dispersion parameter k. We also deduce an expression for the mode of the distribution if it exists. We further derive the Value at Risk (VaR) and Tail Conditional Expectation (TCE) risk measures (Jorion, 1997; Artzner *et al.*, 1999; Landsman and Valdez, 2005). These measures are useful for assessing model performance. In the empirical data analysis, we consider the two fire loss data sets discussed in Cummins *et al.* (1990) and compare the proposed contaminated mixture distributions with two folded and two composite distributions. Parameters are estimated using the Bayesian approach via the user-friendly Bayesian softwares `WinBUGS` and `OpenBUGS`. To estimate loss reserves, VaR and TCE are evaluated from the posterior predictive distribution using the Bayesian approach and our derived formulae for VaR and TCE also allow their evaluation in the likelihood approach.

The rest of this paper is organised as follows. In Section 2, we introduce the GB2 distribution and extend it to the CGB2 family. Properties of the CGB2 distribution are also derived. Section 3 describes model implementation using the Bayesian approach and proposes the log-likelihood and deviance information criterion (DIC) measures for assessing model fit. A simulation study assessing the estimation efficiency is conducted in Section 4. In Section 5, the proposed distributions are applied to fit two real loss data sets and the VaR and TCE of the best fitted distributions are evaluated. Lastly, Section 6 provides some final remarks and a conclusion of this study.

2. Generalised distributions

2.1. Generalised beta type-II

The GB2 distribution is adopted for a wide range of applications (McDonald, 1984) as it nests a number of important distributions as special cases. It favours, in particular, the modelling of loss reserve data which often exhibits heavy-tailed behaviour from long tail business classes. Venter (1983) introduces the GB2 distribution in the actuarial literature and called it *transformed beta distribution*. The GB2 distribution consists of four parameters and its pdf is specified

as follows:

$$f_{\mathrm{GB2}}(y|a,b,p,q)=\frac{\frac{|a|}{b}(\frac{y}{b})^{ap-1}}{B(p,q)[1+(\frac{y}{b})^{a}]^{p+q}},\quad \text{for } y>0, \tag{5}$$

where b is a scale parameter and a, p and q are the shape parameters with b, p and $q>0$. $B(p,q)$ represents the beta function defined by

$$B(p,q)=\int_0^1 t^{p-1}(1-t)^{q-1}dt=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \tag{6}$$

and $\Gamma(\cdot)$ denotes the gamma function. The distribution can be obtained from the beta distribution with a pdf

$$f_B(u|p,q)=\frac{1}{B(p,q)}u^{p-1}(1-u)^{p+q}, \tag{7}$$

using the transformation

$$u=\frac{(\frac{y}{b})^{a}}{1+(\frac{y}{b})^{a}}\equiv\frac{e^{z}}{1+e^{z}}, \tag{8}$$

where $z=(\ln y-\ln b)/a^{-1}$ has a pdf proportional to $\dfrac{e^{zp}}{(1+e^{z})^{p+q}}$. Hence, the cumulative distribution function (cdf) of the GB2 distribution is

$$F_{\mathrm{GB2}}(y|a,b,p,q)=\frac{\int_0^u t^{p-1}(1-t)^{(q-1)}\,dt}{B(p,q)}=\frac{B_i(u,p,q)}{B(p,q)}=F_B(u|p,q), \tag{9}$$

where $B_i(x,p,q)$ is the incomplete beta function and $F_B(\cdot)$ is the cdf of the beta distributions. Generally, p and q determine the skewness of the distribution and negative values of a yield inverse distribution (Cummins *et al.*, 1990). The mode of the GB2 distribution is given by

$$y_m=b\left(\frac{ap-1}{aq+1}\right)^{\frac{1}{a}}=bg^{\frac{1}{a}}, \tag{10}$$

if $ap>1$ and the moments are given by

$$E(Y^h)=\frac{b^h B(p+h/a,q-h/a)}{B(p,q)}=b^h c_h, \tag{11}$$

if $-ap<h<aq$, where

$$c_h=\frac{B(p+\frac{h}{a},q-\frac{h}{a})}{B(p,q)}=\frac{\Gamma(p+\frac{h}{a})\Gamma(q-\frac{h}{a})}{\Gamma(p)\Gamma(q)}$$

(Klugman *et al.* 2008, p. 669). Being a transformed beta random variable, a GB2 distributed random variable Y can be simulated using

$$Y = b\left(\frac{U}{1-U}\right)^{1/a}, \quad \text{where} \quad U \sim \text{beta}(p, q),$$

beta distribution with parameters p and q and mean $p/(p+q)$. While the general moments of the GB2 distribution in (11) are given in many papers, their mean μ_0, variance σ_0^2, skewness γ_0 and excess kurtosis κ_0 are not explicitly expressed. These four moments are often used to describe the characteristic of a distribution. Based on (11), we derive these expressions as follows:

$$\mu_0 = E(Y) = bc_1, \tag{12}$$

$$\sigma_0^2 = E[(Y-\mu)^2] = b^2(c_2 - c_1^2), \tag{13}$$

$$\gamma_0 = \frac{E[(Y-\mu)]^3}{E[(Y-\mu)^2]^{3/2}} = \frac{c_3 - 3c_2c_1 + 2c_1^3}{(c_2 - c_1^2)^{3/2}}, \tag{14}$$

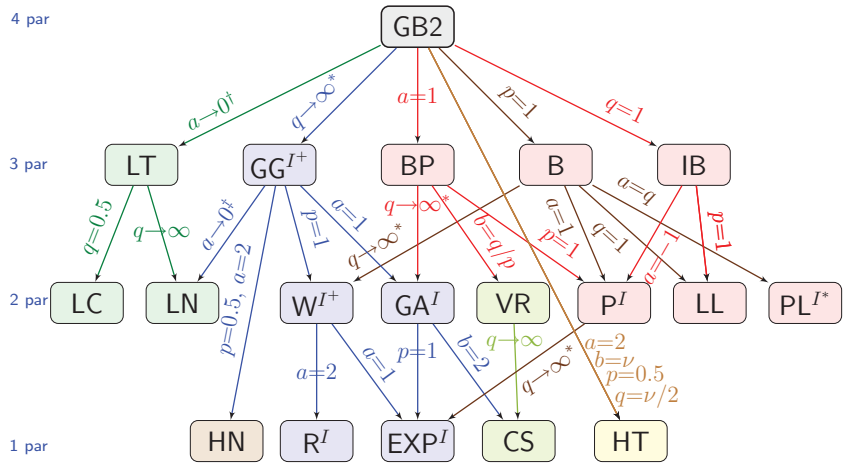
$$\kappa_0 = \frac{E[(Y-\mu)]^4}{E[(Y-\mu)^2]^2} - 3 = \frac{c_4 - 4c_3c_1 + 6c_2c_1^2 - 3c_1^4}{(c_2 - c_1^2)^2} - 3, \tag{15}$$

if $-ap < 4 < aq$, where

$$c_h \simeq \frac{\left(p - 1 + \frac{h}{a}\right)^{p-\frac{1}{2}+\frac{h}{a}} \left(q - 1 - \frac{h}{a}\right)^{q-\frac{1}{2}-\frac{h}{a}}}{(p-1)^{p-\frac{1}{2}}\,(q-1)^{q-\frac{1}{2}}}, \tag{16}$$

using Stirling's approximation $\Gamma(h+1) \simeq \sqrt{2\pi h}(h/e)^h$ (Appendix A). Parameter estimates can be derived by equating (12) to (15) to their sample estimates and solving for the four model parameters. We note that the skewness and kurtosis increase when aq approaches 4 from above. The GB2 distribution nests many popular distributions as summarised in Figure 1. Klugman *et al.* (2008; p. 669–681) provide a clear description of many of these nested distributions, and Table B1 in Appendix B gives a summary of the names and the abbreviations for the distributions in Figure 1, the parameters mapping and pdfs.

As indicated by the colours, the nested distributions can be divided into several sub-groups according to the definition of scale parameter. The red group has the scale parameter b defined consistently with the GB2 distribution (called consistent-scale distribution in Section 5) and any shape parameters a, p, q set to 1. An inverse distribution is suggested if p is large relative to q and log distribution (green group) should be considered if a approaches 0 (Cummins *et al.*, 1990). To quantify risk at different quantile levels, we derive the VaR and TCE risk measures. Based on the cdf in (9), the quantile function for the VaR is given



*$q\to\infty$ with $b=q^{1/a}\beta$ (for GB2 → GG, B → W, BP → GA, P → EXP).
†$a\to 0$ with $b=(q\sigma^2a^2)^{1/a}$, $p=(a\mu+1)/\sigma^2a^2$ (for GB2 → LT).
‡$a\to 0$ with $b=(\sigma^2a^2)^{1/a}$, $p=(a\mu+1)/\sigma^2a^2$ (for GG → LN).
I If the sign of a changes, inverse distribution is obtained (for GA, P, R, EXP).
I^+ If $p\to\infty$ with $b=p^{1/a}\beta$, inverse distribution is obtained (for GG, W).
I^* If $a=p$ and $q=1$, inverse distribution is obtained (for PL → IPL).

FIGURE 1: Family of GB2 distribution. (Color online)

by

$$\mathrm{VaR}_{\mathrm{GB2},\rho} \overset{\mathrm{def}}{=} b\left(\frac{F_B^{-1}(\rho|p,q)}{1-F_B^{-1}(\rho|p,q)}\right)^{\frac{1}{a}},$$

where $\rho\in(0,1)$ denotes the quantile level. Another risk measure is the TCE defined as the upper TCE (see details in Landsman and Valdez, 2005):

$$\mathrm{TCE}_{\mathrm{GB2},\rho} \overset{\mathrm{def}}{=} \frac{b\,B(p+\frac{1}{a},q-\frac{1}{a})}{B(p,q)}\times\frac{1-F_B\left(F_B^{-1}(\rho|p,q)\,\middle|\,p+\frac{1}{a},q-\frac{1}{a}\right)}{1-\rho},$$

if $-ap<1<aq$ (Appendix C). Clearly, $\mathrm{TCE}_{\mathrm{GB2},\rho}$ tends to $E(Y)$ as ρ tends to 0.

2.2. Contaminated generalised beta type-II distribution

The CGB2 distribution has the pdf

$$f_{\mathrm{CGB2}}(y|a,b,p,q,\pi,k)=(1-\pi)\,f_{\mathrm{GB2}}(y|a,b,p,q)+\pi\,f_{\mathrm{GB2}}(y|a,kb,p,q), \tag{17}$$

where $f_{\mathrm{GB2}}(\cdot)$ is given by (5), the parameter for the degree of contamination, $\pi\in(0,0.5)$ and the contaminating dispersion parameter is set such that $k>1$

to avoid identifiability problem. The parameter k indicates the amount of inflated dispersion for the contaminated mixture component. As a special case, the CGB2 distribution reduces to GB2 distribution when $k = 1$. The cdf of CGB2 distribution is

$$\begin{aligned} F_{\text{CGB2}}(y|a,b,p,q,\pi,k) &= (1-\pi)\int_0^y f_{\text{GB2}}(t|a,b,p,q)\,dt \\ &\quad + \pi\int_0^y f_{\text{GB2}}(t|a,kb,p,q)\,dt \\ &= (1-\pi)F_B(u|p,q) + \pi F_B(v|p,q), \end{aligned}$$

where v is given by (8) when b is replaced by kb. As shown in Table B1, some members of the GB2 family such as GG (M5), LT (M6) and VR (M7) have re-parameterised scale parameter

$$b = \beta q^{\frac{1}{a}}, \quad b = (a\sigma)^{\frac{2}{a}} q^{\frac{1}{a}} \quad \text{and} \quad b = q/p, \tag{18}$$

respectively. Their contaminated scale parameters are given by

$$k\beta, \quad k\sigma \quad \text{and} \quad kq, \tag{19}$$

respectively, as indicated in the footnote of Table B1.

In revealing features of the CGB2 distribution, we first show that the mode can be obtained by solving

$$\pi(k^a + y^a)^{p+q+1}[g - y^a] + (1-\pi)k^{aq}(1+y^a)^{p+q+1}(gk^a - y^a) = 0 \tag{20}$$

for y when setting $b = 1$ for standardised distributions and $g = \dfrac{ap-1}{aq+1}$ (Appendix D). Based on (11), the moments of the CGB2 distribution are given by

$$E(Y^h) = (1 - \pi + \pi k^h)b^h \frac{\Gamma(p + h/a)\Gamma(q - h/a)}{\Gamma(p)\Gamma(q)}, \tag{21}$$

provided that $-ap < h < aq$. Using (16), these moments can be alternatively approximated by

$$E(Y^h) \simeq (1 - \pi + \pi k^h)b^h \frac{\left(p - 1 + \frac{h}{a}\right)^{p-\frac{1}{2}+\frac{h}{a}} \left(q - 1 - \frac{h}{a}\right)^{q-\frac{1}{2}-\frac{h}{a}}}{(p-1)^{p-\frac{1}{2}}\,(q-1)^{q-\frac{1}{2}}}. \tag{22}$$

Hence, making use of (22) and (12) to (15), the mean μ, variance σ^2, skewness γ and excess kurtosis κ are, respectively,

$$\mu = (1 - \pi + \pi k)bc_1, \tag{23}$$

$$\sigma^2 = (1 - \pi + \pi k^2)b^2 c_2 - (1 - \pi + \pi k)^2 b^2 c_1^2, \tag{24}$$

$$\gamma = \frac{(1-\pi+\pi k^3)c_3 - 3(1-\pi+\pi k)(1-\pi+\pi k^2)c_1c_2 + 2(1-\pi+\pi k)^3c_1^3}{[(1-\pi+\pi k^2)c_2 - (1-\pi+\pi k)^2c_1^2]^{3/2}}, \tag{25}$$

$$\kappa = \frac{(1-\pi+\pi k^4)c_4 - 4(1-\pi+\pi k)(1-\pi+\pi k^3)c_1c_3}{[(1-\pi+\pi k^2)c_2 - (1-\pi+\pi k)^2c_1^2]^2} + \frac{6(1-\pi+\pi k)^2(1-\pi+\pi k^2)c_1^2c_2 - 3(1-\pi+\pi k)^4c_1^4}{[(1-\pi+\pi k^2)c_2 - (1-\pi+\pi k)^2c_1^2]^2} - 3, \tag{26}$$

for $-ap < 4 < aq$. To show how the parameters affect the shape of the CGB2 distribution, Figure 2 plots the pdf when one parameter varies, keeping other parameters fixed. The modes for the uncontaminated and contaminated mixture components are indicated by vertical dotted lines. The plots show that the distribution is more likely to be bimodal as a, p, q and π increases.

Table 1 reports the modes, mean, variance, skewness and kurtosis of the CGB2 distribution. From the table, both γ and κ increase when each of a, p and q decreases while other parameters remain unchanged. In fact, from (21), γ and κ increase when aq approaches four from above. Obviously, μ and σ^2 tends to infinity as k goes to infinity. This is in contrast to the CEDLM of Landsman and Makov (2003), where the mean μ remains constant across k as the canonical parameter that determines the mean does not change with k. However, we note that in loss data, the smaller proportion of large claims should have larger mean and hence the overall mean of the distribution is often shifted to the right. The limits for γ and κ when k tends to infinity are

$$\gamma_l = \lim_{k\to\infty}\gamma = \frac{c_3 - 3\pi c_1c_2 + 2\pi^2c_1^3}{\pi^{1/2}(c_2 - \pi c_1^2)^{3/2}}, \tag{27}$$

$$\kappa_l = \lim_{k\to\infty}\kappa = \frac{c_4 - 4\pi c_1c_3 + 6\pi^2c_1^2c_2 - 3\pi^3c_1^4}{\pi(c_2 - \pi c_1^2)^2} - 3. \tag{28}$$

As large claims are rare, π often tends to zero when k tends to infinity. Under these conditions, (27) and (28) show that both γ_l and κ_l tend to infinity but at different rates. Figure 3 plots μ, σ^2, γ and κ against k for various levels of π (denoted by pi).

Finally, the VaR can be calculated using the inverse function $F^{-1}_{\mathrm{CGB2}}(\rho|a,b,p,q,\pi,k)$, for quantile level $\rho \in (0,1)$ and some numerical search methods. The TCE is given by

$$\mathrm{TCE}_{\mathrm{CGB2},\rho} = (1-\pi+k\pi)\frac{bB(p+\frac{1}{a},q-\frac{1}{a})}{B(p,q)}\frac{1-F_B\left(F_B^{-1}(\rho|p,q)\,\middle|\,p+\frac{1}{a},q-\frac{1}{a}\right)}{1-\rho},$$

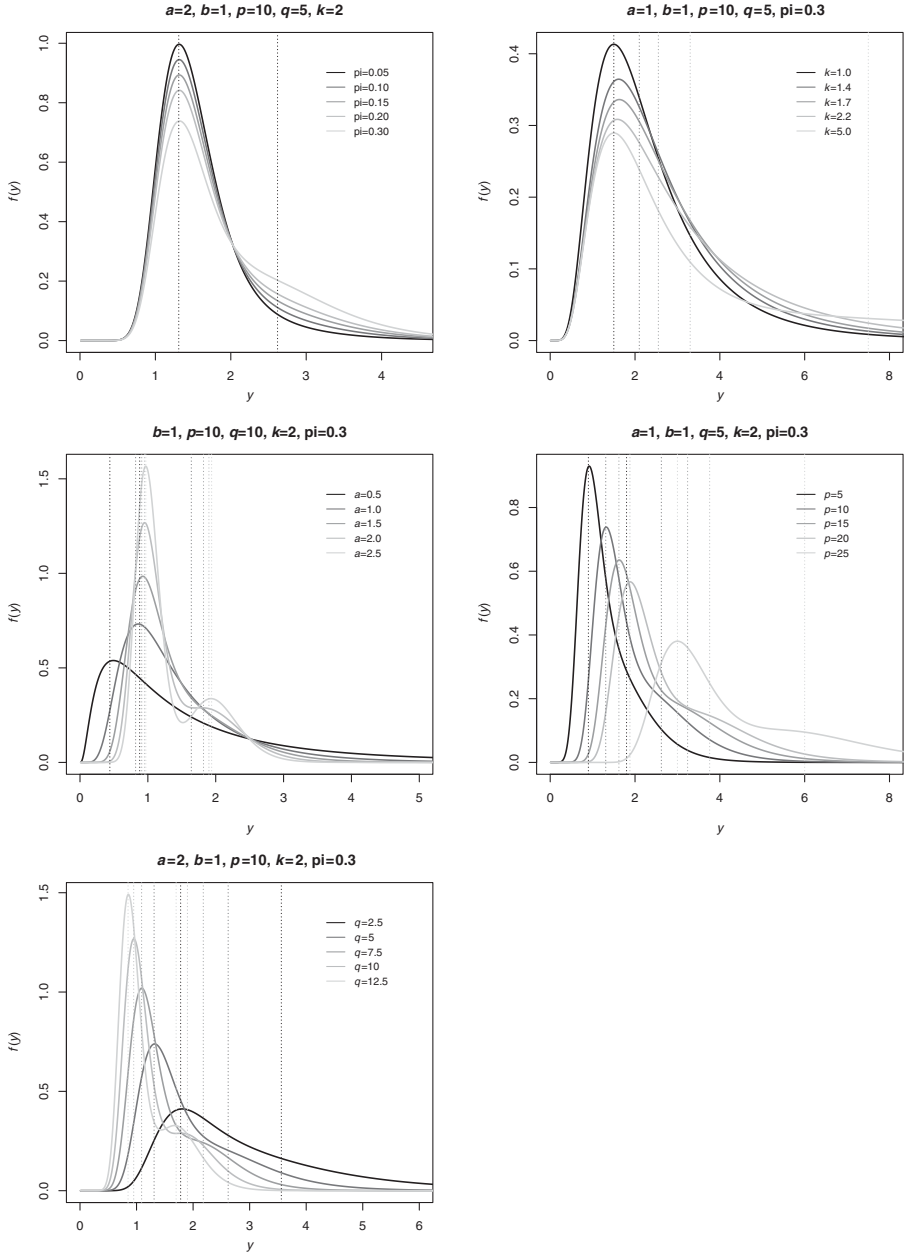


FIGURE 2: Density functions for different CGB2 distributions.

TABLE 1

FOUR MOMENT ESTIMATES AND MODES FOR THE PDFS IN FIGURE 2.

θ Varied	μ	σ^2	γ	κ	†Mode 1	†Mode 2
π	$a = 2, b = 1, p = 10, q = 5, k = 2, \pi$ varied					
0.45	2.195	1.058	1.199	2.178	1.31	2.62
0.35	2.043	0.950	1.418	2.922	1.31	2.62
0.25	1.892	0.795	1.691	4.214	1.31	2.62
0.15	1.741	0.595	2.018	6.427	1.31	2.62
0.05	1.589	0.349	2.184	9.329	1.31	2.62
k	$a = 1, b = 1, p = 10, q = 5, k$ varied, $\pi = 0.3$					
1	2.500	2.917	3.513	43.03	1.50	1.50
1.4	2.800	3.967	3.608	45.13	1.50	2.10
1.7	3.025	5.214	3.744	47.87	1.50	2.55
2.2	3.400	8.167	3.924	50.67	1.50	3.30
5	5.500	44.92	3.965	46.19	1.50	7.50
a	a varied, $b = 1, p = 10, q = 10, k = 2, \pi = 0.3$					
0.5	1.986	6.837	7.615	288.4	0.44	0.88
1	1.444	0.816	2.126	12.10	0.82	1.64
1.5	1.362	0.436	1.475	7.723	0.91	1.82
2	1.335	0.330	1.233	5.697	0.95	1.90
2.5	1.322	0.285	1.112	4.156	0.97	1.94
p	$a = 1, b = 1, p$ varied, $q = 5, k = 2, \pi = 0.3$					
5	1.374	0.487	1.613	4.000	0.90	1.80
10	1.968	0.878	1.547	3.482	1.31	2.62
15	2.420	1.269	1.520	3.270	1.62	3.24
20	2.800	1.659	1.504	3.154	1.88	3.76
50	4.444	4.001	1.475	2.929	3.00	6.00
q	$a = 2, b = 1, p = 10, q$ varied, $k = 2, \pi = 0.3$					
2.5	3.054	3.339	2.850	25.731	1.78	3.56
5	1.968	0.878	1.547	3.482	1.31	2.62
7.5	1.562	0.483	1.324	1.819	1.09	2.18
10	1.335	0.330	1.233	1.212	0.95	1.90
12.5	1.184	0.250	1.183	0.896	0.85	1.70

† Modes 1 and 2 are modes for the uncontaminated and contaminated groups resp.

(Appendix C). Obviously, $\text{TCE}_{\text{CGB2},\rho}$ tends to $E(Y_{\text{CGB2}})$ when ρ tends to 0. Lastly, we remark that the CGB2 distribution family contains all contaminated mixture distributions for the distributions listed in Figure 1 and Table B1, e.g. the contaminated Beta prime (CBP) distribution.

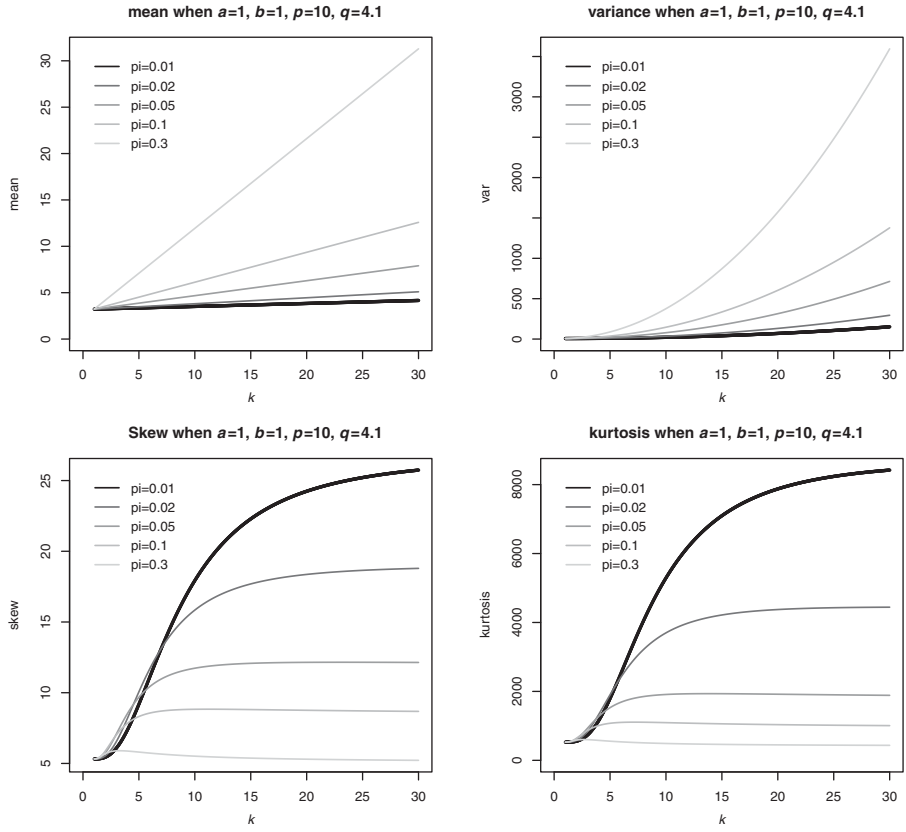


FIGURE 3: Change of μ, σ^2, γ and κ against k for various levels of π.

3. MODEL FITTING

3.1. Bayesian methodology

The fitting of the elaborated six-parameter CGB2 distribution and some of its special cases via the likelihood approach is difficult because the density function (17) and its higher order derivatives are complicated. Fortunately, the Bayesian MCMC and EM approaches make this achievable. The idea is to make use of the following hierarchical structure of the CGB2 distribution:

$$I_i \sim \text{Bern}(\pi), \tag{29}$$

$$Y_i \sim (1 - I_i)\,\text{GB2}(a, b, p, q) + I_i\,\text{GB2}(a, kb, p, q). \tag{30}$$

For example, the EM algorithm estimates the unknown group membership I_i in the E-step and then the model parameters in the M-step condition on $\hat{I}_i$. The group membership estimates $\hat{I}_i$ provide important information useful to

classify claims but they are unavailable in the classical likelihood approach as they are integrated out from the likelihood function. A discussion on the estimation procedures using the EM algorithm in this context is given in Appendix F. The hierarchical structure in (29) and (30) also facilitates the Bayesian MCMC approach which converts the optimisation problem of the likelihood approach to a sampling problem. Apart from the specific observations being analysed, the specification of prior information in the form of a probability distribution provides a powerful mechanism to incorporate expert knowledge which is often available to insurance companies. Such prior knowledge is simply ignored under the frequentist approach where point estimators and their distributions rely solely on large sample theory, which is often not applicable due to insufficient sample size in real applications. England and Verrall (2006) show how distributions of outstanding liabilities can be obtained under the Bayesian MCMC approach. In the situation where there is no agreement on the prior information, non-informative or reference priors can be used. This leads to what is called objective Bayesian inference (Berger, 2006). Our choice of priors in the simulation and empirical studies are

$$a \sim \mathrm{G}(0.04, 0.01)I(0.1, 10),\ b \sim \mathrm{G}(1, 0.001)I(0.01, 10, 000),$$
$$p \sim \mathrm{G}(0.05, 0.01)I(1, 10, 000),\ q \sim \mathrm{G}(0.01, 0.01)I(0.01, 10),$$
$$k \sim \mathrm{G}(0.02, 0.01)I(2, 50),\ \pi \sim \mathrm{Beta}(1, 1)I(0.001, 0.5),$$

where $\mathrm{G}(c, d)I(u, v)$ denotes the gamma distribution with mean c/d, variance c/d^2 and restricted support (u, v). The hyperparameters are chosen to give approximately non-informative priors and the restricted supports are set to ensure stable estimates. Note that some parameters are fixed when fitting members of the CGB2 family, e.g. we fix $a = 1$ in fitting the CBP distribution.

Among many MCMC techniques (Gilks *et al.*, 1996), the Gibbs sampling (Smith and Roberts, 1993) is adopted to generate realisations that mimic a random sample from the joint posterior distribution for approximate posterior inference. Parameter estimates and credible intervals are obtained and a predictive distribution is constructed to estimate future claims. The probability of catastrophic large losses can be evaluated from the predictive distribution based on VaR and TCE measures.

For the contaminated and composite distributions in the empirical study (Sections 5), a Gibbs samplier is run for 55,000 iterations, discarding the initial 5,000 realisations as a burn-in period. In the simulation study (Section 4), we use 6,000 iterations after a burn-in period of 4,000 iterations for each simulated data set. The ergodic average and autocorrelation plots are carefully checked to ensure convergence of the constructed Markov chain. The resulting posterior sample of size 50,000 can be regarded as realisations from the joint posterior distribution. For the folded distributions in Section 5, a much longer chain of 505,000 is run and every 10th simulated value is taken to reduce the high autocorrelation. Parameter estimates are given by the posterior mean or median but

the latter being more robust is preferable for parameters of a typical skewed distribution. All models are implemented using the user-friendly Bayesian software `OpenBUGS`, allowing practitioners to perform Bayesian analysis on complex statistical models easily. The `OpenBUGS` code to estimate the proposed CGB2 distribution is provided in Appendix E.

3.2. Goodness-of-fit statistics

To assess the model-fit and rank competitive models, we consider the log-likelihood (ℓ) and the Bayesian DIC, which measure the model fit without and with penalty for model complexity, respectively. For both measures, the log-likelihood function $\ell = \sum_{i=1}^{n} \ln f(y_i|\widehat{\boldsymbol{\theta}})$ is evaluated using parameter estimates $\widehat{\boldsymbol{\theta}}$, where $f(\cdot|\widehat{\boldsymbol{\theta}})$ is given by (17) for CGB2 distribution, and $\boldsymbol{\theta}$ is the vector of model parameters (a, b, p, q, k, π). The DIC consists of the posterior mean deviance $\overline{D(\boldsymbol{\theta})} = E_{\theta|y}[D(\boldsymbol{\theta})]$, which can be interpreted as a Bayesian measure of model fit, and a penalty accounting for model complexity. This penalty is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters, that is,

$$p_{D1} = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}}). \tag{31}$$

See Spiegelhalter *et al.* (2002) for details. When some posterior distributions are skewed, the posterior mean may not be a good summary statistic and p_{D1} may be negative. In this case, we replace the posterior means by the posterior medians in the calculation of $D(\overline{\boldsymbol{\theta}})$. Alternatively, Gelman *et al.* (2014; P.173) suggest the penalty to be half of the variance estimate for the deviance from the MCMC output, that is

$$p_{D2} = \frac{1}{2}\mathrm{Var}(D(\boldsymbol{\theta})),$$

to lessen the effect of skewed posterior distribution. The DIC is then calculated by

$$\mathrm{DIC}_i = \overline{D(\boldsymbol{\theta})} + p_{Di}, \qquad i = 1, 2.$$

Clearly, models with the smallest DIC_i and largest ℓ are preferred in model selection.

4. Simulation study

First, we assess the performance of Bayes estimators via `OpenBUGS` by simulating $N = 100$ data sets. In each data set, $y_i = (1-I_i)y_{1i} + I_i k y_{1i}$, $i = 1, \ldots, 1{,}000$ are simulated from the CGB2 distribution with $I_i \sim \mathrm{Bern}(\pi)$ and $y_{1i} \sim \mathrm{GB2}(a, b, p, q)$. The true parameters are $a = 2$, $b = 1$, $p = 10$, $q = 5$, $k = 2$ and $\pi = 0.3$. This means $y_i = y_{1i}$ if $I_i = 0$ with probability $1 - \pi$ and $y_i = k y_{1i}$ if $I_i = 1$ with probability π, noting from (5) that if $Y \sim \mathrm{GB2}(a, b, p, q)$,

TABLE 2

PARAMETER ESTIMATES IN THE SIMULATION STUDY.

Parameter	True	Estimate	SD	RMSE	Min	Max
a	2	1.967	0.652	0.649	0.847	3.789
b	1	1.033	0.458	0.457	0.518	4.661
p	10	15.278	9.480	10.809	2.465	40.350
q	5	7.536	6.151	6.625	1.790	34.340
k	2	4.975	0.324	2.993	1.934	5.255
π	0.3	0.301	0.015	0.015	0.667	0.741

$kY \sim \text{GB2}(a, kb, p, q)$. To simulate y_{1i} from the $\text{GB2}(a, b, p, q)$ distribution, u_i is first simulated from the beta(p, q) distribution using the command `u=rbeta(n,p,q)` in `R` and then y_{1i} is obtained through the transformation

$$y_{1i} = b\left(\frac{u_i}{1-u_i}\right)^{\frac{1}{a}} \sim \text{GB2}(a, b, p, q)$$

from (8). The fitting of each data set using `OpenBUGS` is facilitated by calling the `R` package `R2OpenBUGS`. For robustness consideration, we report the average of $N = 100$ posterior median of the parameter estimates and their standard deviation (SD), root mean squared error (RMSE), minimum (Min) and maximum (Max) for each model parameter in Table 2.

Results show that parameter estimates are reasonably accurate for a, b and π but parameters p, q and k show consistent upward biases, a common phenomenon for the shape parameters which may approach infinity when the likelihood surface is flat (Cummins *et al.*, 1990). Despite the discrepancies in these parameter estimates, the estimated pdfs agree closely with the true pdfs. This is demonstrated in Figure 4 which plots for the first simulated data set, the pdfs $f_{\text{GB2}}(y|a, b, p, q)$, $f_{\text{GB2}}(y|a, kb, p, q)$ and $f_{\text{CGB2}}(y|a, b, p, q, \pi, k)$ in blue, red and black lines, respectively, superimposed on the histogram of the simulated data. The dotted lines represent the pdfs with true parameters and the solid lines with parameter estimates.

Simulation results show that the fitted pdfs are close to the true pdfs and that the pdfs for the two latent subgroups display distinct characteristics in terms of spreads, modes and tails. Table 3 reports the mean, variance, skewness and excess kurtosis for the true and estimated pdf calculated using (23) to (26). Results show that these parameter estimates agree closely with their true values, allowing for the sampling errors in the simulated data.

5. APPLICATIONS

We analyse two real data sets. The first data set, given in Table B1 of Cummins and Freifelder (1978) and Table A.1 of Cummins *et al.* (1990), contains 80 fire

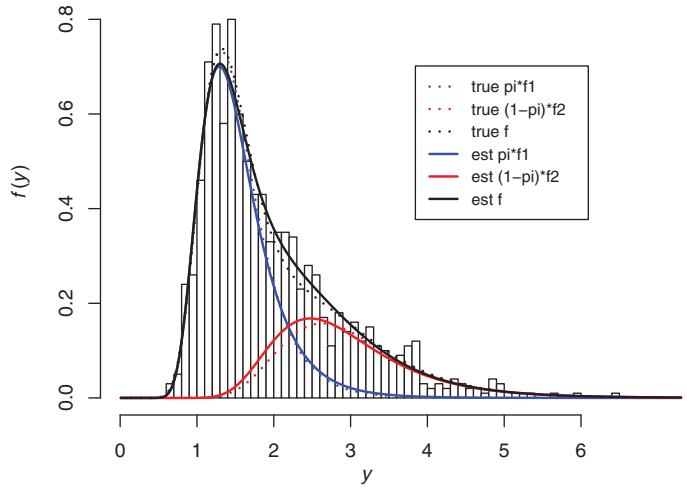


FIGURE 4: True and estimated pdfs for the first simulation data set. `f1` and `f2` are the pdfs of the GB2(a,b,p,q) and GB2(a,bk,p,q), respectively. (Color online)

TABLE 3

FOUR MOMENT ESTIMATES OF THE CGB2 DISTRIBUTION FOR THE FIRST SIMULATED DATA SET.

Based On	μ	σ^2	γ	κ
True pdf f_{CGB2}	1.968	0.878	1.547	3.482
Estimated pdf $\hat{f}_{\text{CGB2}}$	1.973	0.874	1.651	4.612

losses from 500 buildings owned by a large university during the period 1951 to 1973. The second data set includes aggregates of these losses by year for 23 years, given in Table 2 of Cummins *et al.* (1990). These sets of data were analysed in Cummins *et al.* (1990) using the GB2 distribution.

We fit 21 contaminated mixture distributions for the GB2 family (Table B1 Appendix B) to the data. For model comparison, we consider four additional models: models A and B are the folded normal (FN) and FT distributions and models C and D are the composite LP and composite WP distributions. Parameters are estimated using the Bayesian MCMC approach via `OpenBUGS` with codes provided in Appendix E.

5.1. Individual loss severity data

For the 80 loss severities, the sample estimates of μ, σ, γ and κ are 16,950, 71,944, 7.81 and 66.12, respectively. The data as plotted in Figure 5 show the presence of a large outlier enhancing the heavy skewness and kurtosis. Cummins and Freifelder (1978) find that the lognormal and gamma distributions do not have sufficiently heavy tails to model the data and therefore, they consider the GB2 distribution. Parameter estimates and model fit measures for all

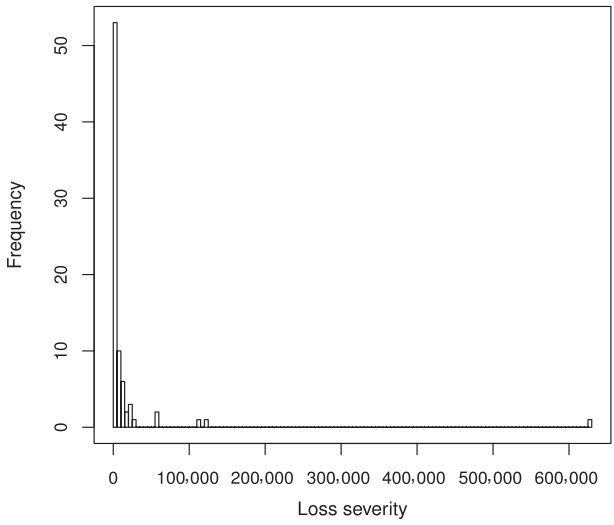


FIGURE 5: Histogram for the individual loss severity data.

contaminated mixture, folded and composite distributions are exhibited in Table 4, and the mean, median and various VaR and TCE estimates for the predictive distributions are reported in Table 5. The VaRs are also plotted in Figure 6 to facilitate comparison.

Table 4 shows that the contaminated proportion π is estimated to be around 10% for most contaminated distributions. Although the contaminated scale k is defined differently in different distributions and hence it cannot be directly compared (see (18) and (19)), it tends to be more consistent within subgroups of distributions as indicated in Figure 1. The contaminated log distributions (M6, M15, M16) do not provide good fit to the data as they give too heavy tails, whereas the contaminated half normal and half t distributions (M20, M21) give even worse fit because of the light tail for half normal distribution and the constraint of equating the scale parameter to the degrees of freedom for the half t distribution. These contaminated half distributions are close to the folded distributions (Models A and B) whose location parameters are estimated to be relatively small but they do not have the contaminated component to account for extreme losses. Hence, the worst fit comes from the FN distribution. Overall, the CB distribution (M3; $p = 1$) provides the best overall model fit but the composite distributions (Models C and D) also provide reasonably good model fit because of the heavy right tail of the Pareto distribution. However, their DIC_1 values are relatively large.

Table 5 presents the posterior mean, median and estimates of VaR and TCE based on the predictive distribution for all fitted models. Since the log and composite distributions have too heavy tails, their VaR and TCE estimates are either extremely large or do not even exist. Figure 6 plots the VaR across models to

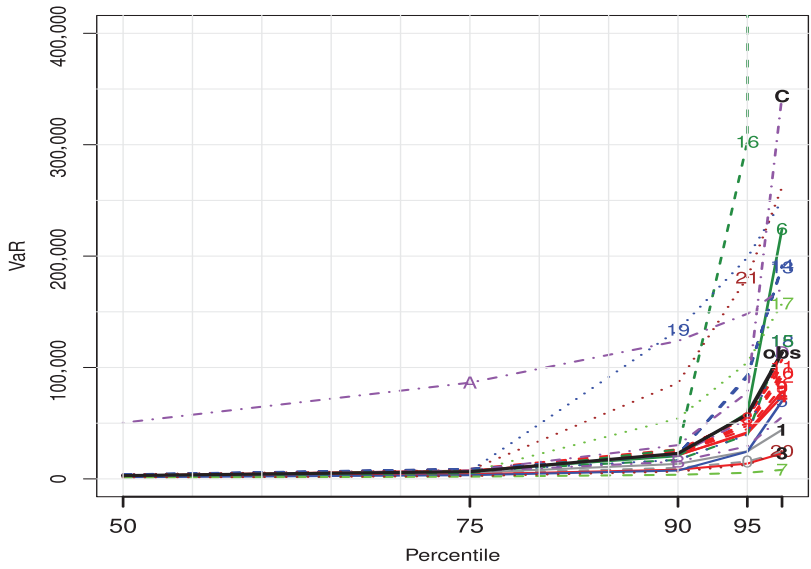


FIGURE 6: VaR across models for individual loss severity data. The solid, dash, dotted and dot-dash lines indicate 5–6, 4, 3 and 2 parameter distributions for the contaminated mixture, folded and composite families. The black, grey, green, blue, red, light green, brown and purple lines indicate observed quantile, CGB2/GB2, log, GG family, consistent-scale (b consistent and any a, p, q set to 1), VR/CS, HT/HN and folded/composite family respectively, consistent with Figure 1. M1, M3 and Model C are highlighted for comparison. (Color online)

facilitate comparisons. The colours of the lines are chosen to be consistent with the colours in Figure 1 to indicate different sub-groups of contaminated distributions. Figure 6 clearly shows that VaR estimates are mostly consistent within the sub-groups (same colour) and the log distributions (green lines) greatly overestimate VaR, whereas the CVR distribution (M7) greatly underestimates VaR giving the second worst fit. The figure also explains why the FN distribution (Model A) provides the worst fit as it overestimates most VaR to accommodate the extreme loss with a thin tail. The chosen CB distribution (M3) underestimates high levels VaR relatively more than the other distributions in the same consistent-scale sub-group (red lines), whereas the composite LP distribution (Model C) overestimates them.

Figure 7 displays the histogram and densities for the posterior predictive distributions using the chosen M3 and compares the densities to the respective fitted densities. Kernel smoothing is applied to obtain the densities for the predictive sample of each component, the weights ($1\text{-}\pi$ or π) are then multiplied by the densities and the degree of smoothing is adjusted to match the height of the peaks for the fitted densities with weights. Figure 7 clearly shows that the peak locations match closely for the main component of the CB distribution which captures mainly the small losses. The densities for the contaminated component are more flat to accommodate larger losses from the second and third clusters. The enhanced tail of the CB distribution can be viewed from the "lift-up" of

TABLE 4

PARAMETER ESTIMATES, SD (ITALIC) AND MODEL FITS FOR INDIVIDUAL LOSS SEVERITY DATA.

M	Dist	Par	a	SD	b	SD	p	SD	q	SD	k	SD	π	SD	ℓ	DIC_1	DIC_2
0	GB2*	4	1.179	*1.32*	1,084	*1,009*	15.84	*28.7*	0.745	*1.87*	1	–	0.0000	–	−267.8	537	538
1	CGB2	6	1.432	*1.56*	1,163	*1,009*	17.04	*31.6*	0.746	*1.80*	5.794	*7.05*	0.2369	*0.14*	**−267.7**	538	538
2	CBP	5	1	–	874	*955*	15.04	*27.8*	1.284	*0.57*	6.108	*7.64*	0.2232	*0.14*	**−267.7**	538	538
3	CB	5	4.597	*2.02*	3,875	*1,058*	1	–	0.165	*0.11*	5.368	*5.80*	0.2980	*0.14*	−268.3	540	541
4	CIB	5	1.202	*0.28*	1,039	*800*	14.45	*16.8*	1	–	6.251	*6.98*	0.2411	*0.13*	**−267.6**	538	538
5	CGG†	5	0.409	*0.18*	310	*5,638*	5.260	*3.77*	∞	∞	8.968	*8.78*	0.2088	*0.13*	−269.5	542	542
6	CLT‡	5	0	–	9.837	*0.28*	1.018	*0.27*	3.348	*2.73*	3.454	*7.70*	0.0919	*0.12*	−269.9	541	545
7	CVR	4	1	–	–	–	2.566	*3.10*	2.017	*0.03*	7.840	*11.92*	0.0294	*0.03*	−340.9	680	688
8	CP	4	1	–	3,973	*1,718*	1	–	0.632	*0.21*	5.525	*6.82*	0.3385	*0.14*	−274.3	550	555
9	CIP	4	1	–	767	*656*	14.29	*1.87*	1	–	5.234	*9.20*	0.1569	*0.14*	**−267.6**	**537**	**537**
10	CLL	4	1.736	*0.45*	11,200	*2,601*	1	–	1	–	6.892	*6.12*	0.3000	*0.13*	−268.7	540	540
11	CPL	4	1.345	*0.27*	12,880	*2,939*	1	–	1.345	*0.27*	7.024	*6.49*	0.3206	*0.13*	−269.7	542	543
12	CIPL	4	1.540	*0.28*	9,306	*2,438*	1.540	*0.28*	1	–	6.223	*7.38*	0.2410	*0.13*	−268.2	538	538
13	CGA†	4	1	–	10,840	*2,793*	1.486	*0.37*	∞	–	11.27	*5.71*	0.2797	*0.10*	−270.4	544	543
14	CW†	4	0.981	*0.22*	13,650	*3,183*	1	–	∞	–	11.56	*6.70*	0.3205	*0.11*	−271.5	546	547
15	CLN‡	4	0	–	9.891	*0.298*	1.156	*0.28*	∞	–	3.711	*9.79*	0.1017	*0.13*	−269.9	539	544
16	CLC‡	4	0	–	9.661	*0.314*	0.676	*0.25*	1	–	3.971	*8.14*	0.1053	*0.14*	−274.3	548	552
17	CCS♭	3	1	–	∞	–	3.298	*0.73*	∞	–	17.76	*4.73*	0.4272	*0.07*	−281.4	247	565
18	CEXP†	3	1	–	13,420	*3,093*	1	–	∞	–	11.62	*6.70*	0.3278	*0.11*	−271.0	544	544
19	CR†	3	2	–	17,500	*2,673*	1	–	∞	–	16.00	*3.97*	0.2817	*0.09*	−276.6	556	558
20	CHT‡ ♯	3	2	–	1.032	*0.049*	0.5	–	1.032	*0.05*	17.76	*4.73*	0.4272	*0.07*	−313.5	627	630
21	CHN‡♯	3	2	–	24,300	*11,150*	0.5	–	∞	–	12.73	*6.19*	0.2461	*0.11*	−272.0	547	546
A	FN⋆	2	−2,482	*31,460*	141,600	*23,200*	–	–	–	–	–	–	–	–	−290.0	581	581
B	FT⋆	3	70.43	*10,620*	22,710	*8,630*	–	–	–	–	–	–	–	–	−270.3	542	544
C	LP♭	2	0.647	*0.13*	11,200	*2,888*	–	–	–	–	–	–	–	–	−268.1	543	538
D	WP♭	2	0.711	*0.15*	13,460	*3,769*	–	–	–	–	–	–	–	–	−268.6	543	539

Note: †: b reports β; ‡: b, p, q report μ, σ, ν; ♯: q reports ν; ♮: b reports σ; ⋆: a, b report μ, σ; ♭: a, b report a, θ.

TABLE 5

MEAN, MEDIAN, VAR AND TCE MEASURES OF THE PREDICTIVE DISTRIBUTION FOR INDIVIDUAL LOSS SEVERITY DATA.

				VaR				TCE			
M	Dist	Mean	Median	75	90	95	97.5	75	90	95	97.5
0	GB2	5,092	2,374	4,902	9,910	15,570	24,230	14,266	25,429	38,459	97,131
1	CGB2	8,067	2,530	5,510	13,430	24,750	44,030	25,747	51,736	85,522	253,829
2	CBP	12,600	2,981	7,432	21,090	41,810	78,390	42,631	88,237	147,044	425,056
3	CB	4,572	2,139	4,091	8,245	13,720	22,860	12,797	23,516	36,638	93,229
4	CIB	11,810	2,915	7,299	21,490	40,760	74,000	39,639	80,622	131,980	372,689
5	CGG	15,030	1,485	3,324	7,506	24,490	69,740	51,945	111,234	180,386	402,182
6	CLT	∞	2,888	7,014	20,380	59,300	224,300	∞	∞	∞	∞
7	CVR	1,982	1,154	2,011	3,692	5,511	8,093	4,955	8,398	12,365	30221
8	CP	18,980	3,046	8,969	26,570	54,250	106,800	68,188	147,427	256,898	862,217
9	CIP	13,190	3,173	8,169	22,740	44,680	82,290	44,513	91,303	151,173	434,887
10	CLL	13,940	3,378	7,868	23,560	50,130	94,970	47,268	98,494	162,473	438,271
11	CPL	13,720	3,492	7,955	25,160	55,710	100,700	46,238	95,439	152,918	376,779
12	CIPL	14,400	3,459	8,253	22,940	47,660	91,520	48,853	102,183	171,950	504,802
13	CGA	18,670	3,914	8,723	21,800	93,290	189,800	65,429	144,242	243,536	515,404
14	CW	18,800	4,046	9,040	22,040	91,140	191,700	65,606	143,979	242,256	500,758
15	CLN	1,298,000	3,049	7,009	17,080	39,580	123,700	5,185,979	12,949,222	25,873,362	128,714,858
16	CLC	∞	2,507	5,132	26,370	302,549	42,298,036	∞	∞	∞	∞
17	CCS	17,360	3,052	6,356	54,260	104,500	157,400	62,064	127,834	178,121	295,127
18	CEXP	18,540	3,049	7,009	17,080	39,580	123,700	64,510	141,741	235,601	494,156
19	CR	31,010	4,153	6,815	133,700	198,600	249,900	113,794	213,598	262,195	348,508
20	CHT	5,293	1,061	2,501	6,495	12,610	25,040	18,573	40,573	72,197	263,045
21	CHN	26,840	4,020	7,854	86,062	180,640	261,825	97,912	210,981	290,586	363,938
A	FN	59,990	50,410	86,380	123,800	148,200	170,700	124,442	156,575	178,324	223,005
B	FT	14,810	3,038	6,550	15,510	29,340	55,590	54,446	121,451	222,044	888,984
C	LP	530,600	2,655	7,560	30,308	77,487	343,645	5,014,609	12,514,789	24,982,036	49,792,298
D	WP	592,600	2,627	6,629	22,059	53,626	113,573	2,226,243	5,548,144	11,061,436	22,046,024

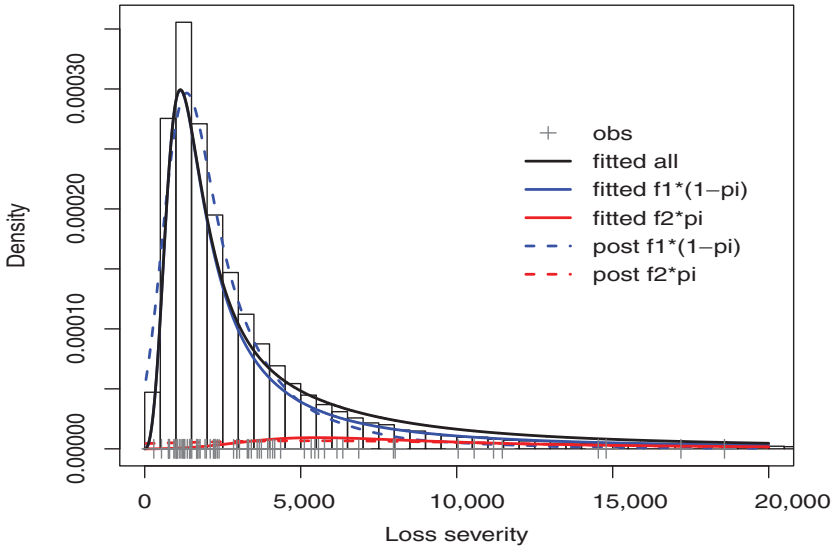


FIGURE 7: Histogram of the posterior predictive distributions using M3 for individual loss severity data, superimposed with the weighted densities of the main component (blue) and contaminated component (red) for the posterior predictive distributions (dash line) and fitted distributions (solid line) and with the overall fitted density (black solid line). The observed data are plotted as "+" along the horizontal axis. (Color online)

the overall density (black solid line) towards the right tail by the contaminated component (red line) as compared to the main component (blue line). To lessen the effect of extreme losses, individual losses are aggregated by year in the subsequent analysis. Despite the smaller sample size, modelling annual losses rather than individual losses allows the estimation of annual loss reserves directly.

5.2. Aggregate fire losses

For the 23 aggregate fire losses, the sample estimates of μ, σ, γ and κ are 59,183, 130,000, 3.94 and 17.84, respectively, exhibiting less skewness and kurtosis than the individual loss data in Section 5.1. Figure 8 depicts the histogram of the data, showing clearly the impact of a large outlier.

Table 6 reports parameter estimates, standard error and model fit criteria for all fitted models. Compared with individual loss data, the estimates $\hat{b}$ and $\hat{\pi}$ are larger, whereas $\hat{k}$ and $\hat{p}$ are smaller in general. The estimate for the degree of contamination $\hat{\pi}$ is around 20% and is more consistent across distributions. Moreover, $\hat{p}$ is much larger than $\hat{q}$ in GB2 (M0) and CGB2 (M1) distributions showing that the inverse distributions such as CIB and CIP (M4, M9; $q = 1$) distributions are more favourable than the CB and CP (M3, M8; $p = 1$) distributions, respectively. Moreover, because $\hat{a}$ and $\hat{q}$ are close to 1 in GB2 and CGB2 distributions, four distributions, namely CGB2, CBP (M2; $a = 1$), CIB and CIP (M9; $a = q = 1$), show similar best fits as compared to the individual loss data with the best model being CB ($p = 1$). Allowing for model complexity,

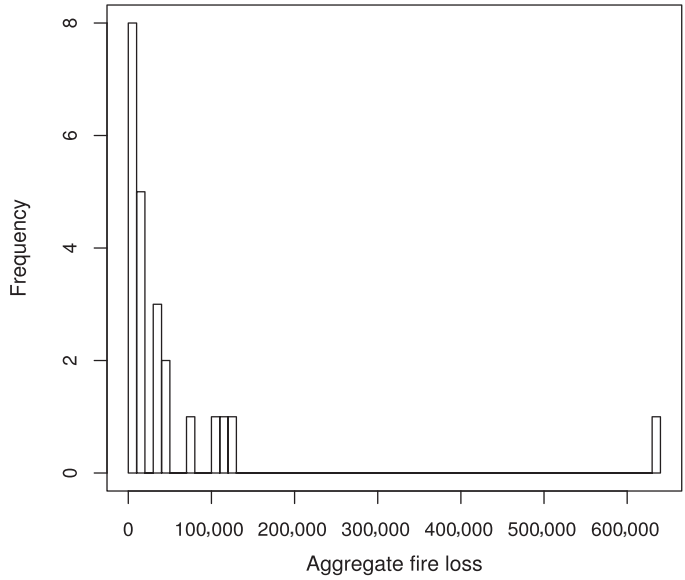


FIGURE 8: Histogram for the aggregate fire losses data.

the simpler CIP distribution provides the best model performance according to DIC_1 and DIC_2, and is marginally superior to the CGB2, CBP and CIB distributions.

Table 7 reports the posterior mean, posterior median and estimates of the VaR and TCE based on the posterior predictive distribution for all fitted distributions. The VaR estimates are also graphed in Figure 9 to facilitate comparison. Again the log distributions (green lines) provide very extreme VaRs, whereas CVR (M7) distribution provides the smallest VaR estimates and the worst fit. On the other hand, CGB2, CBP, CIB and CIP (M1, M2, M4, M9) distributions provide VaR estimates reasonably close to the observe quantiles. From a model fit perspective, all the four distributions provide thick enough tails to accommodate the outlier but the VaR estimates from CGB2 distribution are closer to the observed quantiles in general and hence is chosen to provide the predictive distribution in Figure 10.

The peak locations for the main component of both predictive and fitted distributions match nearly perfectly. This component of the CGB2 distribution captures the vast majority of small losses, whereas the contaminated component accommodates clusters of larger losses with an enhanced tail to capture extreme losses.

6. CONCLUSION

The GB2 family is used to model the fire loss data because it nests many popular distributions and hence provides a great flexibility in describing various

TABLE 6

PARAMETER ESTIMATES, SD (ITALIC) AND MODEL FITS FOR AGGREGATE FIRE LOSSES DATA.

M	Dist	Par	a	SD	b	SD	p	SD	q	SD	k	SD	π	SD	ℓ	DIC_1	DIC_2
0	GB2*	4	1.179	*1.32*	1,084	*1,009*	15.84	*28.7*	0.745	*1.87*	1	–	0.0000	–	−267.8	537	538
1	CGB2	6	1.432	*1.56*	1,163	*1,009*	17.04	*31.6*	0.746	*1.80*	5.794	*7.05*	0.2369	*0.14*	**−267.7**	538	538
2	CBP	5	1	–	874	*955*	15.04	*27.8*	1.284	*0.57*	6.108	*7.64*	0.2232	*0.14*	**−267.7**	538	538
3	CB	5	4.597	*2.02*	3,875	*1,058*	1	–	0.165	*0.11*	5.368	*5.80*	0.2980	*0.14*	−268.3	540	541
4	CIB	5	1.202	*0.28*	1,039	*800*	14.45	*16.8*	1	–	6.251	*6.98*	0.2411	*0.13*	**−267.6**	538	538
5	CGG†	5	0.409	*0.18*	310	*5,638*	5.260	*3.77*	∞	∞	8.968	*8.78*	0.2088	*0.13*	−269.5	542	542
6	CLT‡	5	0	–	9.837	*0.28*	1.018	*0.27*	3.348	*2.73*	3.454	*7.70*	0.0919	*0.12*	−269.9	541	545
7	CVR	4	1	–	–	–	2.566	*3.10*	2.017	*0.03*	7.840	*11.92*	0.0294	*0.03*	−340.9	680	688
8	CP	4	1	–	3,973	*1,718*	1	–	0.632	*0.21*	5.525	*6.82*	0.3385	*0.14*	−274.3	550	555
9	CIP	4	1	–	767	*656*	14.29	*1.87*	1	–	5.234	*9.20*	0.1569	*0.14*	**−267.6**	**537**	**537**
10	CLL	4	1.736	*0.45*	11,200	*2,601*	1	–	1	–	6.892	*6.12*	0.3000	*0.13*	−268.7	540	540
11	CPL	4	1.345	*0.27*	12,880	*2,939*	1	–	1.345	*0.27*	7.024	*6.49*	0.3206	*0.13*	−269.7	542	543
12	CIPL	4	1.540	*0.28*	9,306	*2,438*	1.540	*0.28*	1	–	6.223	*7.38*	0.2410	*0.13*	−268.2	538	538
13	CGA†	4	1	–	10,840	*2,793*	1.486	*0.37*	∞	–	11.27	*5.71*	0.2797	*0.10*	−270.4	544	543
14	CW†	4	0.981	*0.22*	13,650	*3,183*	1	–	∞	–	11.56	*6.70*	0.3205	*0.11*	−271.5	546	547
15	CLN‡	4	0	–	9.891	*0.298*	1.156	*0.28*	∞	–	3.711	*9.79*	0.1017	*0.13*	−269.9	539	544
16	CLC‡	4	0	–	9.661	*0.314*	0.676	*0.25*	1	–	3.971	*8.14*	0.1053	*0.14*	−274.3	548	552
17	CCS♭	3	1	–	∞	–	3.298	*0.73*	∞	–	17.76	*4.73*	0.4272	*0.07*	−281.4	247	565
18	CEXP†	3	1	–	13,420	*3,093*	1	–	∞	–	11.62	*6.70*	0.3278	*0.11*	−271.0	544	544
19	CR†	3	2	–	17,500	*2,673*	1	–	∞	–	16.00	*3.97*	0.2817	*0.09*	−276.6	556	558
20	CHT‡ ♯	3	2	–	1.032	*0.049*	0.5	–	1.032	*0.05*	17.76	*4.73*	0.4272	*0.07*	−313.5	627	630
21	CHN‡♯	3	2	–	24,300	*11,150*	0.5	–	∞	–	12.73	*6.19*	0.2461	*0.11*	−272.0	547	546
A	FN⋆	2	−2482	*31,460*	141,600	*23,200*	–	–	–	–	–	–	–	–	−290.0	581	581
B	FT⋆	3	70.43	*10,620*	22,710	*8,630*	–	–	–	–	–	–	–	–	−270.3	542	544
C	LP♭	2	0.647	*0.13*	11,200	*2,888*	–	–	–	–	–	–	–	–	−268.1	543	538
D	WP♭	2	0.711	*0.15*	13,460	*3,769*	–	–	–	–	–	–	–	–	−268.6	543	539

Note: †: b reports β; ‡: b, p, q report μ, σ, ν; ♯: q reports ν; ♮: b reports σ; ⋆: a, b report μ, σ; ♭: a, b report a, θ.

TABLE 7

MEAN, MEDIAN, VAR AND TCE MEASURES OF THE PREDICTIVE DISTRIBUTION FOR AGGREGATE FIRE LOSSES DATA.

				VaR				TCE			
M	Dist	Mean	Median	75	90	95	97.5	75	90	95	97.5
0	GB2	5,092	2,374	4,902	9,910	15,570	24,230	14,266	25,429	38,459	97131
1	CGB2	8,067	2,530	5,510	13,430	24,750	44,030	25,747	51,736	85,522	253,829
2	CBP	12,600	2,981	7,432	21,090	41,810	78,390	42,631	88,237	147,044	425,056
3	CB	4,572	2,139	4,091	8,245	13,720	22,860	12,797	23,516	36,638	93229
4	CIB	11,810	2,915	7,299	21,490	40,760	74,000	39,639	80,622	131,980	372,689
5	CGG	15,030	1,485	3,324	7,506	24,490	69,740	51,945	111,234	180,386	402,182
6	CLT	∞	2,888	7,012	20,370	59,230	224,200	∞	∞	∞	∞
7	CVR	1,982	1,154	2,011	3,692	5,511	8,093	4,955	8,398	12,365	30221
8	CP	18,980	3,046	8,969	26,570	54,250	106,800	68,188	147,427	256,898	862,217
9	CIP	13,190	3,173	8,169	22,740	44,680	82,290	44,513	91,303	151,173	434,887
10	CLL	13,940	3,378	7,868	23,560	50,130	94,970	47,268	98,494	162,473	438,271
11	CPL	13,720	3,492	7,955	25,160	55,710	100,700	46,238	95,439	152,918	376,779
12	CIPL	14,400	3,459	8,253	22,940	47,660	91,520	48,853	102,183	171,950	504,802
13	CGA	18,670	3,914	8,723	21,800	93,290	189,800	65,429	144,242	243,536	515,404
14	CW	18,800	4,046	9,040	22,040	91,140	191,700	65,606	143,979	242,256	500,758
15	CLN	1,298,000	3,049	7,009	17,080	39,580	123,700	5,185,979	12,949,222	25,873,362	12,871,4858
16	CLC	∞	2,507	5,132	26,370	302,549	42,298,036	∞	∞	∞	∞
17	CCS	17,360	3,052	6,356	54,260	104,500	157,400	62,064	127,834	178,121	295,127
18	CEXP	18,540	3,049	7,009	17,080	39,580	123,700	64,510	141,741	235,601	494,156
19	CR	31,010	4,153	6,815	133,700	198,600	249,900	113,794	213,598	262,195	348,508
20	CHT	5,293	1,061	2,501	6,495	12,610	25,040	18,573	40,573	72,197	263,045
21	CHN	26,840	4,020	7,854	86,062	180,640	261,825	97,912	210,981	290,586	363,938
A	FN	59,990	50,410	86,380	123,800	148,200	170,700	124,442	156,575	178,324	223,005
B	FT	14,810	3,038	6,550	15,510	29,340	55,590	54,446	121,451	222,044	888,984
C	LP	530,600	2,655	7,560	30,308	77,487	343,645	5,014,609	12,514,789	24,982,036	49,792,298
D	WP	592,600	2,627	6,629	22,059	53,626	113,573	2,226,243	5,548,144	11,061,436	22,046,024

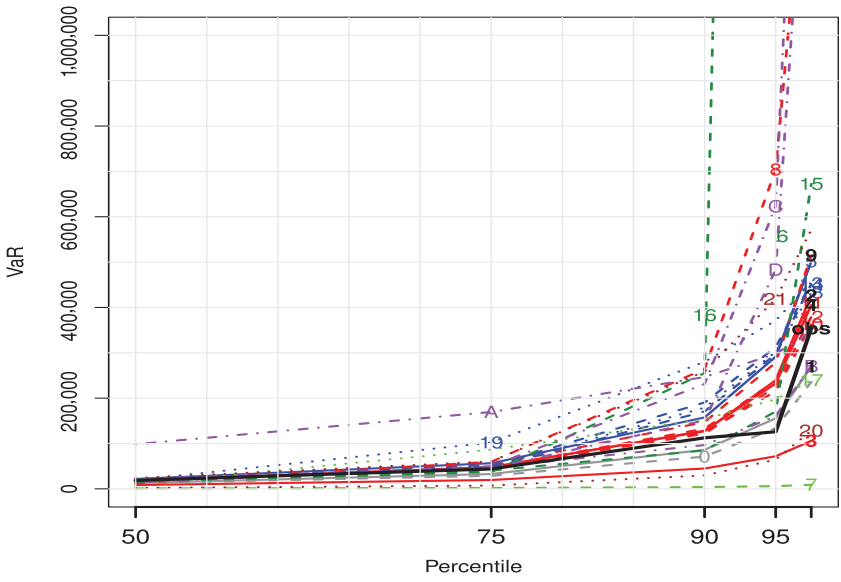


FIGURE 9: VaR across models for aggregate fire losses data. (Color online)

features of the data. This paper extends the GB2 distribution to a richer family, the CGB2 distribution, which provides even thicker tails and allows bimodality, thus accommodating heterogeneity of small and extreme losses. Essentially, this six-parameter distribution is a mixture of two GB2 distributions in which one has an enhanced scale and hence a thicker tail to capture extreme losses.

To describe the characteristics of the CGB2 distribution, its moments, mode and risk measures such as the VaR and TCE are derived. We also examine the impact on the shape of the pdf when one of the shape parameters changes while the others are fixed. We show that the four characterising moments $(\mu, \sigma^2, \gamma, \kappa)$ of the CGB2 distribution cover a wide range of values, allowing flexible shape for modelling any loss data. Models are implemented using the Bayesian approach via `OpenBUGS` and its performance is investigated through a simulation study. Results show that parameter estimates are reasonably close to their true values. Moreover, analysis of the first simulated data set shows that the estimated pdf describes the shape of the distributions for the two latent subgroups fairly closely, and it also provides good matches between the empirical and fitted values of the four characterising moments. We remark that the EM algorithm in the likelihood approach is also an efficient method to estimate the model parameters and details of the algorithm are given in Appendix F.

The applicability of the CGB2 distribution is demonstrated through two real loss data, the individual and the aggregate fire losses, in which both contain a very large outlier. We compare the CGB2 distribution with the GB2 distribution and two additional distributions, the folded distribution and the composite distribution which share some similarities with our proposed

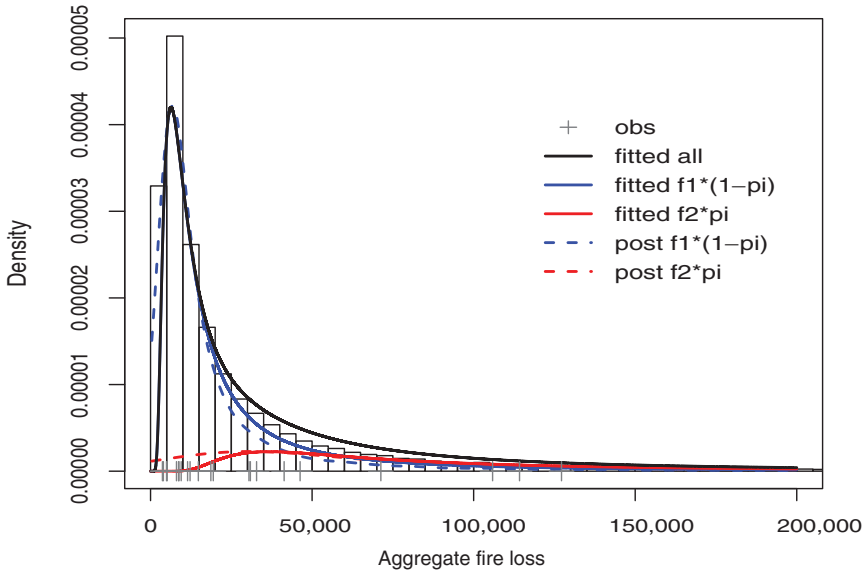


FIGURE 10: Histogram of the posterior predictive distributions using M1 for aggregate fire losses data, superimposed with the weighted densities of the predictive and fitted distributions. (Color online)

contaminated mixture distribution. We find that the contaminated mixture distributions outperform the other distributions in terms of model fit and VaR estimates. Details for the implementation of the two composite distributions can be found in Appendix G.

In summary, the proposed CGB2 distribution possesses properties that make it an attractive and widely applicable loss distribution. Future research can be directed to estimate contaminated mixture of composite distributions, combining the advantages of the contaminated and composite distributions. Moreover, efficient EM algorithms can be developed as alternative estimation methods for the proposed distributions.

REFERENCES

AIUPPA, T.A. (1988) Evaluation of Peason curves as an approximation of the maximum probable annual aggregate loss. *Journal of Risk and Insurance*, **55**, 425–441.

ARTZNER, P., DELBAEN, F., EBER, J.M. and HEATH, D. (1999) Coherent measures of risk. *Mathematical Finance*, **9**, 203–228.

BERGER, J. (2006) The case for objective Bayesian analysis, *Bayesian Analysis*, **1**, 385–402.

BRAZAUSKAS, V. and KLEEFELD, A. (2011) Folded and log-folded-t distributions as models for insurance loss data. *Scandinavian Actuarial Journal*, **2011**, 59–74.

CHAN, J.S.K., CHOY, S.T.B. and MAKOV, U.E. (2008) Robust Baysian analysis of loss reserves data using the generalized-t distribtuion. *ASTIN Bulletin*, **38**, 207–230.

CHAN, J.S.K., KUK, A.Y.C., BELL, J. and MCGILCHRIST, C. (1998) The analysis of methadone clinic data using marginal and conditional logistic models with mixture or random effects. *The Australian and New Zealand Journal of Statistics*, **40**, 1–10.

CIUMARA, R. (2006) An actuarial model based on the composite Weibull–Pareto distribution. *Mathematical Report-Bucharest*, **8**, 401–414.

COORAY, K. (2009) The Weibull–Pareto composite family with applications to the analysis of unimodal failure rate data. *Communications in Statistics-Theory and Methods*, **38**, 1901–1915.

COORAY, K. and ANANDA, M.M. (2005) Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, **2005**, 321–334.

CUMMINS, J.D. and FREIFELDER, L.R. (1978) A comparative analysis of alternative maximum probable yearly aggregate loss estimates. *Journal of Risk and Insurance*, **45**, 27–52.

CUMMINS, J.D., GEORGES, D. and MCDONALD, J.B. and 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. and PHILIPS, R.D. (1999) Pricing excess of loss reinsurance contracts against catastrophic loss. In *The Financing of Catastrophe Risk* (ed. K. Froot), pp. 93–143. Chicago: University of Chicago Press.

CUMMINS, J.D., MCDONALD, J.B. and CRAIG, M. (2007) Risk loss distributions and modelling the loss reserve pay-out tail. *Review of Applied Economics*, **3**, 1–23.

DONG, A.X. and CHAN, J.S.K. (2013) Bayesian analysis of loss reserving using dynamic models with generalized beta distribution. *Insurance: Mathematics and Economics*, **53**, 355–365.

DONG, A.X., CHAN, J.S.K. and PETERS, G.W. (2015) Risk margin quantile function via parametric and non-parametric approaches. *ASTIN Bulletin*, **45**, 503–550.

ELING, M. (2012) Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, **51**, 239–248.

EMBRECHTS, P., KLUPPELBERG, C. and MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Applications of Mathematics. Heidelberg: Springer.

ENGLAND, P.D. and VERRALL, R.J. (2006) Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, **1**(2), 221–270.

FREES, E.W. and VALDEZ, E.A. (2008) Hierarchical insurance claims modeling. *Journal of the American Statistical Assoication*, **103**, 1457–1469.

FREES, E.W., SHI, P. and VALDEZ, E.A. (2009) Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin*, **39**, 165–197.

GELMAN, A., CARLIN, J.B., STERN, H.S., DUNSON, D.B., VEHTARI, A. and RUBIN, D.B. (2014) *Bayesian Data Analysis*, Vol. 2. Boca Raton, FL: CRC Press.

GILKS, W.R., RICHARDSON, S. and SPIEGELHALTER, D.J. (1996) *Markov Chain Monte Carlo in Practice*. UK: Chapman and Hall.

JORION, P. (1997) *Value at Risk: The New Benchmark for Managing Financial Risk*, 1st Edition. New York: McGraw Hill.

KLEIBER, C. and KOTZ, S. (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*. Boboken, NJ: Wiley.

KLUGMAN, S.A., PANJER, H.H. and WILLMOT, G.E. (2008) *Loss Models for Data to Decisions*. Hoboken, NJ: Wiley.

LANDSMAN, Z. and MAKOV, U.E. (2003) Contaminated exponential dispersion loss models. *North American Actuarial Journal*, **7**, 116–127.

LANDSMAN, Z.M. and VALDEZ, E.A. (2005) Tail conditional expectations for exponential dispersion models. *ASTIN Bulletin*, **35**, 189–209.

MCDONALD, J.B. (1984) Some generalized functions for the size distribution of income. *Econometrica*, **52**, 647–665.

MCDONALD, J.B. and BUTLER, R.J. (1987) Some generalized mixture distributions with an application to unemployment duration. *The Review of Economics and Statistics*, **69**, 232–240.

MCDONALD, J.B. and NEWEY, W.K. (1988) Partially adaptive estimation of regression models via the generalized *t* distribution. *Econometric Theory*, **4**, 428–457.

NADARAJAH, S. and BAKAR, S.A.A. (2014) New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal*, **2014**, 180–187.

NADARAJAH, S. and BAKAR, S.A.A. (2015) New folded models for the log-transformed Norwegian fire claim data. *Communications in Statistics-Theory and Methods*, **44**, 4408–4440.

PAULSON, A.S. and FARIS, N.J. (1985) A practical approach to measuring the distribution of total annual claims. In *Strategic Planning and Modeling in Property-Liability Insurance* (ed. J.D. Cumins), pp. 205–223. Norwell, MA: Kluwer Academic Publishers.

RAMLAU-HANSEN, H. (1988) A solvency study in non-life insurance. Part 1. Analysis of fire, windstorm, and glass claims. *Scandinavian Actuarial Journal*, **1988**, 3–34.

SCOLLNIK, D.P. (2007) On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, **2007**, 20–33.

SCOLLNIK, D.P. and SUN, C. (2012) Modeling with Weibull–Pareto models. *North American Actuarial Journal*, **16**, 260–272.

SMITH, A.F.M. and 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. and VAN DER LINDE, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, Series B, **64**, 583–639.

TAYLOR, G.C. (2000) *Loss Reserving: An Actuarial Perspective*. Boston: Kluwer Academic Publishers.

TUKEY, J.W. (1960) A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* (eds. J. Olkin, W. Hoeffding, S. Ghurye, W. Madow and H. Mann), pp. 448–485. Stanford, CA: Stanford University Press.

VENTER, G.C. (1983) *Transformed beta and gamma functions and aggregate losses*. Proceedings of the Acsuality Actuarial Society, Vol. 71. Recording and Statistical Corporation, Boston, MA.

YANG, X., FREES, E.W. and ZHANG, Z. (2011) A generalized beta copula with applications in modeling mulitivariate long-tailed data. *Insurance: Mathematics and Economics*, **49**, 265–284.

J.S.K. CHAN (Corresponding author)
School of Mathematics and Statistics
The University of Sydney
NSW 2006, Australia
Actuarial Research Center
University of Haifa
Haifa, Israel
E-mail: jchan@maths.usyd.edu.au

S.T.B. CHOY
Disciplines of Business Analytics
The University of Sydney
NSW 2006, Australia
Actuarial Research Center
University of Haifa
Haifa, Israel
E-mail: boris.choy@sydney.edu.au

U.E. MAKOV
Department of Statistics
Actuarial Research Center
University of Haifa
Haifa, 31905, Israel
E-mail: makov@stat.haifa.ac.il

Z. LANDSMAN
Department of Statistics
Actuarial Research Center
University of Haifa
Haifa, 31905, Israel
E-mail: landsman@stat.haifa.ac.il

APPENDIX A

The ratio of gamma functions is

$$c_h = \frac{\Gamma(p+\frac{h}{a})\Gamma(q-\frac{h}{a})}{\Gamma(p)\Gamma(q)} \simeq \frac{\sqrt{2\pi(p-1+\frac{h}{a})}\left(\frac{p-1+\frac{h}{a}}{e}\right)^{p-1+\frac{h}{a}}\sqrt{2\pi(q-1-\frac{h}{a})}\left(\frac{q-1-\frac{h}{a}}{e}\right)^{q-1-\frac{h}{a}}}{\sqrt{2\pi(p-1)}\left(\frac{p-1}{e}\right)^{p-1}\sqrt{2\pi(q-1)}\left(\frac{q-1}{e}\right)^{q-1}},$$

$$= \frac{\left(p-1+\frac{h}{a}\right)^{p-\frac{1}{2}+\frac{h}{a}}\left(q-1-\frac{h}{a}\right)^{q-\frac{1}{2}-\frac{h}{a}}}{(p-1)^{p-\frac{1}{2}}\,(q-1)^{q-\frac{1}{2}}},$$

since $\Gamma(h+1) \simeq \sqrt{2\pi h}\left(\frac{h}{e}\right)^h$ from *Stirling's* approximation.

APPENDIX B

TABLE B1

DISTRIBUTIONS NESTED WITHIN THE GB2 DISTRIBUTION FAMILY.

M	P	Distribution			pdf
1	4	Generalised Beta 2 (GB2)		GB2(a, b, p, q)	$\frac{\frac{\lvert a\rvert}{b}(\frac{y}{b})^{ap-1}}{B(p,q)[1+(\frac{y}{b})^a]^{p+q}}$
2	3	Beta Prime	BP(p, q, b)	GB2$(1, b, p, q)$	$\frac{b^q y^{p-1}}{B(p,q)(b+y)^{p+q}}$
3	3	Burr 7 or 12 (Singh–Maddala)	B(a, q, b)	GB2$(a, b, 1, q)$	$\frac{aqb^{aq}y^{a-1}}{(b^a+y^a)^{q+1}}$
4	3	Inverse Burr (Burr 3, Dagum)	IB(a, p, b)	GB2$(a, b, p, 1)$	$\frac{apb^a y^{ap-1}}{(b^a+y^a)^{p+1}}$
5	3	Generalised gamma†	GG(p, β, a)	GB2$(a, \beta q^{\frac{1}{a}}, p, \infty)$	$\frac{ay^{ap-1}}{\beta^{ap}\Gamma(p)}\exp\left[-\left(\frac{y}{\beta}\right)^a\right]$

TABLE B1

CONTINUED.

M	P	Distribution			pdf
6	3	Log-t‡	LT(μ, σ^2, ν)	GB2$(0^+, \beta^* q^{\frac{1}{a}}, p^*, \frac{\nu}{2})$	$\frac{\Gamma(\frac{\nu+1}{2})}{y\sqrt{\nu\pi}\sigma\Gamma(\frac{\nu}{2})}\left[1+\frac{(\ln y-\mu)^2}{\nu\sigma^2}\right]^{-\frac{\nu+1}{2}}$
7	2	Variance Ratio (F)§	VR$(2p, 2q)$	GB2$(1, \frac{q}{p}, p, q)$	$\frac{p^p y^{p-1}}{q^p B(p,q)(1+\frac{py}{q})^{p+q}}$
8	2	Pareto 2 (Lomax)	P(q, b)	GB2$(1, b, 1, q)$	$\frac{qb^q}{(b+y)^{q+1}}$
9	2	Inverse Pareto	IP(p, b)	GB2$(1, b, p, 1)$	$\frac{pby^{p-1}}{(b+y)^{p+1}}$
10	2	Loglogistic (Fisk)	LL(a, b)	GB2$(a, b, 1, 1)$	$\frac{ab^a y^{a-1}}{(b^a+y^a)^2}$
11	2	Paralogistic	PL(a, b)	GB2$(a, b, 1, a)$	$\frac{a^2 b^a y^{a-1}}{(b^a+y^a)^{a+1}}$
12	2	Inverse Paralogistic	IPL(a, b)	GB2$(a, b, a, 1)$	$\frac{a^2 b^a y^{a-1}}{(b^a+y^a)^{a+1}}$
13	2	Gamma†	GA(p, β)	GB2$(1, \beta q, p, \infty)$	$\frac{y^{p-1}}{\beta^p\Gamma(p)}e^{-\frac{y}{\beta}}$
14	2	Weibull†	W(a, β)	GB2$(a, \beta q^{\frac{1}{a}}, 1, \infty)$	$\frac{ay^{a-1}}{\beta^a}\exp\left[-\left(\frac{y}{\beta}\right)^a\right]$
15	2	Log-Normal‡	LN(μ, σ^2)	GB2$(0^+, \beta^* q^{\frac{1}{a}}, p^*, \infty)$	$\frac{1}{y\sigma\sqrt{2\pi}}\exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right]$
16	2	Log-Cauchy‡	LC(μ, σ^2)	GB2$(0^+, \beta^* q^{\frac{1}{a}}, p^*, \frac{1}{2})$	$\frac{1}{y\pi\sigma}\left[1+\frac{(\ln y-\mu)^2}{\sigma^2}\right]^{-1}$
17	1	Chi-Square§	CS$(2p)$	GB2$(1, 2q, p, \infty)$	$\frac{y^{p-1}}{2^p\Gamma(p)}\exp\left(-\frac{y}{2}\right)$
18	1	Exponential†	EXP(β)	GB2$(1, \beta q, 1, \infty)$	$\frac{1}{\beta}\exp\left(-\frac{y}{\beta}\right)$
19	1	Rayleigh†	R(β)	GB2$(2, \beta q^{\frac{1}{2}}, 1, \infty)$	$\frac{y}{\beta^2}\exp\left(-\frac{y^2}{2\beta^2}\right)$
20	1	Half-t	HT$(0, \nu, \nu)$	GB2$(2, \nu, \frac{1}{2}, \frac{\nu}{2})$	$\frac{2}{\nu B(\frac{\nu}{2}, \frac{1}{2})}\left(1+\frac{y^2}{\nu^2}\right)^{-\frac{1+\nu}{2}}$
21	1	Half-Normal‡	HN$(0, \sigma^2)$	GB2$(2, \sigma(2q)^{\frac{1}{2}}, \frac{1}{2}, \infty)$	$\frac{2}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y^2}{2\sigma^2}\right)$

Note: $\beta^* = (a\sigma)^{\frac{2}{a}}$, $p^* = \frac{a\mu+1}{(a\sigma)^2}$; Contaminated scale: †: $k\beta$; ‡: $k\sigma$; §: kq.

APPENDIX C

The TCE for GB2 and CGB2 distributions are

$$\begin{aligned}
\text{TCE}_{\text{GB2},\rho} &= \frac{1}{1-\rho}\int_\rho^\infty F^{-1}_{\text{GB2}}(\alpha)d\alpha = \frac{1}{1-\rho}\int_\rho^\infty F^{-1}_{\text{GB2}}(F_{\text{GB2}}(y))dF_{\text{GB2}}(y) \\
&= \frac{1}{1-\rho}\int_{F^{-1}_{\text{GB2}}(\rho|a,b,p,q)}^\infty y f_{\text{GB2}}(y|a,b,p,q)dy \\
&= \frac{a}{(1-\rho)B(p,q)}\int_{F^{-1}_{\text{GB2}}(\rho|a,b,p,q)}^\infty \frac{(\frac{y}{b})^{a_t p}}{[1+(\frac{y}{b})^a]^{p+q}}dy
\end{aligned}$$

$$= \frac{b}{(1-\rho)\,B(p,q)} \int_{F^{-1}_{\mathrm{GB2}}(\rho|a,b,p,q)}^{\infty} \frac{(\frac{y}{b})^{ap-a+1}}{[1+(\frac{y}{b})^a]^{p+q-2}} \frac{a}{b} \frac{(\frac{y}{b})^{a-1}}{[1+(\frac{y}{b})^a]^2} dy$$

$$= \frac{b\,B(p+\frac{1}{a}, q-\frac{1}{a})}{(1-\rho)\,B(p,q)} \times \frac{\int_{F_B^{-1}(\rho|p,q)}^{\infty} z^{p-1+\frac{1}{a}}(1-z)^{q-1-\frac{1}{a}}}{B(p+\frac{1}{a}, q-\frac{1}{a})} dz$$

$$= \frac{b\,B(p+\frac{1}{a}, q-\frac{1}{a})}{B(p,q)} \times \frac{1 - F_B\Big(F_B^{-1}(\rho|p,q)\,\Big|\,p+\frac{1}{a}, q-\frac{1}{a}\Big)}{1-\rho}$$

and

$$\mathrm{TCE}_{\mathrm{CGB2},\rho} = \frac{1-\pi}{1-\rho} \int_{F^{-1}_{\mathrm{GB2}}(\rho|a,b,p,q)}^{\infty} y f_{\mathrm{GB2}}(y|a,b,p,q)dy + \frac{\pi}{1-\rho} \int_{F^{-1}_{\mathrm{GB2}}(\rho|a,kb,p,q)}^{\infty} y f_{\mathrm{GB2}}(y|a,kb,p,q)dy$$

$$= (1-\pi+k\pi)\frac{b\,B(p+\frac{1}{a}, q-\frac{1}{a})}{B(p,q)} \frac{1 - F_B\Big(F_B^{-1}(\rho|p,q)\,\Big|\,p+\frac{1}{a}, q-\frac{1}{a}\Big)}{1-\rho}.$$

APPENDIX D

The first-order derivative of $f_{\mathrm{GB2}}(y|a,b,p,q)$ and $f_{\mathrm{CGB2}}(y|p,q,a,b,\pi,k)$ are

$$f'_{\mathrm{GB2}}(y|p,q,a,b) = \frac{a}{b^2\,B(p,q)} \frac{(\frac{y}{b})^{ap-2}[(ap-1)-(\frac{y}{b})^a(aq+1)]}{[1+(\frac{y}{b})^a]^{p+q+1}}, \tag{D1}$$

$$\begin{aligned} & f'_{\mathrm{CGB2}}(y|p,q,a,b,\pi,k) \\ &= \frac{a}{B(p,q)} \left\{ \frac{\pi y^{ap-2}[(ap-1)-y^a(aq+1)]}{(1+y^a)^{p+q+1}} + \frac{(1-\pi)\frac{1}{k^2}(\frac{y}{k})^{ap-2}[(ap-1)-\frac{y^a}{k^a}(aq+1)]}{(1+\frac{y^a}{k^a})^{p+q+1}} \right\} \\ &\propto \frac{\pi y^{ap-2}(1+\frac{y^a}{k^a})^{p+q+1}[(ap-1)-y^a(aq+1)] + (1-\pi)\frac{1}{k^2}(\frac{y}{k})^{ap-2}(1+y^a)^{p+q+1}[(ap-1)-\frac{y^a}{k^a}(aq+1)]}{[(1+y^a)(1+\frac{y^a}{k^a})]^{p+q+1}} \\ &\propto \pi(k^a+y^a)^{p+q+1}[g-y^a] + (1-\pi)k^{aq}(1+y^a)^{p+q+1}(gk^a-y^a), \end{aligned} \tag{D2}$$

where $b=1$ for standardised distribution and $g=\frac{ap-1}{aq+1}$.

APPENDIX E

OpenBUGS code for CGB2 distribution.

```
#Model 1 - CGB2
model {
for (i in 1:N){
dummy[i]<-0
dummy[i]~dloglik(logl[i])
logl1[i]<-log(a)+(a*p-1)*(log(y[i])-log(b))-log(b)-logbeta
          -(p+q)*log(1+pow(y[i]/b,a))
logl2[i]<-log(a)+(a*p-1)*(log(y[i])-log(b)-log(k))-log(b)-log(k)-logbeta
          -(p+q)*log(1+pow(y[i]/b/k,a))
logl[i]<-log((1-pi)*exp(logl1[i]) + pi*exp(logl2[i]))
}
logbeta <- loggam(p)+loggam(q)-loggam(p+q)
loglike <- sum(logl[])

# Prior Distributions
a ~ dgamma(0.04,0.01)I(0.1,10)
b ~ dgamma(1,0.001)I(0.01,10000)
p ~ dgamma(0.05,0.01)I(1,10000)
q ~ dgamma(0.01,0.01)I(0.01,10)
pi ~ dbeta(1,1)I(0.001,0.5)
k ~ dgamma(0.02,0.01)I(2,50)

# Predictive distribution
v ~ dbern(pi)
u ~ dbeta(p,q)I(0.001,0.999)
u1 <- u/(1-u)
a1 <- 1/a
ypred <- b*pow(u1,a1)*(1-v) + b*k*pow(u1,a1)*v
}

#initial values
list(a=1.3,b=3000,p=4.2,q=0.85,pi=0.1,k=5,u=0.5,v=0)

list(y=c(290.4,537.19,756.8,769.19,787.69,796.18,933.62,967.97,1010.56,
1017.4,1033.49,1034.33,1056.93,1124.09,1165.73,1248.49,1268.24,1284.56,
1363.85,1436.2,1445.96,1469.48,1507.47,1662.36,1674.58,1690.91,1739.96,
1776.56,1932.09,1975.89,2099.79,1217.64,2202.96,2222.8,2255.72,2274.61,
2328.64,2384.37,2847.83,2947.04,2948.35,3036.51,3287.68,3331.62,3416.67,
3604.66,3671.16,3739.3,3941.3,4017.01,4100,4166.98,4355.02,5117.93,5335.96,
5453.02,5568.96,5761.83,6161.81,6348.69,6859.37,7972.2,8028.32,10047.22,
10560.1,11179.54,11461.39,14538.13,14789.81,17186.09,18582.57,22857.33,
23177.85,23446.13,28409.82,57612.82,59582.78,113164.7,123228.9,626402.8))
#aggregated claims
```

```
list(y=c(71280,3671,18664,8784,3966,30892,631626,11464,127194,4950,30452,
8028,14790,9480,8676,114198,5150,105864,32814,41340,46284,12230,19418))
```

APPENDIX F

Apart from the Bayesian approach, the EM approach can also be used to estimate the model parameters together with the missing group membership indicators I_i for each observation y_i. Taking CGB2 distribution as an example, the missing I_i can be estimated iteratively using conditional expectation

$$\text{E-step:}\quad \hat{I}_i = E(I_i|\widehat{\boldsymbol{\theta}}) = \frac{\pi f_{\text{GB2}}(y_i|\hat{a}, \hat{k}\hat{b}, \hat{p}, \hat{q})}{(1-\hat{\pi}) f_{\text{GB2}}(y_i|\hat{a}, \hat{b}, \hat{p}, \hat{q}) + \pi f_{\text{GB2}}(y_i|\hat{a}, \hat{k}\hat{b}, \hat{p}, \hat{q})},$$

where $f_{\text{GB2}}(\cdot)$ is given in (5). In the M-step, model parameters are estimated by maximising the complete data log-likelihood function

$$\text{M-step:}\quad \widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}|\boldsymbol{y}, \hat{\boldsymbol{I}}),$$

where

$$\begin{aligned}
\ell(\boldsymbol{\theta}|\boldsymbol{y}, \hat{\boldsymbol{I}}) &= \ln\left\{\prod_{i=1}^{n}\prod_{j=0}^{1}[\pi_j f_{\text{GB2}}(y_i|a, b_j^*, p, q)]^{\hat{I}_{ij}}\right\} \\
&= \sum_{i=1}^{n}\sum_{j=0}^{1}\hat{I}_{ij}[\log\pi_i + \log f_{\text{GB2}}(y_i|a, b_j^*, p, q)] \\
&= n\left[\ln(1-\pi) + \ln|a| + \sum_{i=1}^{n}\ln y_i - ap\ln b - \ln B(p,q)\right] \\
&\quad + [\ln\pi - \ln(1-\pi) - ap\ln k]\sum_{i=1}^{n}\hat{I}_i \\
&\quad - (p+q)\left\{\sum_{i=1}^{n}(1-I_i)\ln\left[1+\left(\frac{y_i}{b}\right)^a\right] + \sum_{i=1}^{n} I_i \ln\left[1+\left(\frac{y_i}{kb}\right)^a\right]\right\},
\end{aligned}$$

$\pi_1 = \pi, \pi_0 = 1-\pi$, $I_{i1} = I_i$, $I_{i0} = 1-I_i$, $b_1^* = kb$, $b_0^* = b$, $\boldsymbol{y}$ is a vector of y_i and $\hat{\boldsymbol{I}}$ is a vector of estimated group membership indicator $\hat{I}_i$.

APPENDIX G

Cooray and Ananda (2005) express the density function for the composite LP model in the form of (2) with $f_1(y) = f_{\text{LN}}(y|\mu, \sigma^2)$ and $f_2(y) = f_{\text{P}}(y|a, \theta)$. The two constraints in (3) give

$$\ln\theta - \mu = a\sigma^2 \Rightarrow \mu = \ln\theta - k_1^2/a, \text{ and}$$

$$\exp(-a^2\sigma^2) = 2\pi a^2\sigma^2 \Rightarrow \exp(-k_1^2) = 2\pi k_1^2 \Rightarrow k_1 \approx 0.372238898,\ \sigma = k_1/a.$$

Hence, $F_{\text{LN}}(y) = \Phi(\frac{\ln y-\mu}{\sigma}) = \Phi(\frac{a}{k_1}\ln(\frac{y}{\theta})+k_1)$ and $c = [1+\Phi(k_1)]^{-1}$, where $\Phi(k_1) = F_{\text{LN}}(\theta)$. By making use of the shifted Pareto distribution such that $f_{\text{P}}(y) = 0$ if $y < \theta$, (3) can be expressed, in general, as

$$f(y) = \begin{cases} W\dfrac{f_{\text{LN}}(y)}{F_{\text{LN}}(\theta)}, & \text{if } 0 \le y \le \theta, \\ (1-W)f_{\text{P}}(y), & \text{if } \theta \le y \le \infty, \end{cases} \tag{G1}$$

where $W = F_{\text{LP}}(\theta) = F_{\text{LN}}(\theta)/(1+F_{\text{LN}}(\theta)) = \Phi(k_1)/[1+\Phi(k_1)] = 0.3921499$. The fixed quantile at $y = \theta$ corresponds to the quantile $F_{\text{LN}}(\theta) = 0.6451425$ for lognormal distribution. This gives an interpretation of the composite distribution as a mixture of truncated lognormal and shifted Pareto distributions. In summary, the model LP(a, θ) has pdf

$$f_{\text{LP}}(y) = \begin{cases} \dfrac{a\theta^a}{[1+\Phi(k_1)]y^{a+1}}\exp\left[-\dfrac{a^2}{2k_1^2}\ln^2\left(\dfrac{y}{\theta}\right)\right], & \text{if } 0 < y \le \theta, \\ \dfrac{a\theta^a}{[1+\Phi(k_1)]y^{a+1}}, & \text{if } y \ge \theta, \end{cases}$$

and cdf

$$F_{\text{LP}}(y) = \begin{cases} \dfrac{\Phi(\frac{a}{k_1}\ln(\frac{y}{\theta})+k_1)}{1+\Phi(k_1)}, & \text{if } 0 < y \le \theta, \\ 1-\dfrac{(\frac{\theta}{y})^a}{1+\Phi(k_1)}, & \text{if } y \ge \theta. \end{cases}$$

The composite WP model (Ciumara, 2006 and Cooray, 2009) can be written in the form of (2) with $f_1(y) = f_{\text{W}}(y|r,\lambda)$ and $f_2(y) = f_{\text{P}}(y|a,\theta)$. The two constraints in (3) give

$$1+\frac{r}{a} = \exp\left(1+\frac{a}{r}\right) \Rightarrow 1+k_2 = \exp\left(1+\frac{1}{k_2}\right) \Rightarrow k_2 \approx 2.8573348,\ r = k_2 a \text{ and}$$

$$1+\frac{r}{a} = 1+k_2 = \exp\left[\left(\frac{\theta}{\lambda}\right)^r\right] \Rightarrow \lambda = \frac{\theta}{[\ln(k_2+1)]^{1/r}}.$$

Hence, $F_{\text{W}}(\theta) = 1-\exp\left[-\left(\frac{\theta}{\lambda}\right)^r\right] = k_2/(k_2+1) = 0.74075$ and $c = [1+F_{\text{W}}(\theta)]^{-1} = (k_2+1)/(2k_2+1) = 0.5744638$. Again the model can be written as a truncated Weibull and shifted Pareto mixture with mixture weight $W = F_{\text{WP}}(\theta) = F_{\text{W}}(\theta)/[1+F_{\text{W}}(\theta)] = c\Phi(k_1) = k_2/(2k_2+1) = 0.4255362$ and the fixed quantile at $y = \theta$ corresponds to the quantile $F_{\text{W}}(\theta) = 0.74075$ for Weibull distribution. The pdf of WP(a, θ) model is

$$f_{\text{WP}}(y) = \begin{cases} \dfrac{(k_2+1)^2 a}{(2k_2+1)y}\left(\dfrac{y}{\theta}\right)^{ak_2}\exp\left[-\left(\dfrac{k_2+1}{k_2}\right)\left(\dfrac{y}{\theta}\right)^{ak_2}\right], & \text{if } 0 < y \le \theta, \\ \dfrac{(k_2+1)a}{(2k_2+1)y}\left(\dfrac{\theta}{y}\right)^a, & \text{if } y \ge \theta, \end{cases}$$

and the cdf is

$$F_{\text{WP}}(y) = \begin{cases} \dfrac{k_2+1}{2k_2+1}\left\{1-(k_2+1)\exp\left[-\left(\dfrac{y}{\theta}\right)^{ak_2}\right]\right\}, & \text{if } 0 < y \le \theta, \\ 1-\dfrac{k_2+1}{2k_2+1}\left(\dfrac{\theta}{y}\right)^a, & \text{if } y \ge \theta. \end{cases}$$