

Mixtures of tails in clustered automobile collision claims

Guyonne R.J. Kalb^a, Paul Kofman^a, Ton C.F. Vorst^{b,*}

^a *Department of Econometrics, Monash University, Clayton, Vic.3168, Australia*

^b *Department of Finance and Erasmus Center for Financial Research, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

Received February 1995; revised September 1995

Abstract

Knowledge of the tail shape of claim distributions provides important actuarial information. This paper discusses how two techniques commonly used in assessing the most appropriate underlying distribution can be usefully combined. The maximum likelihood approach is theoretically appealing since it is preferable to many other estimators in the sense of best asymptotic normality. Likelihood based tests are, however, not always capable to discriminate among non-nested classes of distributions. Extremal value theory offers an attractive tool to overcome this problem. It shows that a much larger set of distributions is nested in their tails by the so-called tail parameter.

This paper shows that both estimation strategies can be usefully combined when the data generating process is characterized by strong clustering in time and size. We find that the extreme value theory is a useful starting point in detecting the appropriate distribution class. Once that has been achieved, the likelihood-based EM-algorithm is proposed to capture the clustering phenomena. Clustering is particularly pervasive in actuarial data. An empirical application to a four-year data set of Dutch automobile collision claims is therefore used to illustrate the approach.

Keywords: Extreme value theory; EM-algorithm; Mixtures of distributions

1. Introduction

A statistical problem of main interest to insurance companies is to estimate the probability of claims exceeding some extreme level. Not only is this a prerequisite for calculating the 'cost' of insurance, it is of subsidiary use as well. If policy holders opt for premium deductions (e.g., by incurring non-coverage of small losses), the left-hand side tail of the claims probability distribution reveals the necessary information. Reinsurance contracting based on large losses, on the other hand, requires detailed knowledge on the

behaviour of the right-hand side tail of the underlying claims distribution. Theoretically, two techniques that will be discussed shortly have been proposed to provide answers to these questions. Since most observations fail within the smaller claims area, the lower tail is empirically well defined. That is not the case for the upper tail which will be our focal point. Scarce empirical evidence indicates that these upper tails contain far more probability density than can be explained by (log)normal distributions. A thorough review of potential actuarial distributions that are characterized by fat tails, is given in Cummins et al. (1990) who adopt a maximum likelihood approach, and Beirlant et al. (1994) who focus on extreme value approaches. The

* Corresponding author.

flexibility and asymptotic efficiency of the maximum likelihood approach is considered to be an advantage. Unfortunately, some of the proposed distributions are difficult to compare by maximum likelihood since they are not nested within the same distribution class. MLE estimates will, by implication, be dependent on the maintained hypothesis. Likelihood ratio testing will then be ruled out. Similarly, the Cox test will not be applicable (White, 1982) since it is very likely that the second moment does not exist. A Chi-squared goodness of fit test is not very sufficient either, since it is dependent on the procedure of breaking up the empirical distribution function into arbitrary intervals, see Koedijk et al. (1990).

This paper proposes a unifying approach to assess the shape of actuarial distributions based on a two-step procedure that avoids these problems. The first step consists of an extreme value classification of appropriate distribution classes based on the information provided by empirically observed tails. The second step refines these findings by an iterative maximum likelihood procedure known as the EM-algorithm. Both complementary methods will shortly be discussed in this section and theoretically introduced in Sections 2 and 3.

Recently, numerous articles (see Galambos et al., 1994) employed the Hill tail estimator as a nesting tool that simultaneously conveys evidence on the tail properties. The extreme value theory underlying this estimator proves to be particularly appropriate for specification of the empirical tails of very long time series (e.g., high frequency financial data). Based on a finite large number of empirical observations, the estimator allows a wider set of distributions to be discriminated amongst. In particular, Student- t distributions and GARCH-processes, well known in the financial literature, are included in the discriminatory set. The tail parameter also offers estimation of 'residual life' or, otherwise called exceedance probabilities of observations surpassing the highest empirically observed level. This concept is of great actuarial value, particularly in setting reinsurance premiums, see Benktander and Segerdahl (1960), which requires both the expected number of claims that will exceed some given (high) retention level x , and the average excess claim over retention. This problem requires the application of extreme value theory. Having an estimator for the shape of the tail, then allows probability statements

necessary to calculate the appropriate reinsurance premium. For applications of Hill's tail estimator in an actuarial context we refer to Beirlant and Teugels (1992) and Beirlant et al. (1994).

However, the allegation that distributions are best characterized by their tails deserves some closer scrutiny. Both approaches are dependent on a number of maintained hypotheses. Some of these can easily be relaxed in general, see also Cummins and Freifelder (1978). Actuarial applications, however, are characterized by two clustering features that have so far largely been neglected in the literature¹. First, actuarial claims tend to cluster in time. Damage due to bad weather or month-of-the-year seasonality effects tends to disturb i.i.d. requirements (for both maximum likelihood and tail parameter) of the observations. Or, inflation may lead to a slowly trending claim size. In either case we need to adopt a declustering scheme such that the resulting empirical peaks (tail observations) are driven by time independent events. Fortunately, such declustering techniques are well developed for univariate applications. Whatever technique will be chosen, it should serve to remove any serial correlation in the extremes – to 'prewhiten' the series. To achieve that goal we first have to check where temporal dependency exists and where it originates. We therefore propose to split up severity and frequency components of the individual claims distribution. This leads us to consider a time-varying specification for the frequency distribution, which represents the number of claims per fixed time unit (in our case a day). MLE fitting indicates that the negative binomial seems preferable, particularly after correcting for the time clusters. The latter finding indicates that the Poisson, which is often considered suitable, is not appropriate. Second, claims tend to cluster in size. Certain levels of financial prices seem to occur much more often than predicted by the normal distribution. In the finance literature, this phenomenon is sometimes classified as a psychological barrier, or as an institutionally fixed price increment. In our actuarial case the explanation is much simpler. Either, it concerns claims which are not yet honoured (these data will then reflect reservations based

¹ Beirlant et al. (1994) mention the possibility of clustering, but indicate that '... not much effort has gone into attempts to incorporate clustering in the actuarial analysis'.

on claims), or, it concerns claims that are based on ‘immaterial’ losses which are often denominated in certain discrete multiples. The latter explanation is also known as a cause for heterogeneity in claims. Different components in claim size like personal damage, property damage, and bodily damage, should then be disentangled since they behave according to different distributions. Since we lack the information in the data required for that purpose, we propose an alternative. The empirical distribution ‘betrays’ the different components through discrete ‘peaks’ in a predominantly continuously distributed series. We therefore propose the EM-algorithm which is well suited for simultaneously fitting mixtures of continuous and discrete distributions. This is particularly useful in case of a reinsurance system known as *excess of loss*² which is based on a layered (or sliced) system. The layers are typically bounded by fixed discrete values. Therefore, sudden discrete jumps in probability at these cut-off values will be ignored if the appropriate probability per layer is based on the smooth continuous distribution alone.

We apply a combination of tail parameter estimation and EM-algorithm empirically to automobile collision coverage in the Netherlands. The tail parameter approach indicates that for the continuous severity distribution the Pareto is the most obvious candidate. Using these tail findings as functional input to our EM-algorithm we next find a refined continuous distribution plus a set of discrete logarithmic series distributions that effectively extract the clustered peaks observed in our empirical distribution.

Finally, we combine our findings for the severity and frequency distributions. Temporal aggregation implies a mixture of our severity set with a distribution of claim frequencies per day. A preferable aggregation method accumulates claims over ‘correlated’ events. This is useful when reinsurance is based on *aggregate excess*, i.e. it applies to all losses incurred from a single event. Unfortunately, our dataset does not allow this method. We assume, however, that extreme claims occur in a clustered fashion³. Though not completely

appropriate, under certain assumptions the daily extreme claims distribution may provide similar information as an event claims distribution (including the small and moderate claims) would.

The remainder of this article is organized as follows. Section 2 highlights the tail parameter approach for continuous severity distributions. Section 3 discusses the practical use of the EM-algorithm for mixtures of continuous and discrete severity distributions. In Section 4, temporal aggregation is considered. At this stage we will also discuss time clustering in the frequency distribution. An empirical actuarial application is given in Section 5 for a four-year data set of Dutch automobile collision claims exceeding 25,000 guilders. Section 6 concludes the article with some remarks and suggestions.

2. Residual life and tail parameters

A popular actuarial approach to detect the shape of the underlying distribution of claims (and losses) is the curve fitting method employed in Hogg and Klugman (1983). Their mean residual life model is most easily interpreted in terms of exceedance ‘probabilities’. If X is the length of ‘life’ with probability density function (pdf) $f(x)$, then

$$e(x) = E(X - x | X \geq x) = \int_x^{\infty} (w - x) \frac{f(w)}{\int_x^{\infty} f(w) dw} dw \quad (1)$$

is the mean residual ‘life’ measured in the units of X . Otherwise stated, Eq. (1) gives the expected residual lifetime exceeding some ‘old age’ x . In our case, X becomes a claim (or loss), $f(x)$ is the pdf of a claim distribution, and $e(x)$ is the mean residual claim. Unfortunately, as outlined in the previous section, discrimination among different candidate distributions proves to be very difficult. The distributions that are commonly tested (like the Exponential, Pareto, and Weibull) belong to the generalized beta of the second kind class of distributions (GB2). For an extensive review of this GB2 family we refer to Cummins et al. (1990). Most important, this class encompasses models that

cancel out with small and moderate claims. These ‘ignored’ claims only act as a scaling factor.

² Under this system the reinsurer guarantees the excess amount by which any individual claim or accumulation of claims from one event exceeds the retention level.

³ This implies that simply accumulating the extreme claims per day still identifies extreme daily claims, i.e. they do not

are characterized by fat tails. In general, distributions belonging to the GB2 class have no upper limit for X and are therefore better fit to describe loss distributions which often have infinite (theoretical) endpoints. Even though the relation between densities and parameters is often complex in the GB2, in general the smaller the value of the shape parameter (denoted as α), the fatter the tails of the distribution.

This α , also called the tail parameter, has been estimated by MLE as well as by extreme value methods. To avoid the MLE nesting problems, discussed in the previous section, we will focus on the latter. This approach has also found its way into the actuarial literature, see Beirlant and Teugels (1992), Weba (1993), and Beirlant et al. (1994). We will therefore restrict the exposition of the underlying theory to its main features.

Extreme value theory gives the exact form for the distribution of maxima (i.e. the tails) of random variables regardless of the processes that generate realizations of these random variables. It analyses the limiting distribution of the order statistic M_n , the maximum of a stationary sequence of n i.i.d. random variables. This sequence has an unknown underlying distribution function F . We are particularly interested in the probability of M_n below a certain exceedance level x :

$$P\{\gamma_n(M_n - \delta_n) \leq x\} \xrightarrow{w} G_i(x) \quad (2)$$

where M_n is appropriately transformed by γ (scale parameter) and δ (location parameter). According to Eq. (2), this probability is weakly (w) converging to G_i which typically belongs to a type II limit law:

$$G_{II}(x) = e^{-x^{-\alpha}} \quad (3)$$

for x positive, and zero elsewhere. Of course, the specific underlying distribution, $F(x)$, determines the relevant limit law⁴. However, even without knowledge of this distribution, the qualitative characteristics of the empirical process may yield sufficient information to track down $G_i(x)$. This distinguishing characteristic of type II limit laws is that the tails decline by a power instead of exponentially (as e.g., for the normal and exponential). That implies fat-tailedness of the distribution function $F(x)$. Fat-tailed distributions

like the sum-stable Pareto class, but also Student- t distributions (which are still fat-tailed but have finite variance) are nested within this limit law. The nesting tail parameter α is, respectively, the characteristic exponent and the degrees-of-freedom. An appropriate estimator for α (or its inverse ν), which is consistent with the properties in Eqs. (2) and (3) is given in Hill (1975) as:

$$\hat{\nu} \left(= \frac{1}{\hat{\alpha}} \right) = \left(\frac{1}{m-1} \right) \sum_{i=1}^{m-1} \log X_{(i)} - \log X_{(m)} \quad (4)$$

where m is the number of ordered observations $X_{(i)}$ classified as the upper tail of the distributions. With the condition of regular variation in the tails satisfied, $\hat{\nu}$ is a consistent estimator of ν . In addition: $\sqrt{m}(\hat{\nu} - \nu) \sim N(0, \nu^2)$. As compared to a maximum likelihood estimator, this enhances an efficiency gain⁵.

Once we have an estimate for the tail index (and its inverse $\hat{\nu}$), we can specify the 'residual life' function of the underlying distribution. For given probability levels p , the matching exceedance (or retention) level, \hat{x}_p is given as:

$$\hat{x}_p = \frac{(1/2pn)^{\hat{\nu}} - 1}{1 - 2^{-\hat{\nu}}} \left[X_{(n-\frac{m}{2})} - X_{(n-m)} \right] + X_{(n-\frac{m}{2})}. \quad (5)$$

This level extrapolates the empirical distribution function outside its empirical domain by combining the tail shape at $X_{(n-m/2)}$ and the step size $X_{(n-m/2)} - X_{(n-m)}$, with the way the steps of the limit law change according to the multiplicative factor.

The tail estimator $\hat{\nu}$ has one major drawback in its dependence on the specification of what constitutes the tail of the empirical distribution. For Eqs. (4) and (5) we first need to specify the number of ordered observations m to be included in the tail estimation. Optimally, $m(n) \rightarrow \infty$ but for finite samples this is not feasible. In that case, a large value for m increases efficiency in estimating (4), but simultaneously increases bias by potentially including part of the centre of the distribution. The increment in bias seems to be penalized more than a loss in efficiency. Koedijk et al. (1990) employs a minimizing MSE Monte Carlo experiment but, this approach is rather vulnerable to the

⁴ Dekkers et al. (1989) have extended the Hill estimator from this type II limit law to all cases in (2) above.

⁵ Smith (1987) shows that maximum likelihood can give consistent estimates of ν as well, which are asymptotically normal with mean zero and variance $(1 + \nu)^2$.

underlying distributional assumption. Two simple alternatives are given by Boos (1984), and Galambos (1987):

Boos	$m/n = 0.1$	for	$500 < n \leq 5,000$
	$= 0.05$		$5,000 < n \leq 50,000$
	$= 0.025$		$50,000 < n \leq 500,000$
Galambos	$m = 2\sqrt{n}$		

Simulation estimates in Hasover and Wang (1992) indicate that the Boos measure tends to overestimate whereas Galambos is slightly underestimating the ‘true’ number of tail observations.

An alternative approach is given in Zelterman (1993), and is based on the Gini statistic, calculated as follows:

$$g_m = \frac{\sum_{i=1}^{m-1} i(m-1)(d_{(i,m)} - d_{(i+1,m)})}{(m-1) \sum_{i=1}^m d_i} \tag{6}$$

where $d_{(1,m)} \geq \dots \geq d_{(i,m)} \geq \dots \geq d_{(m,m)}$ are the ordered normalized spacings between the m largest observations of the original series given by:

$$d_i = i(X_{(i)} - X_{(i+1)}), \quad \text{for } i = 1, \dots, m-1. \tag{6a}$$

These normalized spacings d_i are assumed to behave as i.i.d. exponential random variables. Then, a sequential test (where $m = 1, \dots, \infty$) based on the g -estimates of Eq. (6) searches for a value m where these normalized spacings are found to reject the hypothesis of an exponential distribution. Large values for g_m point towards rejection. For each sequential m , the approximate significance level p_m is: $1 - \Phi\{(g_m - 0.5)\sqrt{[12(m-1)]}\}$ where $\Phi\{\cdot\}$ is the cumulative distribution function of the standard normal. It should, of course, be realized that (once again) the test is dependent on a distributional assumption.

We opt for the heuristic procedure proposed by Loretan and Phillips (1994), by estimating α for different m -values, and select an m -value where $\hat{\alpha}$ is relatively stable. It seems that for n very large (as it is in our application), the estimate is not very vulnerable to errors in the choice of m .

2.1. Tail stability

To test whether the tail behaviour is stable over time, we can make a comparison by means of the Q_α test

statistic for equivalence across sub-periods:

$$Q_\alpha = \left(\frac{\alpha}{\hat{\alpha}_1} - 1\right)^2 m_1 + \left(\frac{\alpha}{\hat{\alpha}_2} - 1\right)^2 m_2 \tag{7}$$

where the α estimate is given by Eq. (4), and the appropriate m -values are generated by one of the above outlined procedures. Q_α is χ^2 distributed with 2 degrees-of-freedom. Tails equality is rejected if there is no $\alpha = \alpha_1 = \alpha_2$ for which Q_α is below the critical value (5.99 at the 5% significance level). If such an α does exist, however, a confidence interval of α 's exists for tail parameters that both subsets have in common.

An alternative sample split prediction test is given in Loretan and Phillips (1994):

$$\tilde{V}(\hat{\alpha}_1 - \hat{\alpha}_2) = \frac{\sqrt{m} \cdot (\hat{\alpha}_1 - \hat{\alpha}_2)}{\sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2}} \tag{8}$$

where the samples are of equal size. If $m \rightarrow \infty$ and $n \rightarrow \infty$ then $\tilde{V}(\cdot) \sim N(0, 1)$ under the null hypothesis of equal tail parameters. Loretan and Phillips show that $\tilde{V}(\cdot)$ has good empirical size properties under the null of equality. Furthermore, the test is consistent under the alternative that the tails are strictly Paretian. Alternatively, the test is inappropriate if higher than second order moments do exist. This is unlikely in our application.

3. Mixtures of severity distributions revealed by an EM-algorithm

Once we have an indication of the most appropriate distribution class, as revealed by the tail parameter, we can proceed by focussing in more detail on particular characteristics of the tail. For that reason we switch to maximum likelihood estimation. There are, however, a couple of problems that have to be dealt with first. Unfortunately, actuarial data sets are hardly ever complete. In our case we will be restricted to observations exceeding some pre-determined level. This need not be a severe problem since we are predominantly interested in the probability behaviour of the tails of the distribution. In fact, our tail parameter approach assumed that inference from the tails reveals the most appropriate underlying distribution. However, the censoring implies that we have no information whatsoever on the probability distribution in the main ‘body’ of the

distribution⁶. This is quite unlike many other actuarial applications where truncation takes place at some maximum level due e.g., to maximum liability provisos. Truncations of our minimum-type can and should be taken into account by simply modifying the likelihood function, see also Hogg and Klugman (1983):

$$f_{s,A}(x) = \frac{f_s(x)}{1 - F_s(A)} \quad (9)$$

where A is the cut-off level, $f_s(x)$ is the pdf and $F_s(A)$ is the cumulative probability evaluated at a pre-specified exceedance or retention level A .

In addition, we will have to deal with any clustering biases that may be present in our application. We will first focus on clustering aspects relevant for individual claims, i.e., the severity distribution F_s in Eq. (9). Time dependency of claims distributions will be dealt with in Section 4, and is shown to be of minor importance for F_s . More importantly, certain claim sizes occur much more often than can be explained by any of the proposed continuous (Exponential, Pareto, Weibull etc.) distributions. This phenomenon is known as *size clustering*. We are especially interested in the case where certain claim sizes become the sum of several different distributions. Part of the number of occurrences of that particular claim size can be captured by the continuous distributions, but this will fall short of the observed number. Hence, we have to fit alternative (discrete) distributions to the excess occurrence. That implies a mixture of distributions approach where the size clusters, observed in the empirical density function, are fitted separately from (or in addition to) an underlying continuous distribution.

The EM-algorithm allows us to achieve exactly that. Dempster et al. (1977) give an extensive treatment of this iterative likelihood maximizer. Ruud (1991) shows that the application possibilities extend to missing data problems. We will consider frequencies at certain discrete multiples of the smallest observed cluster to be unknown.

Consider claim size as the random variable X with a Pareto probability density function derived from our extreme value analysis (though we can substitute the Pareto for any fat-tailed alternative):

$$f_s(x_i|\alpha) = \frac{\alpha\delta^\alpha}{x_i^{\alpha+1}} \quad \text{for } \delta < x_i < \infty. \quad (10)$$

According to Eq. (9), we will interpret δ as the truncation value A . If the continuous Pareto fully captures $\{X\}$, we could maximize the log likelihood function:

$$\log L(\alpha|x) = \sum_i [\log(\alpha) + \alpha \log(\delta) - (\alpha + 1) \log(x_i)], \quad \delta < x_i < \infty \quad (11)$$

or alternatively:

$$\log L(\alpha|w) = \sum_{cs} w_{cs} [\log(\alpha) + \alpha \log(\delta) - (\alpha + 1) \log(cs)], \quad \delta < cs < \infty$$

and $w_{cs} \geq 0$ (12)

where w_{cs} is the number of occurrences of claim size cs that can be fitted to the continuous pdf f_s . In our case w_{cs} is latent, since we do not know its value if cs is some multiple k of the smallest prespecified discrete value D which constitutes a size cluster. We will indicate the actually observed number of occurrences of claim size cs by w_{cs}^* . Dempster et al. (1977) show that we can write:

$$\log L(\alpha|w^*) = \int \log L(\alpha|w^*, w) f(w|\alpha', w^*) dw - \int \log f(w|\alpha, w^*) f(w|\alpha', w^*) dw + C. \quad (13)$$

Parameter α' can take any value allowed in the parameter space. C is a constant with respect to the parameter α . The first term on the right-hand side of (13) is usually denoted by $Q(\alpha, \alpha')$, which is an expression for the expected value of the log likelihood, $L(\alpha|w)$. Dempster et al. show that it is sufficient to iteratively maximize $Q(\cdot, \cdot)$ to find stationary points of $L(\alpha|w^*)$. These iterations proceed as follows:

- (1) Choose a value for $\alpha' = \alpha_0$.
- (2) Calculate $Q(\alpha, \alpha')$ for this chosen value of α' .
- (3) Maximize $Q(\alpha, \alpha')$ with respect to $\alpha (= \alpha_1)$.
- (4) Replace α' by the optimized value obtained in step 3.
- (5) Return to step 2 and repeat this procedure until $\alpha_i = \alpha_{i-1} = \alpha^*$.

In our case, based on Eq. (11), if w_{cs} is observed ($= w_{cs}^*$):

⁶ This implies that we cannot deal with reinsurance schemes that are pro rata (or proportional). These require knowledge on the complete distribution.

$$Q(\alpha, \alpha'; w_{cs}) = w_{cs}^* [\log(\alpha) + \alpha \log(\delta) - (\alpha + 1) \log(cs)] \tag{14}$$

and, if w_{cs} is not observed:

$$Q(\alpha, \alpha'; w_{cs}) = E(w_{cs} | w^*, \alpha') [\log(\alpha) + \alpha \log(\delta) - (\alpha + 1) \log(cs)] \tag{15}$$

where ⁷.

$$E(w_{cs} | w^*, \alpha') = n [f(cs | \alpha')] \tag{16}$$

since w_{cs} follows a binomial distribution with probability:

$$f(cs | \alpha') = \frac{\alpha' \delta^{\alpha'}}{cs \alpha'^{+1}}$$

of ‘success’ (=observing claim size cs), and n trials, we know that:

$$f(w_{cs} | \alpha') = \binom{w_{cs}}{n} \left(\frac{\alpha' \delta^{\alpha'}}{cs \alpha'^{+1}} \right)^{w_{cs}} \times \left(1 - \frac{\alpha' \delta^{\alpha'}}{cs \alpha'^{+1}} \right)^{n-w_{cs}} \tag{17}$$

where n is the sum of w_{cs} , including the expected values $E(w_{cs} | w^*, \alpha')$. Both n and $E(\cdot)$ are determined iteratively and are updated in each iteration in the EM-algorithm. Having found the expected number, we next sum $Q(\alpha, \alpha' | w_{cs})$ over all claim sizes cs and maximize the sum $Q(\alpha, \alpha')$ with respect to α . Then, calculate $Q(\alpha, \alpha' | w_{cs})$ for the new value of α' and go once again through the maximization steps, etc., until α' converges to α^* .

To calculate the standard errors for the estimated parameter, we use Louis’ method as given in Tanner (1993). Tanner gives four alternative methods, but we prefer Louis’ for its computational ease in calculating the information matrix (for our case):

$$\frac{\partial^2 \log L(\alpha | w^*)}{\partial \alpha^2} \Big|_{\alpha^*} = \frac{\partial^2 Q(\alpha, \alpha^* | w)}{\partial \alpha^2} \Big|_{\alpha^*} + \text{var} \left(\frac{\partial \log L(\alpha | w_{cs}^*, w_{cs})}{\partial \alpha} \Big|_{\alpha^*} \right) \tag{18}$$

⁷ To calculate $E(\cdot, \cdot)$ we do have to discretize the pdf. We will therefore apply a continuity correction to the probability calculation.

$$= \sum_{cs} -E(w_{cs} | w_{cs}^*, \alpha^*) / \alpha^{*2} + \sum_{cs} \left(\frac{1}{\alpha^*} + \log(\delta) - \log(cs) \right)^2 \times n \left(\frac{\alpha^* \delta^{\alpha^*}}{cs \alpha^{*+1}} \right) \left(1 - \frac{\alpha^* \delta^{\alpha^*}}{cs \alpha^{*+1}} \right).$$

Note that we will discretize this continuous distribution as mentioned in footnote 7. The standard error then becomes:

$$\text{Standard error} = \left(-\frac{\partial^2 \log L(\alpha | w_{cs}^*)}{\partial \alpha^2} \Big|_{\alpha^*} \right)^{-\frac{1}{2}}. \tag{19}$$

The same steps as described so far, will then be repeated for the most likely discrete mixture of candidate distributions. These should capture the observed peaks in the empirical distribution function. Since some discrete multiples k seem to occur more often than others, we will further distinguish between different ‘layers’ of discrete distributions. Appropriate candidates for a distribution require maximum probability at its minimum value and long (fat) tails. A truncated discrete lognormal distribution displays the first characteristic, but may be inappropriate in taking account of the excessive tail probability mass. On the other hand, a truncated negative binomial distribution seems to have too little mass at the smaller values. The logarithmic series distribution is more likely to satisfy both prerequisites: high probability at the minimum, steeply declining, but still long tails. Its probability mass function (pmf):

$$P_s(x_i | \theta) = -(\log(1 - \theta))^{-1} \left[\frac{\theta^x}{x_i} \right] \tag{20}$$

has shape parameter θ , where $0 < \theta < 1$. The closer θ is to its bounds, the larger the kurtosis and skewness of the distribution.

Repeating the previous steps, we get:

$$Q(\theta, \theta'; w_{cs}) = E(w_{cs} | w^*, \theta') [cs \log(\theta) - \log(cs) - \log(-\log(1 - \theta))]$$

where

$$E(w_{cs} | w^*, \theta') = n \left[\frac{(\theta')^{cs}}{-cs \log(1 - \theta')} \right] \text{ when } w_{cs} \text{ is not observed} = w_{cs}^* \text{ when } w_{cs} \text{ is observed} \tag{21}$$

instead of Eqs. (15) to (17). Likewise, we can find an expression for the standard error of θ^* , our optimized (set of) parameter(s).

4. Time clusters in frequency distributions

After disentangling the mixture of severity distributions, we finally consider temporal aggregation issues. In fact, if time dependency would disappear under temporal aggregation, we can easily avoid the modelling complications of Section 3 by using more aggregated data. Of course, for efficiency considerations we would like to use as many observations as possible. Furthermore, if the underlying distribution is not stable under addition, we cannot make tail inferences from the aggregated estimates that will also be valid for higher (disaggregated) frequencies as well. That poses some problems if the tails have to be used for setting reinsurance premiums.

It is, however, also possible that we want tail information for different purposes, like e.g., ruin probability measurement. In that case we would not be so much interested in individual claims but much more in accumulated claims over some time period. Using the maximum amount of information would imply that we have observations that are not measured in constant time-spaced intervals. Individual claims tend to arrive in irregular order. This means that if an insurance company wants to know the probability of a claim exceeding some predetermined level, this cannot automatically be expressed in terms of ‘on average once every so many days’⁸. This shortfall has to be accommodated for by aggregating claims over days or even longer periods. The compound distribution F_C for total claim size per day, X_C , is given as:

$$F_C(X_C) = \sum_{y=0}^{\infty} p_y F_s^y(x) \tag{22}$$

where $F_s^y(x)$ is the y th convolution of the individual claims severity distribution and p_y is the probability that the daily number of claims is y . The cumulative X_C should then be decomposed into its frequency

component (the number of claims per day, y) and the already discussed severity component (the size of each individual claim X). This decomposition allows an optimal exploitation of sample size from the available database.

If the aggregated X_C provides a sufficiently large sample, one can still try to fit a GB2 distribution to this set as a further check on the reliability of the compounding procedure that is presented in the next section. This does, however, require some strong assumptions, like a Poisson distributed frequency distribution⁹. We will avoid such prior assumptions and, instead, suggest a mixing process. Since we already fitted the most appropriate severity distribution, we will now separately fit a frequency distribution and, consequently compound the two.

Obvious discrete candidates for the frequency distribution are the Poisson:

$$f_f^{\text{Poisson}}(y) = \frac{\mu^y e^{-\mu}}{y!}$$

$$\log L(\mu|y) = \sum_i \{-\mu + y_i \log(\mu) - \log(y_i!)\} \tag{23}$$

for $0 \leq y_i$ with mean and variance $\mu > 0$; and the not so often used Negative Binomial:

$$f_f^{\text{negbin}}(y) = \frac{\Gamma(\kappa + y)}{\Gamma(\kappa)y!} p^\kappa (1 - p)^y$$

$$\log L(\kappa, p|y) = \sum_i \{\log[\Gamma(\kappa + y_i)] - \log[\Gamma(\kappa)] - \log(y_i!) + \kappa \log(p) + y \log(1 - p)\} \tag{24}$$

for $0 \leq y_i$ with mean $\kappa(1 - p)/p$, and variance $\kappa(1 - p)/p^2$. Whereas the Poisson is appropriate when the mean of y equals its variance, the negative binomial is better suited for cases where the variance exceeds the mean.

In this case, time dependency may appear. Time clustering may enter the severity distribution through inflationary pressures. A discounted claims solution is given in Willmot (1989) for a number of distributions, including fat-tailed ones that have regularly varying

⁸ It would mean ‘on average every so many submitted claims’. Actually, this distinction is equivalent to the Maximum Probable Yearly Aggregate Loss (MPY) versus Probable Maximum Loss, see Cummins and Freifelder (1978).

⁹ In that case, the aggregated claims and individual claims follow a compound Poisson process which have identical tail indices. The tail parameter will then be stable under addition.

tails. For our application this turns out to be a negligible effect, in part due to the short (4-year) time span that we consider. For the frequency distribution of claims, however, time clustering may be of much more importance. Bad weather days or seasons, tend to increase the number of claims (including large claims) per day, hence make the frequency time dependent. Or, the number of policies sold increases over time, and (with a constant claim ratio, i.e. ratio of claims to premiums) likewise the number of claims. In either case, we will have to correct for temporal dependency in the frequency distributions. This can be achieved, e.g. by giving κ in Eq. (24) a conditional interpretation:

$$\kappa_t = C + \rho * y_{t-1} \quad (25)$$

where y is the number of claims per day. Such a specification can, of course, easily be extended to more general ARMA-type settings, or particular seasonal models if necessary.

4.1. Compounding claims distributions

After fitting the individual (declustered) components, the next step is to compound frequency and severity distributions, according to Eq. (22). Unfortunately, only in certain cases a closed form solution will be available. A set of available approximations is given in Pentikainen (1987). These are broadly distinguished into normalizing (analytical) and moments generating (numerical) approaches. A pragmatic solution, based on a simulated percentiles approach, is given in Aiuppa (1988). However, his approach is prone to a downward bias since this technique fits the theoretical moments of the compound distribution so that they equal their empirical counterparts. For distributions with infinite (theoretical) higher moments this is obviously not correct. In general, it seems that these techniques perform particularly well for scenarios that also have closed form solutions. For highly kurtotic or skewed claims distributions they turn out to be less helpful.

Furthermore, the discreteness hidden in the continuous severity distribution as displayed by our application, may considerably bias both the normalizing and empirical moments approaches. We propose therefore the following simple experiment. Based on our severity and frequency findings in the previous sections we employ the fitted probability distributions in a two-step method. First, we generate some large number

of daily frequencies according to (time dependent adjustments of) Eqs. (23) and (24). Next, we generate severities for each generated frequency. For this purpose, we will generate random numbers in the interval $[0,1]$ where each drawing is related to a probability interval. The first part of the interval will relate to the discrete proportion of the severity mixture. Say, we find the probability for the smallest discrete value $P(x_s = D) = x_D\%$, we declare the value D to the (sub)interval $[0, x_D\%]$. For each multiple k of D , this will immediately generate the required claim size. For the remaining $1 - \sum P(x_s = kD)$, we will draw randomly from the appropriate (truncated) continuous distribution.

Finally, we can compare this mixture of data generating processes with the observed daily claim sizes. To check whether the two distributions are different we use a Kolmogorov–Smirnov goodness-of-fit test, where we are comparing one data set's empirically estimated cumulative distribution function (cdf) $F^*(X_C)$ with our known data generating process, with cumulative distribution function $F(X_C)$:

$$KS = \max_{-\infty < X_C < \infty} |F^*(X_C) - F(X_C)|. \quad (26)$$

However, due to the mixture setup we may also prefer to evaluate the cdf's at the discrete multiples for both known and unknown functions, to evaluate whether these blips are accurately modelled.

5. Automobile collision claims

We apply the previously discussed techniques to a dataset acquired from the Centre for Actuarial Statistics (CVS). The CVS database currently represents almost 50% of Dutch automobile insurance companies. In addition to the fact that coverage has been growing over time, the included reporting companies need not be the same over the considered time period. The data pertain to automobile collision claims for the period 1989 to 1992. Our set is restricted to claims exceeding 24,999 Dutch guilders. Table 1 indicates that this truncation covers 2.6% (2.2 plus 0.4%) of all 1992 claims registered by the CVS (for 1990 and 1991 this was 1.8 and 2.0% respectively).

To identify the tail (i.e., the required number of observations in our tail estimation) we use the only

available complete set of observations for 1992. The first four moments of this non-truncated set are:

Mean (1992) = 4,659 Skewness (1992) = 37
 St. dev. (1992) = 20,673 Kurtosis (1992) = 2,217.

This leads to a heavily skewed and fat-tailed frequency distribution for 1992, see Fig. 1. Visual inspection reveals a rather short lower (left) tail whereas the upper (right) tail seems extremely long. This reflects the excessive kurtosis measure.

Next, we classify the tail part of Fig. 1. In Table 2 some evidence is given for the optimal tail size according to the four techniques that have been proposed in Section 2.

The Gini-coefficient tends to stabilize around $m > 1,750$. In addition to the Gini-coefficient, which could only be calculated for 1992, we also give the optimal m -values for the Boos, Galambos and MSE procedures. Since we do not know the total number of claims for 1989, none of these techniques can be applied to that year. The rather significant different outcomes may reflect the relative importance of the discussed size clusters. This can be highlighted by constructing a quantile plot on the log-transformed data, where the log of the upper empirical quantiles is set out against theoretical exponential quantiles. If the data are Pareto distributed (anticipating on the results in Table 3) this should be a straight line. However, if some values appear too often, this will show up as horizontal lines, the length of which reflects the size of the cluster. Obviously, any linear regression through this plot will be wrong¹⁰.

¹⁰ We are grateful to the referee for pointing this out. The referee also suggested that the model on which the different techniques in Table 2 are based, will only be valid for the most extreme part of the distribution, after the last size cluster.

Table 1
 Frequency distribution of claim sizes

Claim size	1992	
0 < 1,000	39,721	24.2%
1,000 < 2,500	61,843	37.7%
2,500 < 10,000	50,942	31.0%
10,000 < 25,000	7,425	4.5%
25,000 < 100,000	3,590	2.2%
≥ 100,000	662	0.4%
Number of claims	164,183	

Since the four-year data set contains claims in excess of 24,999 guilders, the last column gives the number of observations per year exceeding that level. The likelihood application is evaluated at truncation point A which we will define at 25,000 guilders (thereby maximizing the use of our data-set). For the tail parameter approach, we will experiment with different measures for m , as suggested by Loretan and Phillips (1994).

The complete 1992-set, as well as the >24,999-set (and the cumulative daily set) pass the standard stationarity tests (Augmented Dickey–Fuller and Phillips–Perron unit root tests) despite some evidence of time clustering in the aggregated daily claims. This allows us to apply the tail parameter estimator.

5.1. Parameterizing the severity distribution tail

Since the EM-algorithm depends on the maintained hypothesis of underlying distribution, we first want to indicate the appropriate class. This is achieved by applying the extreme value theory and estimating the tail parameter. The first three rows in Table 3 give the estimates for Eq. (4) for three different lengths of tails. These choices more or less coincide with the different m -levels in Table 2, respectively for Boos/MSE, Gini, and Galambos.

Surprisingly, all the information on the tail shape seems to be conveyed in a small number of tail observations relative to the complete size of the empirical distribution. In addition, tail parameters are given for two bi-annual subsets to test whether the α estimate is stable over time. Even though the estimates are apparently quite close, equivalence is rejected by both the Q_α -test and \tilde{V}_α sample split prediction test. Nevertheless, all estimates in Table 3 are significantly different from both 1 (which would have implied an infinite mean) and 2 (implying infinite variance) indicating that the tail is Pareto-like.

We agree in principle, however, the most extreme observations also turn out to be size-clustered values. We therefore opt for a rather pragmatic approach to estimate the tail parameter in the first stage.

Table 2
Estimators for m , the number of tail observations.

Total number of claims (year)	Gini-coefficient	Boos	Galambos	MSE	$\geq 25,000$
unknown (1989)	n.a.	n.a.	n.a.	n.a.	2,321
166,469 (1990)	n.a.	4,162	816	3,330	3,045
153,880 (1991)	n.a.	3,847	785	3,110	3,050
164,183 (1992)	1,750	4,105	810	3,320	4,252

n.a.= not available.

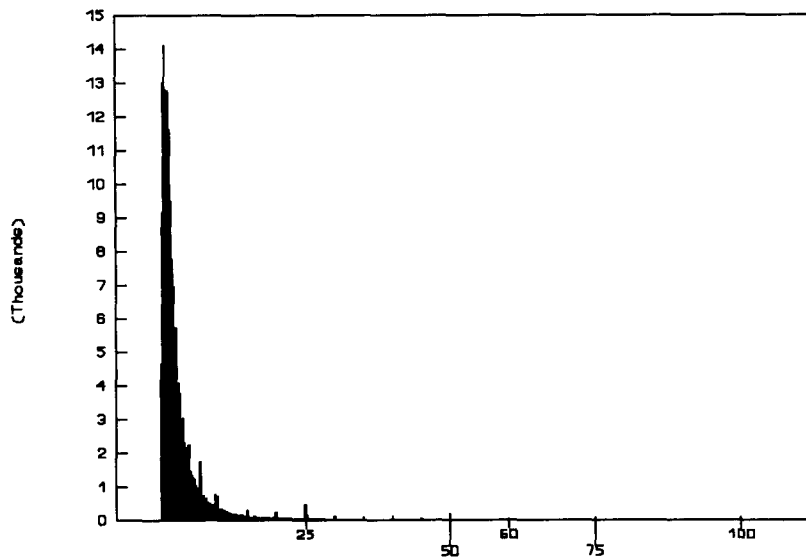


Fig. 1. Frequency distribution for all 1992 claims.

Table 3
Tail parameter estimates

Tail samples	Tail parameter (confidence interval)	Q_α -test	$\tilde{V}(\cdot)$ -test
Full sample ($m = 12,666$)	1.370 (1.346–1.394)		
Full sample ($m = 8,000$)	1.322 (1.287–1.345)		
Full sample ($m = 4,000$)	1.379 (1.346–1.432)		
Subset 1 (89–90) ($m = 5,366$)	1.320 (1.285–1.355)	Subset 1 vs Subset 2	3.629**
Subset 2 (91–92) ($m = 7,300$)	1.408 (1.375–1.440)	Rejected	Rejected ^a

^a We have assumed that the samples are of equal size. This means that we have used $m = 6,333$ in Eq. (8).

5.2. The EM-algorithm approach for the severity distributions

Having identified an appropriate continuous distribution class, we can now proceed with the further refinement of the estimates. Basically, this means that we will investigate whether some peaks in the em-

pirical frequency distribution (or horizontal lines in the quantile plots) can explain the fat-tailedness as revealed by the tail parameter. These size clusters are typically due to immaterial loss claims (usually bodily damage) or an occasional disaster loss claims. In those cases claims will be denominated in large fixed amounts (round numbers), very often coinciding with

lengthy delays in final claim settlements. In that case they may also reflect reservations for future claims.

Size clustering is clearly indicated in Figs. 2 and 3, where only the larger claims are given. The obvious horizontal lines in Fig. 2 and spikes in Fig. 3 represent ‘excess’ occurrence at multiples of 5, 10, 25, 50 and 250 thousand guilders (hence, $D = 5000$ and $k = 1, 2, 5, 10, 50$ left-truncated at 25000). The EM-finding will help us measure the extent of bias in using simple procedure, like the tail parameter, for analyzing the tail distribution.

Table 4 gives the EM-estimates for the continuous Pareto-distribution and the set of 5 discrete logarithmic series distributions as outlined in Section 3. The first column gives the appropriate parameters for each distribution, the second column gives the log-likelihood and the last column gives the expected number of observations per distribution. Obviously not all of the discrete layers perform equally well. The similarity in shape for each multiple is, however, remarkable. Each of them is heavily fat-tailed, most of all for the smallest multiples, $k=1, 2$. To check the relative performance of this mixture versus a simple single continuous distribution, Fig. 4 indicates the comparative error of each fit.

The vertical axis measures the difference between the cumulative density error (fit minus empirical) of the simple continuous and the mixture, respectively. The predominantly positive values indicate that the mixture outperforms the continuous (except for a somewhat prolonged claim layer: 250,000–500,000).

Having discovered the mixture of distributions for the severity, we can utilize this information to find appropriate exceedance probabilities that are necessary to determine reinsurance premiums. Fig. 5a, gives some information.

The sudden probability drops indicate the importance of taking the discrete blips into account. The information revealed in the graph tells us that e.g. there is a probability of 1.95% on claims exceeding a retention level of 24,999 guilders. Shifting the retention level by one guilder to 25,000 leads to a (relatively large) drop in probability: 1.76%. According to a ‘simple’ continuous fit, we find a probability of 1.94%.

We can also exploit the mixtures information in calculating probabilities or estimated number of claims for a sliced (or layered) excess loss system. In some cases it may be attractive for the ceding insurer to

Table 4
EM-estimates for severities

Distributions	Parameters ^a	– log Likelihood	Expected number ^b
Pareto	$\alpha = 1.3634$ (0.0147)	39,960.78	8,604
Logarithmic series for 5,000 multiples	$\theta = 0.9217$ (0.0057)	3,982.82	1,739
Logarithmic series for 10,000 multiples	$\theta = 0.9306$ (0.0075)	932.89	392
Logarithmic series for 25,000 multiples	$\theta = 0.7928$ (0.0127)	1,832.90	1,125
Logarithmic series for 50,000 multiples	$\theta = 0.8483$ (0.0127)	1,460.33	768
Logarithmic series for 250,000 multiples	$\theta = 0.7893$ (0.0488)	81.14	46

^a Standard errors according to Eq. (19), are given in parentheses.

^b Expected numbers are calculated according to Eqs. (16), (17) and the explanation given on the calculation of n , for the optimized parameter values.

reinsure part of the tail while tolerating some small excess probability. Table 5 illustrates the importance of extracting the discrete peaks from a continuous distribution, and modelling them separately.

If we were to determine the expected number of claims, simply based on the continuous distribution, we could expect an underestimate of about 33% of the ‘true’ number of expected claims for that particular layer.

Using the extrapolation property of the extreme value approach in Eq. (5) we can also estimate the probability on claims exceeding catastrophic levels e.g. 3 million guilders. This is important, since the empirical distribution function may not contain any claims larger than such a catastrophic level and will therefore not be informative for these purposes. Even though catastrophic claim information may be missing in the empirical distribution, we can still extrapolate the tail (since we know its shape) to include these non-observed claim levels. According to Eq. (5), and the information on the probability contained in the discrete peaks we construct Fig. 5b.

The top line indicates the ‘smooth’ continuous fit, the bottom line incorporates the discrete distributions as well. The 250,000 blips ($k = 50$) are still visible. Obviously, the continuous fit overestimates the extreme tails, due to its smoothing of the discrete probability mass at lower claim levels. It is found e.g., that claims larger than 3 million guilders would occur on average once every 55,377 claims (0.00181% proba-

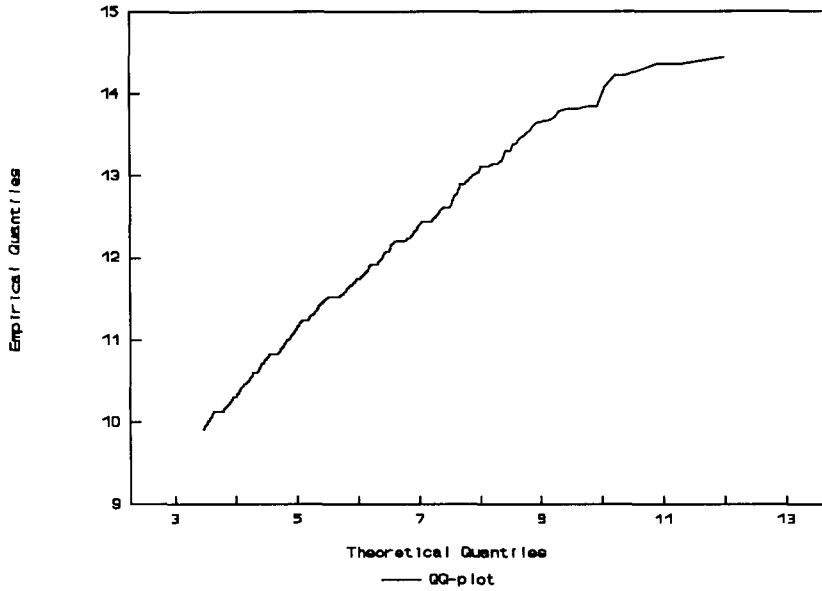


Fig. 2. Quantile plot for the upper tail of 1992 claims.

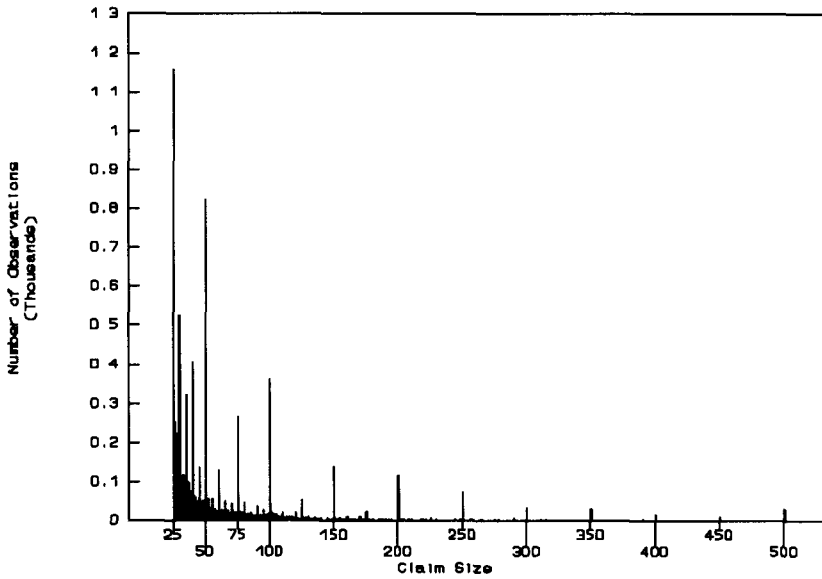


Fig. 3. Spiked frequency distribution for claims exceeding 24,999 guilders (four-year sample period).

bility of exceedance). This is substantially less than its ‘simple’ continuous equivalent (once every 35,050 claims). However, it may still be a rather unpleasant finding for insurance companies. Combined with ruin probability theory this information can be helpful in determining maximum liability levels.

5.3. Dependency in the frequency distribution

Having estimated a specification for the severity distribution, we next evaluate the number of submitted claims per day. The empirical frequency distribution of these submitted claims frequencies, y in Eqs. (23)

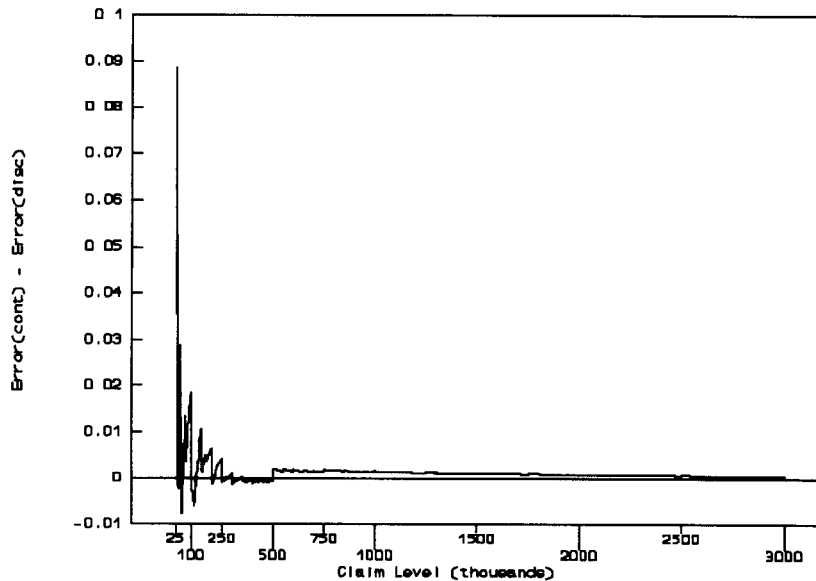


Fig. 4. Cumulative density errors.

Table 5
Continuous versus mixtures probabilities in retention layers

Layers in guilders (including the limits)	Expected continuous number of claims (a)	Expected continuous + discrete number of claims (b)	Relative percentage error $1 - (a)/(b)$
25,000–50,000	5,260	8,018	0.34
25,000–100,000	7,304	10,757	0.32
100,000–250,000	927	1,384	0.33
100,000–1,000,000	1,244	1,853	0.33

Columns (a) and (b) are based on the EM-findings for the iterated value of n .

and (24), is given in Fig. 6 with a mean value of 8.66, and a variance of 16.29. As suggested in Cummins et al. (1990), we fit both Poisson, and Negative Binomial to this set. Since the variance is about twice the size of the mean, we expect the Negative Binomial to be more appropriate. Maximum likelihood estimates are given in the first two rows of Table 6, marked 'independent'.

The Negative Binomial does indeed give a superior fit with estimated mean $\kappa(1-p)/p = 8.66$, and variance $\kappa(1-p)/p^2 = 16.13$, which compares nicely to the empirical moments. This is an interesting finding since many applications commonly assume that frequency distributions behave Poisson-like without any formal testing whether that is indeed the case. It also implies that the alternative to compounding as suggested by footnote 7, is not appropriate.

So far, we have given estimates without taking account of potential clustering in time. Box–Jenkins

testing for the severity data rejects serial correlation ($Q_{BJ}(12) = 7.4$), whereas autocorrelation is strongly present in both daily claim aggregates as in the daily frequencies series (respectively $Q_{BJ}(12) = 48$ and $Q_{BJ}(12) = 1238$). This suggests that the frequency distribution is 'driving' the temporal clustering phenomenon. Some evidence for this is provided in Fig. 7, where daily number of claims seems to be slowly upward trending.

There are two possible sources for this rise. Either the coverage of the Centre for Actuarial Statistics has risen in terms of more insurance agencies or in terms of the number of policies (implying that the 'real' number of claims may be constant). Or there has been a real rise in extreme claims per policy. Obviously, more information is required as to the contents of the 'real' tails (those that represent the exposure of the insurance agencies).

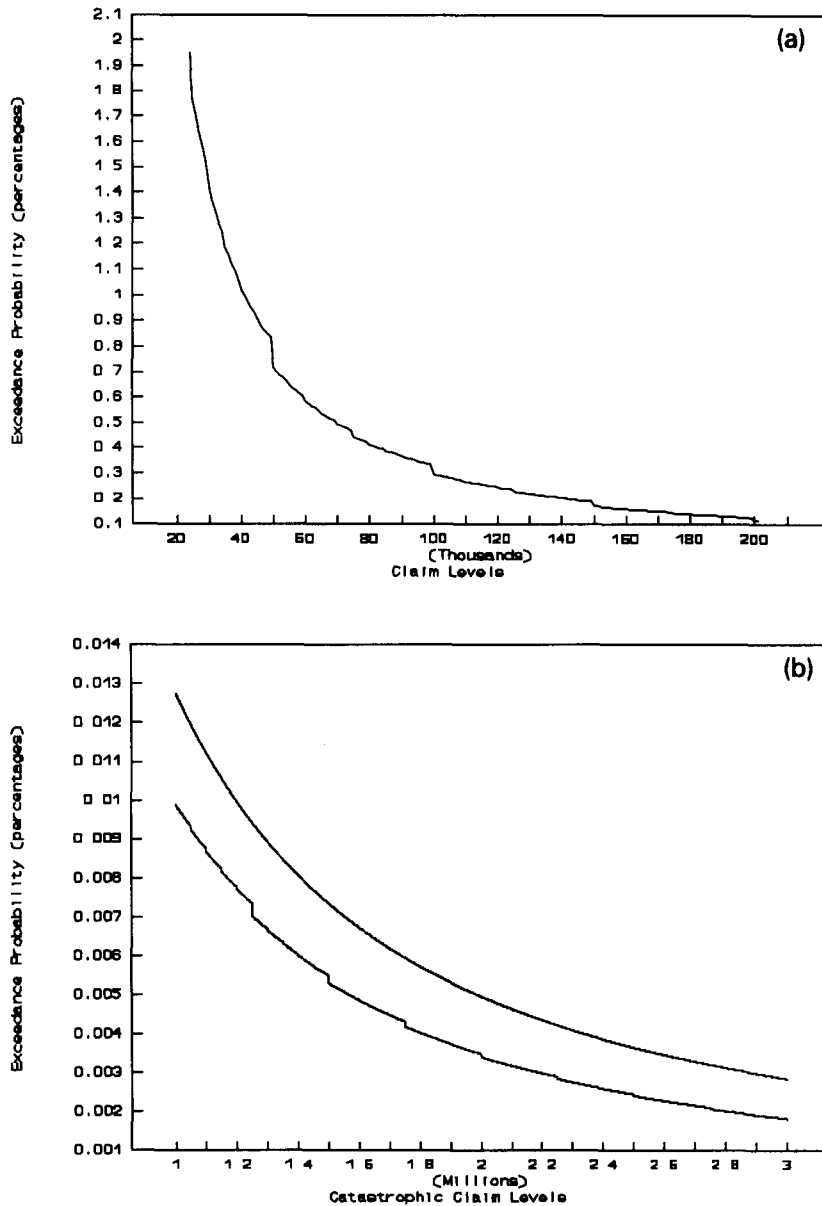


Fig. 5. Retention exceedance probabilities (a) and Catastrophic claim probabilities (b).

To gauge the effects of temporal dependency, the same likelihood fitting has been applied after substituting Eq. (25) in Eq. (24). The results are given in rows 3 through 7, Table 6, marked ‘dependent’. Row 3 gives the dependent estimates for the full sample, and is therefore comparable to row 2 by a $\chi^2(1)$ likelihood ratio test. To check whether this marked improvement is due to any particular year, rows 4 through 7 indi-

cate that ρ is only significant for 1992. However, for each year Eq. (24) is (marginally) improved upon by giving it a time-dependent interpretation.

5.4. A final compounding mixture

Having identified both severity and frequency components we can now proceed by compounding them.

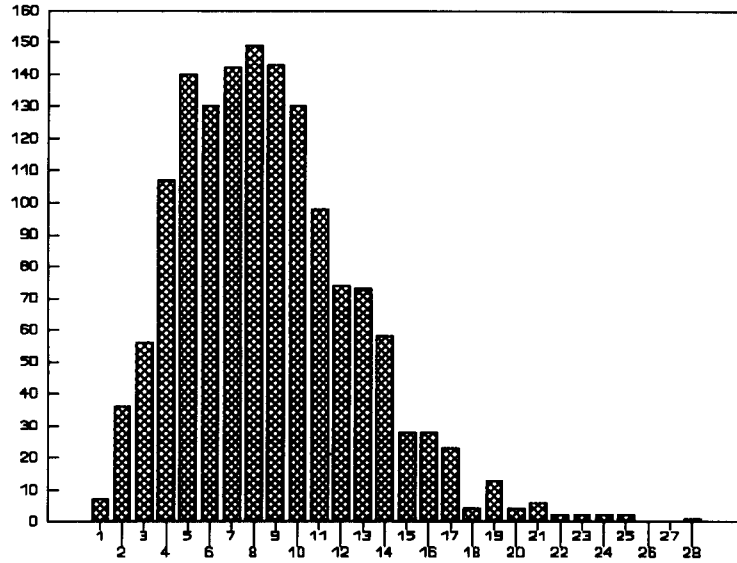


Fig. 6. Frequency distribution of daily claim frequencies.

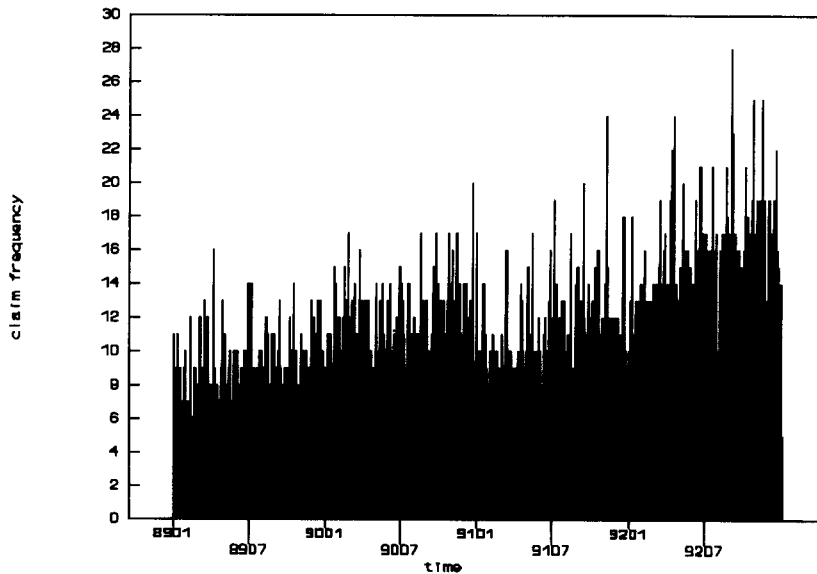


Fig. 7. Frequency–Daily number of submitted claims.

A mixture of distributions model follows where the claims count acts as the conditioning (and in this case secondary) mixing variable. A four-year sample (of 1461 daily observations) for the experiment described in Section 4 provides a simulated frequency distribution for daily claims where class size is set equal to 10,000 guilders: Remember that we only draw from claims larger than or equal to 25,000 guilders. We can,

however, still have outcomes in the interval $[0, 25000]$ since the frequency count can be equal to zero. What we observe in Fig. 8 is that the discreteness tends to disappear for aggregated daily claims. It is, however, still visible in the upper tail.

Finally, we check whether this experiment fits the empirical observations. For that purpose we give the difference between $F(X_C)$ and $F^*(X_C)$ in Fig. 9.

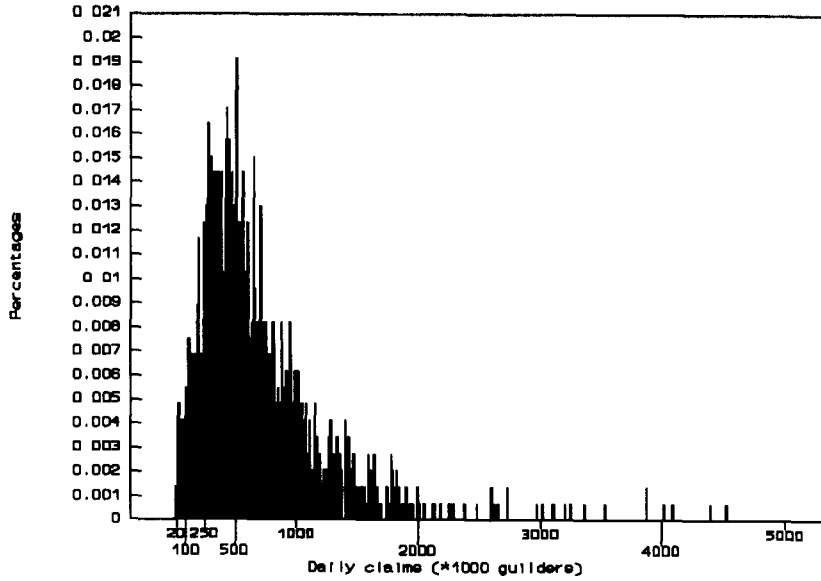


Fig. 8. Simulated daily claims distribution.

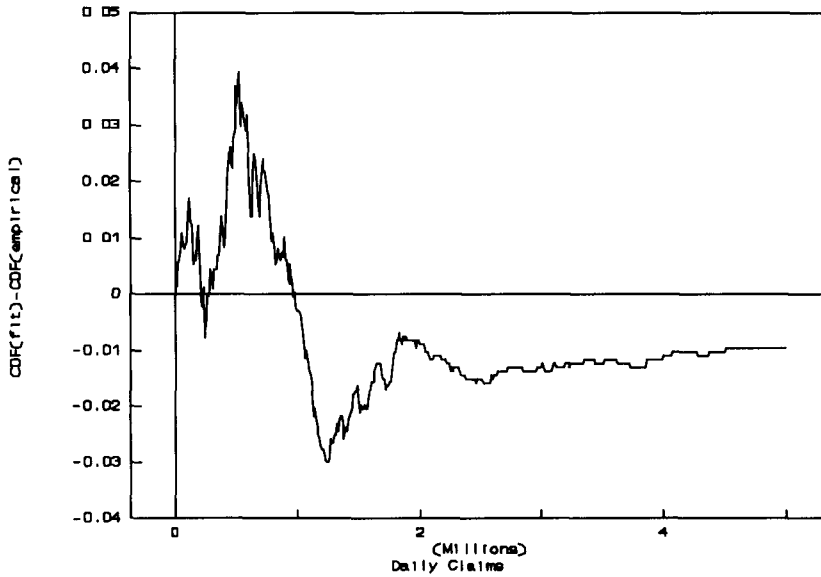


Fig. 9. Cumulative probability errors.

Since we limited our experiment to $n = 1461$ days, the error does not converge very quickly. However, the Kolmogorov–Smirnov KS-test statistic in Eq. (26) can not be rejected at the 95% significance level. Hence, we cannot reject the null hypothesis of equal distributions.

6. Conclusions

This paper has shown that both traditional maximum likelihood fitting and the tail parameter approach can be usefully combined to capture the basic characteristics of a (typical) actuarial dataset. Whereas the

Table 6
Likelihood estimation for frequencies

Distribution	Parameters ^a	– log Likelihood ^b
Poisson independent	$\mu = 8.662(0.106)$	4,214
Neg. binomial independent	$\kappa = 10.048(0.8167)$ $p = 0.537(0.0205)$	4,040
Neg. bin. full sample dependent	$C = 8.8453(0.899)$ $\rho = 0.5004(0.064)$ $p = 0.6035(0.023)$	3,955
Neg. bin. 1989 dependent	$C = 28.717(15.0)$ $\rho = 0.7675(0.48)$ $p = 0.8412(0.068)$	873 [880]
Neg. bin. 1990 dependent	$C = 14.341(3.47)$ $\rho = 0.3622(0.24)$ $p = 0.6747(0.056)$	971 [977]
Neg. bin. 1991 dependent	$C = 28.429(11.97)$ $\rho = 0.3632(0.25)$ $p = 0.7907(0.07)$	943 [947]
Neg. bin. 1992 dependent	$C = 13.488(2.91)$ $\rho = 0.2210(0.092)$ $p = 0.5805(0.047)$	1,056 [1,063]

^a Standard errors are given in parentheses.

^b Comparable estimates for independent annual samples are given in brackets.

extreme value approach gives an indication of the most appropriate class of distributions, maximum likelihood estimation can be exploited to capture certain clustering phenomena. The EM-algorithm turns out to be useful in detecting the optimum layer of distributions in cases where mixtures of distributions prevail.

Detecting the properties of the tails of a distribution is of particular importance for insurers that consider reinsurance. Reinsurance is a transfer of risk that can take several forms. This paper is particularly concerned with cases where this transfer is based on tail risk. Not only does it show the importance for catastrophic risk scenarios, but also does it show how neglecting the discrete blips may lead to severe underestimation of true risk.

Further simulations are required to get a better idea of the impact of temporal aggregation for mixtures of (truncated) distributions. Currently this issue has only received limited attention. For empirical evidence we would need complete datasets to check whether some assumptions made in this paper are actually valid. For example, in Fig. 1, there is some evidence that the discrete blips can also be found at levels below 25,000 guilders. How would that affect our EM-results; will the logarithmic series distribution still be an appropriate discrete candidate in that case? Likewise, we

would like to have information on the event aspect of extreme claims. Obviously, clustering aspects can take many forms. It is not always clear how to deal with them.

Acknowledgements

We are very grateful to the Centre for Actuarial Statistics (CVS, Zoetermeer), in particular to Age Willemse and Marjan van Kraaij for the data and comments. Special thanks are also due to Max King, Adrian Pagan, participants at the 7th World Congress of the Econometric Society and an anonymous referee for several suggestions and remarks.

References

- [1] Aiuppa, T.A. (1988). Evaluation of Pearson curves as an approximation of the maximum probable annual aggregate loss. *Journal of Risk and Insurance* 55, 425–439.
- [2] Barton, D.E., F.N. David and M. Merrington (1963). Table for the solution of the exponential equation $\exp(b) - b/(1 - p) = 1$, *Biometrika* 50, 169–176.
- [3] Beirlant, J. and J.L. Teugels (1992). Modeling large claims in non-life insurance, *Insurance: Mathematics and Economics* 11, 17–30.
- [4] Beirlant, J., J.L. Teugels and P. Vynckier (1994). Extremes in non-life insurance. In: J. Galambos et al., eds., *Extreme value theory and applications*. Kluwer Academic Publishers, Dordrecht.
- [5] Benktander, G. and C.O. Segerdahl (1960). On the analytical representation of claim distributions with special reference to excess of loss reinsurance. *Comptes Rendus du XVI Congrès International d'Actuaires* I, 626–648.
- [6] Boos, D.D. (1984). Using extreme value theory to estimate large percentiles. *Technometrics* 26, 33–39.
- [7] Cummins, J.D. and L.R. Freifelder (1978). A comparative analysis of alternative maximum probable yearly aggregate loss estimators. *Journal of Risk and Insurance* 45, 27–52.
- [8] Cummins, J.D., G. Dionne, J.B. McDonald and B.M. Pritchett (1990). Applications of the GB2 family of distributions in modelling insurance loss processes. *Insurance: Mathematics and Economics* 9, 257–272.
- [9] Dekkers, A.L.M., J.H.J. Einmahl and L. de Haan (1989). A moment estimator for the index and large quantile estimation. *The Annals of Statistics* 3, 743–756.
- [10] Dempster, A.P., N. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39, 1–38.
- [11] Galambos, J. (1987). The selection of the domain of attraction of an extreme value distribution from a set of data. Lecture notes in statistics. Springer Verlag, Berlin.

- [12] Galambos, J., J. Lechner and E. Simiu (1994). *Extreme value theory and applications*. Kluwer Academic Publishers, Dordrecht.
- [13] Hall, P. (1990). Asymptotic properties of the Bootstrap for heavy-tailed distributions. *Annals of Probability* 18, 1342–1360.
- [14] Hasover, A.M. and Z. Wang (1992). A test for extreme value domain of attraction. *Journal of the American Statistical Association* 87, 171–177.
- [15] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3, 1163–1173.
- [16] Hogg, R.V. and S.A. Klugman (1983). On the estimation of long tailed skewed distributions with actuarial applications. *Journal of Econometrics* 23, 92–102.
- [17] Koedijk, K.G., M.A. Schafgans and C.G. de Vries (1990). The tail index of exchange rate returns. *Journal of International Economics* 29, 93–108.
- [18] Loretan, M. and P.C.B. Phillips (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* 1, 211–248.
- [19] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). *Extremes and related properties of random sequences and processes*. Springer Verlag, New York.
- [20] Pentikäinen, T. (1987). Approximate evaluation of the distribution function of aggregate claims. *Astin Bulletin* 17, 15–39.
- [21] Ruud, P.A. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics* 49, 305–341.
- [22] Tanner, M. (1993). *Tools for statistical inference: methods for the exploration of posterior distribution and likelihood functions*, 2nd edition. Springer-Verlag, New York.
- [23] Weba, M. (1993). Fitting a parametric distribution for large claims in case of censored or partitioned data. *Insurance: Mathematics and Economics* 12, 155–165.
- [24] White, H. (1982). Regularity conditions for Cox's test of non-nested hypothesis. *Journal of Econometrics*, 301–318.
- [25] Willmot, G.E. (1989). The total claims distribution under inflationary conditions. *Scandinavian Actuarial Journal*, 1–12.
- [26] Zelterman, D. (1993). A semiparametric bootstrap technique for simulating extreme order statistics. *Journal of the American Statistical Association* 88, 477–485.