Insurance: Mathematics and Economics 9 (1990) 257-272
North-Holland

Applications of the GB2 family of distributions in modeling insurance loss processes *

J. David CUMMINS
University of Pennsylvania, Philadelphia, PA 19104, USA

Georges DIONNE
University of Montreal, Montreal, Que. H3C 3J1, Canada

James B. McDONALD
and B. Michael PRITCHETT
Brigham Young University, Provo, UT 84602, USA

Received November 1988
Revised January 1990

This paper investigates the use of a four parameter family of probability distributions, the generalized beta of the second kind (GB2), for modeling insurance loss processes. The GB2 family includes many commonly used distributions such as the lognormal, gamma and Weibull. The GB2 also includes the Burr and generalized gamma distributions. Members of this family and their inverse distributions have significant potential for improving the distributional fit in many applications involving thin or heavy-tailed distributions. Members of the GB2 family can be generated as mixtures of well-known distributions and provide a model for heterogeneity in claims distributions. Examples are presented which consider models of the distribution of individual and of aggregate losses. The results suggest that seemingly slight differences in modeling the tails can result in large differences in reinsurance premiums and quantiles for the distribution of total insurance losses.

Keywords: Generalized beta of the first kind, Generalized beta of the second kind, Log *t*; Generalized gamma, Pearson, Burr, Kappa, Pareto, Lomax, Log Cauchy, Lognormal, Beta, Gamma, Weibull, Fisk, Rayleigh; Uniform, Exponential, Inverse distributions, Maximum likelihood estimation, Maximum probable yearly aggregate loss, Mean residual life, Insurance loss processes, Reinsurance.

* The authors appreciate the excellent comments received from Stuart Klugman and two anonymous referees who also made substantive comments and suggestions.
In early 1988 the editor encountered two similar papers. One was 'Distribution of Aggregate Loss' by Pritchett and McDonald and the other was 'Applications of the GB2 Family of Probability Distributions in Collective Risk Theory' by Cummins, Dionne and Maistre. At the request of the editor the two papers were combined.

1. Introduction

One of the problems in collective risk theory is the estimation of the distribution of total claims, $X = X_1 + \cdots + X_N$, where N is the number (frequency) of claims and X_i is the amount (severity) of the ith claim. The random variables N, X_1, X_2, ... are assumed to be mutually independent. Estimation of the distribution function, $F(x)$, usually involves fitting the frequency and severity distributions, testing for goodness of fit, and then combining the distributions to yield $F(x)$. The model of $F(x)$ is then used to estimate premiums, risk loadings, reinsurance premiums, and other decision variables such as the maximum probable yearly aggregate loss (MPY).

Distributions used for frequency include the Poisson, the negative binomial and the logarithmic series distribution [see for example, Ferreira (1970), Seal (1969), and Cummins and Wiltbank (1983)]. Among the distributions that have been considered for severity are the exponential, gamma, log-gamma, lognormal, Pareto, and log-*t* [Beard, Pentikainen and Pesonen (1965), Benckert (1962), Cummins and Freifelder (1978), Cummins and Wiltbank (1983), Hewitt (1970), Mandlebrot (1964), Paulson (1984), Seal (1969), and Shpilberg (1977)].

Traditionally, calculating $F(x)$ directly was considered a difficult problem, and various approximation formulas, such as the normal-power and gamma, received considerable attention [see, for example, Beard, Pentikainen, and Pesonen [1984)]. In recent years, developments in risk theory and advances in computing have made virtually exact calculation of $F(x)$ quite feasible. Prominent approaches are simulation [e.g., Roy and Cummins (1985)], fast Fourier transforms (FFTs) [Paulson (1984)], and the Panjer recursion algorithm [Panjer (1981) and Panjer and Willmot (1987)]. An approach often used in the United States is numerical inversion of the characteristic function [Heckman and Myers (1983), Paulson and Dixit (1989)].

Paralleling the developments in the calculation of $F(x)$ have been more sophisticated approaches to the estimation of the frequency and severity distributions. The traditional approach was to use tractable one or two parameter distributions such as those mentioned above. In many cases, these distributions were selected not necessarily because they were in some sense 'best' but because of the feasibility of estimating parameters and computing quantiles. Because insurance claims distributions are often heavy-tailed, restricting the set of candidate distributions considered in modeling can lead to serious underestimation of tail quantiles, reinsurance premiums, and other variables [Cummins and Freifelder (1978)].

Recent advances have opened up a much wider range of probability distributions for use in modeling insurance claims processes. Hogg and Klugman (1984) discuss many alternative models for loss distributions as well as related issues of estimation and inference. Paulson has made extensive use of the stable family of distributions, which includes the one-tail Pareto, the normal, and the Cauchy distributions, among others, as special cases [e.g., Paulson and Faris (1985)]. Aiuppa (1988) utilizes a computer program that can estimate parameters and compute percentile points of the distribution functions for any member of the Pearson family. [1] McDonald (1984) considers generalizations of the beta of the first type and of the second type which include the Pearson Type I and VI as special cases. These generalized beta distributions will be denoted GB1 and GB2, respectively. Venter (1984) introduced the GB2 in the actuarial literature as the transformed beta. Applications of the GB1 and GB2 in the economics literature are described in McDonald (1984), McDonald and Butler (1987), and McDonald and Richards (1987).

The purpose of this article is to investigate the use of the GB2 family as a model for continuous distributions in 'non-life' insurance. Thus, the GB2 is proposed as a potential model for both aggregate losses and loss severity. The GB2 provides an extremely flexible functional form that can be used to model highly skewed loss distributions such as those typically observed in 'non-life' insurance. The use of the GB2 is illustrated below by modeling fire losses of a major university using the Cummins and Freifelder (1978) data.

The paper is organized as follows: The generalized distributions are discussed in Section 2. Two applications of these distributions are then considered: estimating the distribution of aggregate losses in Section 3, and estimating the severity distribution in Section 4. Severity distributions are estimated with both grouped and ungrouped data to explore the issue of accuracy loss from using grouped data. Section 5 discusses the impact of model selection and estimation techniques on severity quantiles, reinsurance premiums, and simulated total claims distributions. Section 6 concludes the paper.

2. The generalized distributions

2.1. The density functions

It is useful to define three very flexible distributions: the generalized gamma (GG), the generalized beta of the first kind (GB1) and the generalized beta of the second kind (GB2). The density functions are:

$$GG(x;\, a,\, b,\, p) = \frac{|a|\, x^{ap-1}\, \mathrm{e}^{-(x/b)^a}}{b^{ap}\Gamma(p)}, \quad 0 \le x, \qquad = 0, \quad \text{otherwise}; \tag{1}$$

$$GB1(x;\, a,\, b,\, p,\, q) = \frac{|a|\, x^{ap-1}\left(1-(x/b)^a\right)^{q-1}}{b^{ap}B(p,\, q)}, \quad 0 \le x^a \le b^a, \qquad = 0, \quad \text{otherwise}; \tag{2}$$

and

$$GB2(x;\, a,\, b,\, p,\, q) = \frac{|a|\, x^{ap-1}}{b^{ap}B(p,\, q)\left(1+(x/b)^a\right)^{p+q}}, \quad 0 \le x, \qquad = 0, \quad \text{otherwise}; \tag{3}$$

where b, p and $q > 0$ and $a \neq 0$.

[1] This includes the Pearson Type IV, which statisticians have traditionally considered intractable [see, for example, Johnson and Kotz (1970, p. 12)].

2.2. Interpretation of parameters

The generalized gamma is a three-parameter distribution and is a limiting case of both the four-parameter GB1 and GB2 distributions. The parameters in these distributions determine the shape and location of the density. The parameter 'b' is a scale factor; 'b' is also an upper (lower) bound for GB1 variables as the parameter 'a' is positive (negative). Unlike the GB1 or the beta of the first kind (B1), which is mentioned in risk theory texts [for B1 see Buhlmann (1970)], the GB2 has no upper limit and hence is likely to be applicable for severity distributions and other risk theory applications where the upper tail has no theoretical boundary. For the GB1 and GG, $E(X^h)$ exists when $-p < h/a$. For the GB2, $-p < h/a < q$ is required. The possibility of 'a' negative for the GG admits all orders of negative moments and limited positive moments and thus is a good candidate for loss severity.

The GB2 provides models for distributions characterized by thick tails. The relationship between the density and parameters is complex, but generally speaking the larger the value of the parameter 'a', the thinner the tails of the density function. Negative values of 'a' yield inverse distributions. The relative values of 'p' and 'q' are important in determining the skewness of the distribution and the GB2 permits positive as well as negative skewness.

Additional insight about the GG, GB1 and GB2 can be gained from the following construction. Let X and Y denote independent gamma random variables having common scale parameters of one and shape parameters p and q. It can be shown that

$$Z_1 = bX^{1/a} \sim GG(\cdot;\ a,\ b,\ p),$$

$$Z_2 = b\left(\frac{X}{X+Y}\right)^{1/a} \sim GB1(\cdot;\ a,\ b,\ p,\ q),$$

$$Z_3 = b(X/Y)^{1/a} \sim GB2(\cdot;\ a,\ b,\ p,\ q).$$

Hence, the GG, GB1 and GB2 variates are easily expressed in terms of simple transformations of independent gammas. Several distributional characteristics can be inferred from these relationships. For $q > 1$ the mean of X/Y is $p/(q-1)$. The mode of X/Y is $(p-1)/(q+1)$ for $p > 1$ and zero otherwise. The distinction between the GG and GB2 distributions can be thought of as the division of X by Y in the construction of Z_3. Thus, the tail of X/Y will be thinner (heavier) relative to the tail of X alone as the parameter q is large (near or less than one). The parameter b is seen to be a scale parameter in all three constructions.

2.3. Moments and distribution functions

The moments and expressions for the distributions of the GG, GB1, BR3, BR12 and GB2 are given in Table 1. The expressions $P(a, z)$ and $I_z(a, b)$, respectively, denote the incomplete gamma and beta functions. These functions can be approximated using numerical integration or approximations of series representations and can be used in determining quantiles. Abramowitz and Stegun (1965) provide an excellent source of references. The Burr 12 (GB2 with $p = 1$) and Burr 3 (GB2 with $q = 1$) have closed form expressions for the distribution functions. These are given in Table 1.

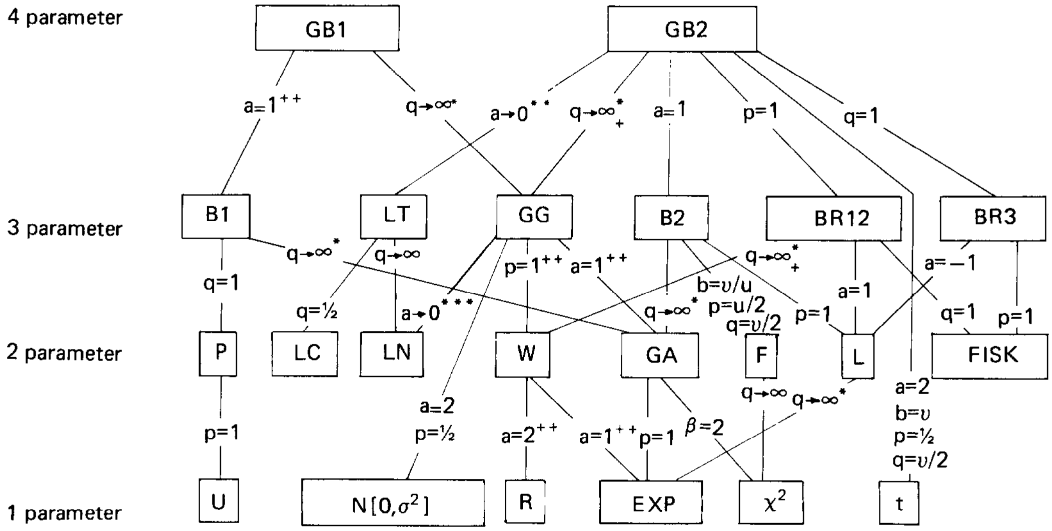
2.4. Relationships with other distributions

The flexibility of these distributions can also be illustrated using Fig. 1 [from McDonald and Richards (1987) or McDonald and Butler (1987)]. The GB2 is seen to include the log-t (LT), generalized gamma (GG), beta of the second kind (B2), Burr types 3 and 12 (BR3 and BR12), log-Cauchy (LC), lognormal (LN), Weibull (W), gamma (GA), variance ratio (F), Lomax or shifted Pareto (L),

Table 1

Distribution and moments.

Model	Distribution function	Moments
GG	$P(p, z=(x, b)^a)$	$\frac{b^h\Gamma(p+h/a)}{\Gamma(p)}$
BR3	$1/(1+(b/x)^a)^p$	$b^h pB(p+h/a, 1-h/a)$
BR12	$1-(1+(x/b)^a)^{-q}$	$b^h qB(1+h/a, q-h/a)$
GB1	$I_z(p, q)$, $z=(x/b)^a$	$\frac{b^h B(p+h/a, q)}{B(p, q)}$
GB2	$I_z(p, q)$, $z=\frac{(x/b)^a}{1+(x/b)^a}$	$\frac{b^h B(p+h/a, q-h/a)}{B(p, q)}$



* $q \to \infty$ with $b = q^{1/a}\beta$

** $a \to 0$ with $b = (q\sigma^2 a^2)^{1/a}$, $p = (a\mu + 1)/\sigma^2 a^2$

*** $a \to 0$ with $b = (\sigma^2 a^2)^{1/a}$, $p = (a\mu + 1)/\sigma^2 a^2$

\+ If $p \to \infty$ with $b = p^{1/a}\beta$, the inverse form of the distribution is obtained.

\+\+ The inverse form is obtained if the sign of a is changed.

Fig. 1. Distribution family tree.

Fisk or loglogistic, the half normal (N), Rayleigh (R), exponential (EXP), Chi-square and half-t distributions (t) as special or limiting cases. The GB1 includes the beta of the first kind (B1), power (P), uniform (U) and generalized gamma with related distributions appearing as special or limiting cases.

When the parameter 'a' is negative, (1), (2) and (3) admit inverse distributions. These are distributions of the random variable Y, obtained by making the reciprocal transformation $Y = 1/X$. It is easily shown that $GB2(x;\ -a,\ b,\ p,\ q) = GB2(x;\ a,\ b,\ q,\ p)$; hence the density is unaltered if p and q are interchanged if the sign of 'a' is also changed. Other examples of 'inverse' distributions are the inverse Burr and the inverse gamma, which includes the Pearson Type V as a special case, as well as inverse functions of other special and limiting cases related to the generalized gamma. Thus, the GB1 and GB2 include many members of the Pearson family of distributions [Ord (1972) and Elderton (1969)]. Other models in the GB2 family, which are not Pearson distributions, include the Burr types 3 and 12, Fisk, Weibull, generalized gamma and log-t. Since moving down the 'distribution tree' corresponds to special and limiting cases of the more general distribution, it follows that the more general distributional form will provide at least as good a fit as any of its special cases. Additional details are included in McDonald (1984), McDonald and Butler (1987), McDonald and Richards (1987) or Venter (1984).

2.5. Heterogeneity

Heterogeneity is a problem which is often encountered in insurance data. Heterogeneity frequently results in distributions with thick tails. Hogg and Klugman (1983) indicate how mixture distributions provide an approach to modeling unobservable heterogeneity. The GB2 provides a mixture interpretation which allows, but does not require, heterogeneity. The GB2 can be shown to arise from a structural distribution which is $GG(x;\ a,\ \theta,\ p)$ where the scale parameter θ is distributed as a $GG(\theta;\ -a,\ b,\ q)$. The result may be of particular interest in the case of $a < 0$ for severity applications. The limiting case of large values of q corresponds to homogeneity.

Each special case of the GB2 can be interpreted as a mixture. Some important cases are summarized in Table 2. Heterogeneity can be tested by estimating the models in the first two columns of Table 2 using maximum likelihood estimation and testing for significant differences using a likelihood ratio test. See McDonald and Butler (1987)

Table 2

Some mixture distributions.

Observed distribution	Structural distribution	Parameter mixing distribution
$GB2(x;\ a,\ b,\ p,\ q)$	$GG(x;\ a,\ \theta,\ p)$	$GG(\theta;\ -a,\ b,\ q)$
$LT(x;\ \mu,\ \sigma^2,\ q)$	$LN(x;\ \mu,\ \theta)$	$GG(\theta;\ a=-1,\ \sigma^2 q,\ q)$
$BR3(x;\ a,\ b,\ p)$	$GG(x;\ a,\ \theta,\ p)$	$W(\theta;\ -a,\ b)$
$BR12(x;\ a,\ b,\ q)$	$W(x;\ a,\ \theta)$	$GG(\theta;\ -a,\ b,\ q)$
$B2(x;\ b,\ p,\ q)$	$GA(x;\ \theta,\ p)$	$GG(\theta;\ a=-1,\ b,\ q)$
$L(x;\ b,\ q)$	$EXP(x;\ \theta)$	$GG(\theta;\ a=-1,\ b,\ q)$

and Venter (1984) for more details. Thus, the GB2 has a theoretical justification as a representation of claims arising from a heterogeneous population of exposures.

The importance of the GB2 distribution for risk theory is that it has great flexibility due to the availability of four parameters. In addition, it encompasses many of the traditional distributions as special cases. The log-t is a limiting case of the GB2 as $a \to 0$; whereas the other models in Table 2 are special cases of the GB2.

3. Aggregate losses: The GB2 and MPY

3.1. Methodology

Several approaches to modeling the distribution of X have been considered. [see Cummins and Freifelder (1978)]. As mentioned above, one method of obtaining the aggregate loss distribution is by estimating and then compounding frequency and severity distributions. One reason for decomposing the problem in this way has to do with sample size. One year's observations on a relatively large pool may prove sufficient to estimate frequency and severity distributions with a high degree of confidence, providing that data have been maintained in the appropriate degree of detail. [2] By contrast, the aggregate losses of the pool for that year constitute only one observation from $F(x)$.

If the pool is relatively small and/or if frequency and severity data have not been maintained in a sufficient degree of detail, it is sometimes possible to estimate $F(x)$ from a time series of observations on total claims, X. [3] If a sufficient number of observations on X is available, it may be advisable to fit the GB2 distribution to this sample even if frequency and severity data have also been maintained. In this instance, the GB2 distribution fitted to the aggregate data provides an alternative to the distribution obtained by compounding frequency and severity. Thus, fitting the GB2 to the aggregate data will be useful, at a minimum, to check the reasonableness and stability of the estimates obtained from alternative methods.

Maximum likelihood estimates can be obtained by maximizing the usual loglikelihood function for a sample consisting of individual observations on X or, for grouped data, the loglikelihood function of the corresponding multinomial distribution.

The method of moments offers the possibility of using the GB2 to model total claims when separate frequency and severity distributions are available and when an adequate number of observations on X is not available. In this regard, the method of moments GB2 would be an alternative to an approximation method such as the normal power or to a compounding approach. Method of moments estimators may be obtained by solving the system of four non-linear equations obtained by setting the first four theoretical moments of the GB2 equal to the corresponding empirical moments. [4] The moments of X can be computed directly from the sample or inferred

[2] To estimate the frequency distribution, the number of claims experienced by each exposure unit must be maintained. This permits the tabulation and fitting of a frequency distribution. If only the total number of claims is available, one year's data constitutes only one observation from the frequency distribution.

[3] A cross section of risk pools might also be used, but this situation is less likely to arise in practice.

[4] The equations that must be solved to yield method of moments estimators for the GB2 are the following [moment formulas for other members of the family are given in McDonald (1984) and McDonald and Richards (1987)]:

$$E_{GB2}(X^h) = b^h \frac{B(p+h/a,\ q-h/a)}{B(p,q)},$$
$$= \hat{\mu}'_h(X),$$
$$= \sum_{t=1}^{n} X_t^h/n, \quad h=1,2,3,4,$$

for a, b, p and q, where $\hat{\mu}'(X)$ denotes the estimated hth-order moment of X about the origin. Cummins and Wiltbank (1983), Kottas and Lau (1979) and Lau (1984) report some formulas which relate the first four moments of aggregate loss to the first four moments of severity and frequency.

from severity and frequency data. [5] It would also be possible to fit a GG using three sample moments and use the mixing property summarized in Table 2 to infer a GB2 which accounts for the uncertain impact of inflation on nominal values. The statistical efficiency of method of moments estimation is questionable if the data are from a population without the requisite number of theoretical moments [Ord (1972)]. [6]

3.2. An example: Fire losses

In this section we analyze the data on aggregate fire loss at a major university reported in Cummins and Freifelder (1978). These data are reproduced in Table 3. The reported sample moments indicate that the distribution of these data are quite skewed and have thick tails. The GB2, LT, GG, B2, BR3, BR12, LN, W, GA, Lomax and exponential are fit to the data in Table 3 using maximum likelihood estimation. Inverse generalized gamma, Weibull, gamma, Lomax and exponential distributions are also fit and are denoted I-GG, I-W, I-GA, I-LOMAX, and I-EXP, respectively. The maximum likelihood estimates and corresponding loglikelihood values are given in Table 4. [7]

We note from Table 4 that none of the special cases of the GB2 has a loglikelihood value greater than that for the unconstrained GB2, as would be expected from their 'nested' relationships. The likelihood ratio statistic can be used to test for statistically significant differences between 'nested' models. For such cases, twice the difference between the loglikelihood values is asymptotically distributed as a Chi-square with degrees of freedom equal to the difference between the number of 'free' parameters in the two models being compared. Some caveats associated with the distribution of the likelihood ratio test statistic are discussed in the appendix. Several of the fitted distributions appear to fit the data equally well based upon reported values of the loglikelihood function. The best fitting distributions are the GB2, I-GG, B2, BR3, I-W, I-GA, I-LOMAX and I-EXP distributions. The large estimate for p, relative to q for the GB2, suggests the consideration of inverse distributions. The parameter 'a' in the GB2 also permits considerable flexibility in the distribution. One can think of the distribution of the reciprocal of the variable arising for $a < 0$ and the logarithm of the variable corresponding to the limiting case of $a = 0$. It is somewhat remarkable that the loglikelihood value for the one parameter inverse exponential agrees so well with the four-parameter GB2 in this particular application. Other distributions could not be rejected as popu-

Table 3

Fire loss experience of a major university.

Year	Total losses
1950	71280
1951	3671
1952	18664
1953	8784
1954	3966
1955	30892
1956	631626
1957	11464
1958	127194
1959	4950
1960	30452
1961	8028
1962	14790
1963	9480
1964	8676
1965	114198
1966	5150
1967	105864
1968	32814
1969	41340
1970	46284
1971	12230
1972	19418
Average	59183

Mean = 59183   Skewness = 3.94
Var = 1.61878×10^{11} Kurtosis = 17.84

Source: Cummins and Freifelder (1978).

[5] Empirical moments also have provided the basis for selecting members of the Pearson family. Applications of the Pearson family in modeling aggregate losses are discussed by Lau (1984) and Aiuppa (1988).

[6] If the aggregate approach is adopted, great care must be taken to allow for trends, cycles, and other sources of non-stationarity that may be present in the data. Of course, the non-stationarity problems will be reduced if data are available on a quarterly or monthly basis, provided the pool size is sufficient. Methods to adjusting for trends, inflation, and other data problems are discussed in Cummins and Freifelder (1978).

[7] The table corresponds to the unadjusted data [unadjusted for the varying number of exposure units or other factors discussed by Cummins and Freifelder (1978) and reported in Table 3]. Thus, the results in this section abstract from these adjustment problems and are only intended to illustrate estimation of the MPY.

Table 4

Aggregate fire loss data MLE parameter estimates. [a]

Model	$a(\mu)$	$b(\sigma)$	p	q(d.f.)	LL	Rank (LL)
Four-parameter						
GB2	1.2688	4.3336	14078.1	0.68389	−266.5	1
Three-parameter						
LT	(9.9788)	(1.2492)		(60.166)	−268.0	10
GG	0.15271	0.00001614	25.310		−268.6	12
I-GG	−1.2680	8068.4	0.6844		−266.5	1
B2	1.0000	0.7986	14731.3	0.9742	−266.5	1
BR3	0.9900	4.1936	2671.26	1.0000	−266.5	1
BR12	3.1856	6161.52	1.00	0.2301	−267.1	9
Two-parameter						
LN	(9.9933)	(1.2704)			−268.0	10
W	0.70115	42927.7	1.0000		−271.9	14
I − W	−0.9898	12127.0	1.0000		−266.5	1
GA	1.0000	95736.5	0.6182		−273.6	15
I − GA	−1.0000	11763.0	0.9742		−266.5	1
LOMAX	1.0000	38471.5	1.0000	1.6088	−269.2	13
I − LOMAX	−1.0000	0.4526	1.0000	26679.0	−266.5	1
One-parameter						
EXP	1.0000	59183.3		1.0000	−275.7	16
I-EXP	−1.0000	12075.0	1.0000		−266.5	1

[a] LL refers to the log of the likelihood function. Parameters in parentheses correspond to estimated parameters associated with the lognormal and log-t distributions.

lation distributions at the five percent level of significance, e.g., the LT, GG, Burr 12, Lognormal, and Lomax distributions. However, 23 observations could hardly be expected to validate the use of asymptotic tests.

Table 5 reports the Maximum Probable Yearly Aggregate Loss (MPY) estimates (quantiles) based on the estimated models reported in Table 4. The estimates for the 0.01 level reported by Cummins and Freifelder using normal approximations, Chebyshev, and normal power methodologies are also included in Table 5. [8]

[8] The Chebyshev inequality yields an upper bound on MPY for distributions with finite first and second moments. The inequality does not provide this information if the moments do not exist. In applications, the Chebyshev inequality may not provide an upper bound even for distributions with finite first and second moments. If, for example, the Chebyshev formula were based on the sample variance whereas the estimated variances of the parametric distributions were based on parameters obtained by the method of maximum likelihood (MLE), the Chebyshev method would not necessarily give an upper bound unless the parametric variance estimate is the same as the sample variance.

The alternative models provide very different results for the MPY. Given the great diversity in the estimates of MPY, there appears to be an open question as to the appropriateness of some distributions. The results suggest that the 0.01 MPY could exceed 1,250,000 which is much larger than estimates obtained from the normal and normal power approximations. However, we should recall that these results are based on 23 observations which would give limited information about extreme tail behavior.

Of the eight distributions having very similar loglikelihood values, only two distributions (GB2 and I − GG) have a 0.01 MPY out of the range (1,200,000, 1,350,000). The GB2 and I − GG suggest an MPY of 1,800,000. It is interesting to note that some of the commonly used distributions such as the lognormal, Weibull or gamma yield 0.01 MPY's of only about 400,000. The case of the BR12 is an anomaly for the aggregate loss data. In the absence of any additional information, we would recommend using an MPY corresponding to the simplest model having an acceptable fit and theoretical properties.

Table 5

Aggregate fire loss data estimated quantiles.

Model	$\alpha = 0.5$ [a]	0.1	0.01
Four-parameter			
GB2	16850	126600	1823000
Three-parameter			
LT	21560	108830	426900
GG	22870	114300	1386100
I-GG	16850	126600	1823000
B2	17610	120400	1345000
BR3	17560	117790	1264000
BR12	15610	142540	3297000
Two-parameter			
LN	21880	111500	420300
W	25450	141040	379000
I-W	17560	117800	1265000
GA	31800	152890	350100
I-GA	17600	120370	1338000
LOMAX	20720	122490	635000
I-LOMAX	17420	114600	1202000
One-parameter			
EXP	41020	136270	272500
I-EXP	17420	114600	1201000
Sample [b]	18664	121996	> 631626
Normal			397200
Chebyshev			1373000
Normal Power			737500

[a] $\alpha = 1 - \Pr(X \leq x)$

[b] Sample quantiles are estimated by the method suggested in Hogg and Klugman (1984, p. 63). The 0.01 quantile could not be estimated by this method due to the sample size.

The estimated MPY's reported in Table 5 are, of course, subject to sampling variation. Conditional on correct model specification, this sampling variation will depend on the variance–covariance matrix of the parameter estimators. The complexity of this relationship increases with the number of parameters estimated. The inverse exponential can be used to illustrate the relationship between the MPY's and parameter estimates. The MLE of the single parameter in the inverse exponential is $b = 12075$. The 0.01 MPY's corresponding to $0.9b$, $0.99b$, b, $1.01b$ and $1.1b$ are 1081000, 1189000, 1201000, 1213000 and 1322000. Thus, we see that one and ten percent changes in b can be important.

4. The severity of loss distribution

The Cummins and Freifelder (1978) severity data consist of 80 fire claims. [9] These data are reported in grouped form in Table 6. The ungrouped data are presented in the appendix. Cummins and Freifelder found that the lognormal and gamma distributions did not have sufficiently heavy tails to describe the data. The log-t distribution was found to provide a better fit to the tail data.

Maximum likelihood estimation was used in the current paper to estimate the members of the GB2 family. Both grouped and ungrouped data were used. In general, individual claim data are preferable to grouped data because information is lost through grouping. However, especially in small samples, the use of grouped data provides a way to reduce the impact of potential outliers. Both grouped and ungrouped data are evaluated here to provide an indication of the differences in the estimates that are likely to result from grouping. In estimating the models based upon the grouped data, the last three groups in Table 6 were combined as in Cummins and Freifelder.

The estimation results are reported in Tables 7 and 8. Of the 'non-inverse' distributions, there appears to be very close agreement between the GB2, the two Burr distributions and the B2 for both the grouped and ungrouped data. The corresponding likelihood ratio tests are not statistically significant at the five percent level. Large estimated values of p relative to q in the GB2 suggest that 'inverse' forms should be seriously considered. In this example the estimated value of p is more than four times as large as the estimated value of q for individual data and more than 140 times as large for grouped data. Accordingly, the inverse distributions I-GG, I-W, I-GA, I-LOMAX and even the one parameter inverse exponential are seen to have loglikelihood values very close to that of the estimated GB2.

The Chi-square goodness of fit test does not provide the basis for rejecting either the LN or LT as being consistent with the data. However, the

[9] The data cover several years and have been adjusted to a common time point using a claims cost index maintained by the university from which the claims were obtained. The data are described in more detail in Cummins and Freifelder (1978).

Table 6

Cummins–Freifelder (1978) severity distribution (grouped data).

Grouped format:	# of losses
(0,800]	6
(800,1442]	15
(1442,2093]	10
(2093,2820]	7
(2820,3696]	9
(3696,4845]	6
(4845,6527]	7
(6527,9471]	3
(9471,17124]	6
(17124,31158]	6
(31158,49803]	0
(49803,∞)	5
	80

GB2, B2, and the two Burr distributions provide statistically significant improvements over either the LN or the LT. Ranking the distributions based on Chi-square values gives very similar rankings relative to those based on loglikelihood values for grouped data. The differences between the Chi-square and loglikelihood rankings appear larger for individual data than for grouped data, as might be expected. However, values of the Chi-square statistics only differ by 0.7 for the nine best fitting distributions obtained from individual data. The Chi-square test should be used with caution in this type of analysis, however, for two reasons: (1) It is not a very powerful test for goodness of fit of continuous distributions, and (2) it may fail to reject a distribution which provides an adequate fit in the body of the distribution but a very poor fit in the tail. Underestimating the tail can have serious consequences in an insurance context as is reported in sections three and five.

The Burr 12 results are very close to the GB2 results for severity data. Given the flexibility and relative simplicity of the Burr 12, it deserves to be given careful consideration in empirical work in this area. The inverse forms of the Weibull, gamma, Lomax and exponential also deserve serious consideration. Another interesting observation is that while the estimated parameter values for the GB2 in Tables 7 and 8 appear to be quite different, the estimated distributions are very simi-

Table 7

MLE parameter estimates severity data: Grouped data. [a]

Model	$a(\mu)$	$b(\sigma)$	p	q(d.f.)	LL	χ^2	Rank (LL)
Four-parameter							
GB2	1.5308	67.3722	71.1851	0.5039	−17.6	2.6	1
Three-parameter							
LT	(7.9885)	(9.750)		(2.6064)	−21.3	10.0	10
GG	0.06068	$3.309.10^{-34}$	176.66		−22.1	11.1	12
I-GG	−1.5087	1095.8	0.51210		−17.6	2.6	1
B2	1.0000	0.00359	520180.0	0.9307	−17.8	3.0	4
BR3	0.9687	0.0247	57748.0	1.0000	−17.9	3.1	6
BR12	2.9525	1095.4	1.0000	0.2540	−17.6	2.8	1
Two-parameter							
LN	(8.1378)	(1.2378)			−21.6	10.3	11
W	0.8275	5982.1	1.0000		−28.6	24.4	14
I-W	−0.96865	2035.8	1.0000		−17.9	3.1	6
GA	1.0000	7896.6	0.82068		−29.7	27.3	15
I-GA	−1.0000	1869.2	0.93065		−17.8	3.0	4
LOMAX	1.0000	5839.7	1.0000	1.5507	−24.4	15.4	13
I-LOMAX	−1.0000	0.07317	1.0000	27588.0	17.9	3.2	6
One-parameter							
EXP	1.0000	6310.3	1.0000		−30.5	29.9	16
I-EXP	−1.0000	2018.5	1.0000		17.9	3.2	6

[a] LL refers to the log of the likelihood function. Parameters in parentheses correspond to estimated parameters associated with the lognormal and log-t distributions.

Table 8

MLE parameter estimates severity data: Individual observations. [a]

Model	$a(\mu)$	$b(\sigma)$	p	q(d.f.)	LL	χ^2	Rank (LL)
Four-parameter							
GB2	3.9658	1097.4	0.8524	0.1866	−784.6	3.6	1
Three-parameter							
LT	(8.0159)	(1.0284)		(4.4832)	−791.6	10.6	10
GG	0.15378	0.00001072	21.0437		−798.4	15.1	13
I−GG	−1.0913	1656.1	0.84013		−785.5	2.9	3
B2	1.0000	20.437	96.444	0.9664	−785.6	3.1	4
BR3	0.9918	31.991	62.194	1.0000	−785.6	3.2	4
BR12	3.5284	1062.0	1.0000	0.2125	−784.6	3.4	1
Two-parameter							
LN	(8.2151)	(1.3490)			−794.7	11.4	10
W	0.5810	7757.7	1.0000		−815.1	40.1	14
I-W	−0.9811	2034.8	1.0000		−785.6	3.1	4
GA	1.0000	39513.0	0.4280		−830.9	68.0	15
I-GA	−1.0000	1935.4	0.95886		−785.6	3.0	4
LOMAX	1.0000	4640.1	1.0000	1.2941	−796.6	15.8	12
I-LOMAX	−1.0000	47.1533	1.0000	43.8093	−785.6	3.2	4
One-parameter							
EXP	1.0000	16950.0	1.0000		−859.0	110.0	16
I-EXP	−1.0000	2018.4	1.0000		−785.6	3.2	4

[a] LL refers to the log of the likelihood function. Parameters in parentheses correspond to estimated lognormal and log−t parameters. The Chi-square tests are based on the groupings in Table 6, with the last three categories combined.

lar. It is not uncommon to find examples, such as this one, in which the likelihood surface is quite flat. [10]

5. Applications of estimated distributions

The estimated frequency, severity, and total claims distributions are useful in numerous practical applications. The selection of distributional forms can have a significant impact on estimates of reinsurance premiums, tail quantiles (e.g., for MPY or ruin calculations), and other important statistics. In this section, we investigate the effects of model selection on the tails of the severity distribution, excess of loss reinsurance premiums, and simulated total claims distributions. We compare the GB2, Burr 12, and inverse generalized gamma with the lognormal distribution. The first three distributions were selected because they fit the data best in terms of likelihood function values, while the lognormal was chosen because it has been used frequently in the prior literature. Recall that the lognormal could not be rejected using the Chi-square goodness of fit test.

5.1. *Tails of severity distributions*

The tails (last 15 observations) of the estimated GB2, Burr 12, inverse generalized gamma, and lognormal severity distributions based on grouped and ungrouped data are shown in Table 9 along with the empirical distribution function (EDF). The tails of the grouped and ungrouped distributions differ noticeably, and the differences are large enough to have a significant impact on reinsurance premiums and other quantities. Visual inspection reveals that the grouped and un-

[10] Depending upon the starting values chosen in the maximum likelihood estimation, it is possible to 'converge' to different parameter estimates for the GB2 from the same data set. For example, the authors obtained the following alternative set of GB2 parameters from the ungrouped data on one estimation:

$a = 1.1132, \quad b = 20.9366, \quad p = 126.9381, \quad q = 0.8202.$

This parameter set yielded a loglikelihood function value of −785.5, which is not as good as the results reported in Table 8, but the estimated distribution functions are very close.

Table 9

Tails of selected severity distributions (last 15 observations). [a]

ln(loss)	EDF	Ungrouped data (Table 8)				Grouped data (Table 7)			
		GB2	Burr12	Inverse gen. gamma	Log-normal	GB2	Burr12	Inverse gen. gamma	Log-normal
9.3218	0.8148	0.8283	0.8288	0.8258	0.7940	0.8155	0.8137	0.8144	0.8306
9.3467	0.8272	0.8315	0.8320	0.8295	0.7992	0.8190	0.8171	0.8179	0.8356
9.5845	0.8395	0.8587	0.8594	0.8612	0.8450	0.8490	0.8459	0.8480	0.8788
9.6017	0.8519	0.8604	0.8612	0.8633	0.8480	0.8509	0.8478	0.8500	0.8815
9.7519	0.8642	0.8751	0.8760	0.8801	0.8727	0.8671	0.8635	0.8662	0.9039
9.8300	0.8765	0.8821	0.8830	0.8881	0.8844	0.8748	0.8710	0.8740	0.9142
10.0370	0.8889	0.8989	0.8999	0.9068	0.9116	0.8932	0.8889	0.8925	0.9375
10.0510	0.9012	0.8999	0.9009	0.9080	0.9132	0.8944	0.8901	0.8936	0.9389
10.0625	0.9136	0.9008	0.9018	0.9089	0.9146	0.8953	0.8910	0.8946	0.9400
10.2545	0.9259	0.9139	0.9149	0.9232	0.9347	0.9097	0.9051	0.9090	0.9564
10.9615	0.9383	0.9490	0.9499	0.9594	0.9791	0.9476	0.9431	0.9472	0.9887
10.9951	0.9506	0.9502	0.9512	0.9606	0.9803	0.9490	0.9445	0.9486	0.9895
11.6366	0.9630	0.9690	0.9698	0.9780	0.9944	0.9689	0.9651	0.9687	0.9976
11.7218	0.9753	0.9709	0.9717	0.9797	0.9953	0.9709	0.9672	0.9707	0.9981
13.3477	0.9877	0.9913	0.9916	0.9954	0.9999	0.9917	0.9899	0.9916	1.0000

[a] EDF = empirical distribution function = $i/(N+1)$, where i = observation number and N = sample size.

grouped GB2 tails are more comparable than those of the other distributions. [11] The potential generality of this result would be an interesting topic for future research.

The distribution function values, including the tails, of the GB2 and Burr 12 based on ungrouped data are virtually identical. For grouped data, the Burr 12 has a heavier tail than the GB2. In the grouped case, the GB2 appears to fit the data slightly better than the Burr 12. It is apparent that the tail of the lognormal is much too light to represent the data. Visual inspection also suggests that the Burr 12 fits the tail better than the inverse generalized gamma.

5.2. Reinsurance premiums

Reinsurance premium calculations were based on excess of loss reinsurance with both upper and lower limits. The calculations estimate the reinsurance company's expected severity. The expected severities would be multiplied by expected (total) frequency to give the reinsurance pure premium. The reinsurer's obligation for any given claim is the following:

$$X_{\mathrm{RE}} = \begin{cases} 0 & \text{for } X \le M \\ X - M & \text{for } M < X < U, \\ U - M & \text{for } X \ge U \end{cases}$$

where

X_{RE} = the reinsurer's obligation,
X = loss amount, and
M, U = lower and upper boundaries of the reinsurance layer.

For illustrative purposes, the upper boundary was set at \$1,000,000 and the lower boundary allowed to vary between ln $M = 7$ (approximately \$1,000) and \$1,000,000. The upper boundary point was chosen to be slightly higher than the largest observed claim (X_{MAX} = \$626,000).

The differences between expected reinsurance severity based on grouped and ungrouped models tend to be quite noticeable. Consider, for example, the expected severities for the grouped and ungrouped Burr 12 distributions, shown in Fig. 2. The differences in the premiums range from about 15 percent for lower values of M to about 22 percent for higher values of M. Since the expected severities translate directly into pure premiums, the pure premium differences would be of the same magnitude. This is especially significant be-

[11] For example, the ratio of the ungrouped tail probabilities to the grouped tail probabilities is generally closer to 1 for the GB2 than for the other distributions.

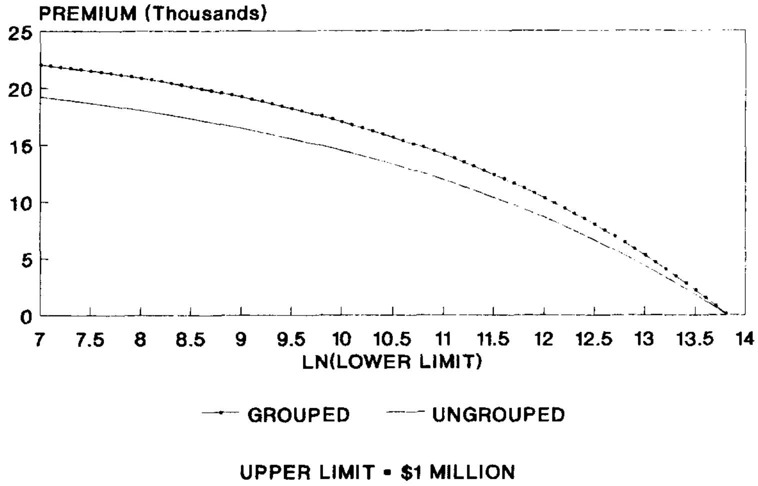


Fig. 2. Burr 12: Expected reinsurance severity.

cause the differences in the estimates could easily exceed the expected profit loading. Thus, the choice of the 'wrong' distribution could result in inadvertently writing a policy at an expected loss.

The importance of the choice of a severity model is reinforced by Fig. 3, which presents the expected reinsurance severities based on ungrouped data for the EDF, GB2, inverse generalized gamma, and lognormal distributions. [12] The expected severities for the GB2 and Burr 12 are at least twice as large as those for the lognormal, over all values of M. The differences between the inverse generalized gamma and the GB2 and Burr 12 are approximately 25 percent. This is especially significant in view of the fact that the inverse generalized gamma distribution is not significantly different from the GB2 at the 10 percent level of significance (likelihood test). The difference is statistically significant at about the 17 percent level. Thus, in selecting a severity distribution model, it may be necessary to consider weaker significance criteria than the usual 5 or 10 percent levels. If the tail is of primary importance, visual inspection of a graph such as fig. 2 may be more effective than formal significance tests. [13]

5.3. Simulated total claims distributions

The final applications involved simulating the total claims distribution $F(x)$ for various severity distributions. Two negative binomial frequency distributions were used:

low frequency: Failure parameter = 0.03714,

Shape parameter = 95.1968,

high frequency: Failure parameter = 0.03714,

Shape parameter = 3240.6435.

[12] The Burr 12 was omitted because it plots very close (slightly below) to the GB2. The severities for the EDF are sensitive to the choice of $F(1{,}000{,}000)$. This probability was varied between 0.994 and 0.999. The EDF severity graph remained between the GB2 and inverse generalized gamma graphs for all probabilities in this range. The EDF curve in Fig. 3 corresponds to $F(1{,}000{,}000) = 0.999$.

[13] The other moments of reinsurance severity such as the variance also are affected by the choice of model. Differences in higher moments (e.g., the variance) reflect differences in the distributional models and would exist entirely apart from any sampling error present in the estimated distributions. Because reinsurance risk loadings often are based on the variance or standard deviation, the choice of the wrong model may lead to errors in both the pure premium and the risk loading. Sampling error is another potential problem that may affect the choice of model and the accuracy of estimated moments. It would be possible to estimate confidence bounds for the estimated quantiles using the method discussed in Hogg and Klugman (1984, pp. 93–94). This is beyond the scope of the present study.

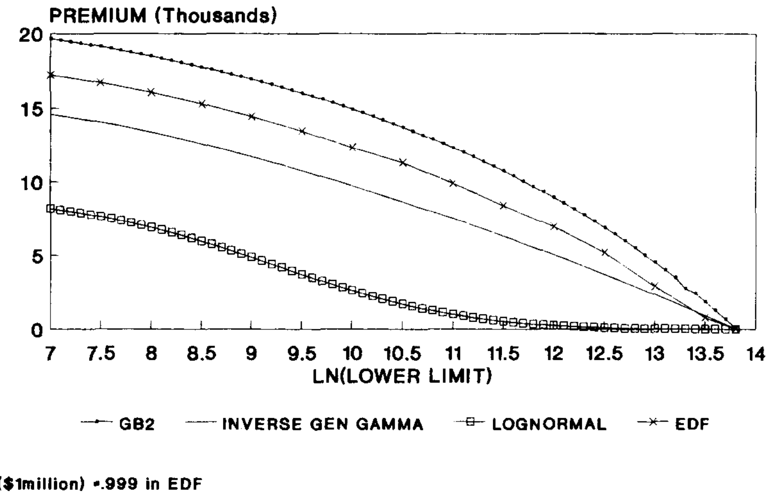


Fig. 3. Reinsurance premiums (severity): Ungrouped data, upper limit = $1 million.

The low frequency distribution is similar to the one used by Cummins and Freifelder (1978).[14] The expected number of claims per year for this distribution is 3.67, corresponding to the experience of the university supplying the fire insurance data. Since most insurance pools would have higher frequency, the high frequency distribution is also simulated. The expected number of claims in this case is 125 per year. This might correspond to a corporate self-insurance program. The number of frequency draws used in all of the simulations is 50,000.

The simulated total claims distributions are presented in Tables 10 through 12. Tables 10 and 11 give low frequency results for grouped and ungrouped data, while Table 12 gives the results for the high frequency distribution. The empirical moments of the estimated distributions also are given in the tables. These statistics should be interpreted carefully, since virtually none of the theoretical moments exist for the estimated in-

[14]The parameterization of the negative binomial is $p(k) = p^r q^k \Gamma(r+k)/[\Gamma(r)\Gamma(k+1)]$, where q is the failure parameter and r is the shape parameter. The parameter values used in the simulations are comparable to those in the Cummins–Freifelder article (1978, Table 2). The shape parameter for one exposure unit is identical to the one used by Cummins and Freifelder (0.2978). The failure parameter used here (0.03714) is slightly larger than the one used by Cummins and Freifelder (0.0371). The latter parameter does not appear in Table 2 of Cummins–Freifelder because they used a different parameterization of the negative binomial. For purposes of comparison, the failure parameter (0.0371) has been computed from the second parameter given in Table 2 of Cummins–Freifelder. The shape parameter used in the low frequency simulations in the present article represents approximately 320 exposure units while the high frequency parameter implies about 10,882 exposure units.

Table 10
Quantiles of simulated total claim distributions: Low frequency, 50,000 observations, ungrouped severity. [a,b]

Quantile	GB2	Burr12	Generalized Gamma	Log-normal
0.50	17.68	17.71	18.50	21.57
0.60	25.24	25.15	25.31	28.21
0.70	37.66	37.34	35.92	37.12
0.80	63.67	62.69	55.36	50.16
0.90	155.75	150.61	109.36	75.35
0.91	175.79	170.49	121.10	79.35
0.92	203.25	196.08	137.39	84.20
0.93	241.46	232.45	156.11	89.93
0.94	294.48	283.01	181.39	96.56
0.95	375.00	358.52	221.18	104.32
0.96	497.30	473.29	279.12	115.52
0.97	729.09	689.19	370.84	129.39
0.98	1275.38	1196.90	560.30	152.38
0.99	3050.23	2826.37	1166.10	193.97
0.999	72841.54	64728.64	13458.12	435.48
Mean	2535.61	2119.49	152.54	33.84
Std dev	261895.11	214385.12	5070.00	44.78
Skewness	166.68	165.65	118.76	7.83

[a] Frequency is negative binomial with failure parameter 0.03714 and shape parameter 95.1968. Expected number of claims per year is 3.67. Table entries are in 000's except for quantiles and skewness.
[b] Severity distribution parameters estimated from ungrouped data.

Table 11

Quantiles of simulated total claim distributions: Low frequency, 50,000 observations, grouped severity. [a,b]

Quantile	GB2	Burr12	Generalized gamma	Log-normal
0.50	19.21	19.35	19.38	18.70
0.60	27.31	27.92	27.42	24.01
0.70	40.47	42.26	40.94	30.98
0.80	67.17	72.67	68.12	40.82
0.90	156.80	182.93	154.39	59.22
0.91	177.28	206.65	175.19	62.24
0.92	202.58	240.65	205.62	65.55
0.93	238.24	287.44	239.91	69.67
0.94	288.39	352.40	289.90	74.12
0.95	362.86	452.17	367.92	79.69
0.96	474.81	602.98	487.81	87.07
0.97	681.36	891.45	688.08	96.69
0.98	1168.71	1588.54	1142.17	111.02
0.99	2662.35	3882.27	2782.93	138.41
0.999	55309.24	99831.36	49193.65	284.25
Mean	1578.79	3883.75	650.78	27.11
Std dev	151354.39	415080.41	38923.73	30.91
Skewness	163.15	168.45	136.30	5.38

[a] Frequency is negative binomial with failure parameter 0.03714 and shape parameter 95.1968. Expected number of claims per year is 3.67. Table entries are in 000's except for quantiles and skewness.

[b] Severity distribution parameters estimated from grouped data.

verse generalized gamma, GB2, and Burr 12 distributions.

The results presented in Tables 10 through 12 reveal significant differences in the estimated quantiles depending upon the data used to estimate the parameters and the choice of model. For example, using ungrouped severity estimates and low frequency (Table 10), the 99th percentiles of the total claims distributions range from \$193,970 for the lognormal to \$3,050,230 for the GB2. The GB2 and Burr 12 results are similar in Table 10 but differ substantially from the inverse generalized gamma. Based on the grouped data (Table 11), the GB2 and inverse generalized gamma results are comparable at the 99th percentile, but the Burr 12 yields a significantly larger value for X. The differences between the grouped and ungrouped low frequency estimates also are substantial. The GB2 performs best in terms of giving similar results using grouped and ungrouped data. It is difficult to say whether this result has general implications. However, the additional parameter of the GB2 may lead to enhanced reliability.

Due to the high costs of simulating the high frequency case, a full set of results was obtained only for the distributions based on ungrouped data. These results appear in Table 12. The overall conclusions are similar to those based on the low frequency case, i.e., there are substantial differences in the MPY values produced by the alternative models of $F(x)$. For example, the 99th percentile of $F(x)$ is \$1,912,000 for the lognormal and \$373,981,000 for the GB2. The inverse generalized gamma also produces a 99 percent MPY much lower than that of the GB2. The Burr 12 results are comparable to those of the GB2. High-frequency simulation results with the Burr 12, inverse generalized gamma, and lognormal reveal significant differences in MPYs based on the grouped and ungrouped data similar to those observed in the low frequency case.

These results reinforce the conclusion of Cummins and Freifelder (1978) that the choice of a severity distribution can have a substantial impact on the estimation of MPY and ruin probabilities. The potential error in using the lognormal distri-

Table 12

Quantiles of simulated total claim distributions: High frequency, 50,000 observations, ungrouped severity. [a,b]

Quantile	GB2	Burr12	Generalized gamma	Log-normal
0.50	3023.76	2893.43	1857.77	1116.58
0.60	3903.74	3711.30	2171.01	1175.80
0.70	5409.47	5099.43	2662.45	1244.25
0.80	8469.46	7893.35	3574.66	1332.23
0.90	18614.66	17042.23	6129.56	1467.12
0.91	21114.32	19268.32	6718.28	1486.49
0.92	24782.75	22555.95	7408.53	1509.26
0.93	29468.11	26717.69	8232.88	1534.46
0.94	35786.90	32330.38	9336.65	1563.81
0.95	44956.99	40467.63	10875.72	1598.11
0.96	58537.12	52494.64	13301.50	1642.23
0.97	86803.76	77472.09	17613.00	1694.32
0.98	147382.83	130296.87	26626.59	1775.58
0.99	373980.84	325678.38	54672.50	1911.62
0.999	7836815.00	6552029.50	628293.38	2488.61
Mean	67399.30	180539.81	7878.91	1148.67
Std dev	2238155.45	21393692.13	238635.71	256.04
Skewness	70.67	200.71	133.11	1.46

[a] Frequency is negative binomial with failure parameter 0.03714 and shape parameter 3240.6435. Expected number of claims per year is 125. Tables entries are in 000's except for quantiles and skewness.

[b] Severity distribution parameters estimated from ungrouped data.

bution is especially serious. Thus, the range of severity distributions considered in risk management applications should be significantly greater than in the past. The GB2 family of distributions, together with their inverse distributions, should be sufficiently flexible for most insurance applications. The use of grouped rather than individual claim data also can lead to substantially different results. This reinforces the importance of collecting individual claim data, even when grouped estimates will be obtained as an alternative.

6. Summary

We have investigated several methods of estimating probability distributions and their inverses using the GB2 family of distributions. The GB2 has been shown to encompass many useful distributions and to provide a systematic way of considering them. The four parameters of this distribution provide sufficient flexibility to model most insurance loss and severity distributions. The GB2 provides at least as good a fit as any distribution 'nested' within it, though it may be equaled when a special or limiting case adequately describes the process being modeled. In many cases, a three-parameter member of the family may be adequate. The Burr 12 is especially attractive in this regard because it has a closed form distribution function.

In modeling insurance loss distributions, we have shown that the choice of a distributional model for severity can have substantial effects on important statistics such as expected severities under reinsurance contracts and MPY estimates. The use of grouped individual claims data in estimating the severity distribution also can produce noticeably different results.

It is advisable to compute the relevant statistics using several different distributions in order to obtain an indication of the range of results that can be obtained using different models. When testing among alternative distributions, the Chi-square test should be deemphasized and likelihood ratio tests should be used, with lower tolerances (e.g., 20 percent) than the usual 1, 5, or 10 percent because different tail behavior does not necessarily imply significant differences based upon comparisons involving the entire distribution.

Continuing improvements in computational hardware and software have made maximum likelihood techniques much more feasible than in the past. The gains from adopting these techniques and considering a broader range of probability distributions are potentially quite significant. The argument in favor of adopting these more general methodologies is especially strong when one considers the potential for error and the incomplete theoretical basis inherent in some techniques that have been used in the past to achieve computational simplicity.

The validity of the Chi-square distribution as the asymptotic distribution of the likelihood ratio statistic is conditional on a number of regularity conditions, including the parameters being in the interior of the parameter space. This condition is violated in a number of hypotheses considered in this paper. Recent research by one of the authors suggests that the likelihood ratio test may be too conservative in these cases and is less likely to reject the hypothesis than is indicated by the critical level of the Chi-square.

Appendix

Table A.1

Cummins and Freifelder (1978) Severity data: Individual observations.

290.40	1248.49	2202.96	3941.30	10560.10
537.19	1268.24	2222.80	4017.01	11179.54
756.80	1284.56	2255.72	4100.00	11461.39
769.19	1363.85	2274.61	4166.98	14538.13
787.69	1436.20	2328.64	4355.02	14789.81
796.18	1445.96	2384.37	5117.93	17186.09
933.62	1469.48	2847.83	5335.96	18582.57
967.97	1507.47	2947.04	5453.02	22857.33
1010.56	1662.36	2948.35	5568.96	23177.85
1017.40	1674.58	3036.51	5761.83	23446.13
1033.49	1690.91	3287.68	6161.81	28409.82
1034.33	1739.96	3331.62	6348.69	57612.82
1056.93	1776.56	3416.67	6859.37	59582.78
1124.09	1932.09	3604.66	7972.20	113164.70
1165.73	1975.89	3671.16	8028.32	123228.90
1217.64	2099.79	3739.30	10047.22	626402.80

References

Abramowitz, Milton and Irene A. Stegun (1965). *Handbook of Mathematical Functions*. Dover Publications, New York.

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

Beard, R.E., T. Pentikainen and E. Pesonen (1965). *Risk Theory*. Methuen, London.

Benckert, L.G. (1962). The log-normal model for the distribution of one claim. *ASTIN Bulletin* 2, Part I, 9–23.

Buhlmann, Hans (1970). *Mathematical Methods in Risk Theory*. Springer-Verlag, New York.

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

Cummins, J.D. and Laurel J. Wiltbank (1983). Estimating the total claims distribution using multivariate frequency and severity distributions. *Journal of Risk and Insurance* 50 377–403.

Elderton, W.P. and N.L. Johnson (1969). *Systems of Frequency Curves*. Cambridge University Press, London (1.3, 1.7, 1.11).

Feller, William (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd ed. Wiley, New York.

Ferreira, Joseph, Jr. (1970). *Quantitative Models for Automobile Accidents in Insurance*. U.S. Department of Transportation Automobile Insurance and Compensation Study. U.S. Government Printing Office, Washington, D.C.

Heckman, P.E. and G.G. Meyers (1983). The calculation of aggregate loss distributions from claim severity and claim count distributions. *Proceedings of the Casualty Actuarial Society*, Vol. 70.

Hewitt, C.C. (1970). Credibility for severity. *Proceedings of the Casualty Actuarial Society*, Vol. 57.

Hogg, R.V. and S.A. Klugman (1983). On the estimation of long-tailed distributions with actuarial data. *Journal of Econometrics* 23, 91–102.

Hogg, R.V. and S.A. Klugman (1984). *Loss distributions*. Wiley, New York.

Johnson, Norman L. and Samuel Kotz (1970). *Continuous Univariate Distributions*, Vol. 1. Wiley, New York.

Kottas, J.F. and H.S. Lau (1979). A realistic approach for modeling stochastic lead time distributions. *AIIE Transactions* 11, no. 1, 54–60.

Lau, Hon-Shiang (1984). An effective approach for estimating the aggregate loss of an insurance portfolio. *The Journal of Risk and Insurance* 50, 21–30.

McDonald, James B. (1984). Some generalized functions for the size distribution of income. *Econometrica* 52, 647–663.

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

McDonald, James B. and Dale O. Richards (1987). Model selection: Some generalized distributions, *Communications in Statistics* 16, 1049–1074.

Mandelbrot, B. (1964). Random walks, fire damage and other Paretian risk phenomena. *Operations Research* 12.

Ord, J.K. (1972). *Families of frequency distributions*. Hafner Publishing Company, New York.

Panjer, H.H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* 12, 22–30.

Panjer, H.H. and G.E. Willmot (1987). Difference equation approaches in evaluation of compound distributions. *Insurance: Mathematics and Economics* 6, 43–56.

Paulson, A.S. (1984). Estimating the total claims distribution in property–liability insurance. Working paper. S.S. Huebner Foundation, University of Pennsylvania, Philadelphia, PA.

Paulson, A.S. and R. Dixit (1989). Some general approaches to computing total loss distributions and the probability of ruin. In: J.D. Cummins and R.A. Derrig, eds., *Financial Models of Insurance Solvency*. Kluwer Academic Publishers, Norwell, MA.

Paulson, A.S. and N.J. Faris (1985). A practical approach to measuring the distribution of total annual claims. In: J.D. Cummins, ed., *Strategic Planning and Modelling in Property–Liability Insurance*. Kluwer Academic Publishers, Norwell, MA.

Roy, Yves and J. David Cummins (1985). A stochastic simulation model for reinsurance decision making by ceding companies. In: J.D. Cummins, ed., *Strategic Planning and Modelling in Property–Liability Insurance*. Kluwer Academic Publishers, Norwell, MA.

Seal, Hilary L. (1969). *Stochastic Theory of a Risk Business*. John Wiley and Sons, New York.

Shpilberg, D.C. (1977). The probability distribution of fire loss amount. *Journal of Risk and Insurance* 44, 103–115.

Venter, Gary C. (1984). Transformed beta and gamma functions and aggregate losses. Reprinted from: *Proceedings of the Casualty Actuarial Society*, Vol. 71. Recording and Statistical Corporation, Boston, MA.