EXTREME VALUE THEORY AND LARGE FIRE LOSSES

G. RAMACHANDRAN
U.K.

INTRODUCTION

The statistical theory of extreme values well described by Gumbel [1] has been fruitfully applied in many fields, but only in recent times has it been suggested in connection with fire insurance problems. The idea originally stemmed from a consideration of the ECOMOR reinsurance treaty proposed by Thepaut [2]. Thereafter, a few papers appeared investigating the usefulness of the theory in the calculation of an excess of loss premium. Among these, Beard [3, 4], d'Hooge [5] and Jung [6] have made contributions which are worth studying. They have considered, however, only the largest claims during a succession of periods. In this paper, generalized techniques are presented which enable use to be made of all large losses that are available for analysis and not merely the largest. These methods would be particularly useful in situations where data are available only for large losses.

EXTREME VALUE DISTRIBUTIONS

Suppose there were n losses or claims during a given period. These observations constitute a sample of size n from a probability distribution $F(Z)$. If they are arranged in decreasing order of magnitude let $Z_{(m)n}$ be the mth loss with $Z_{(1)n}$ the largest. Over a succession of periods $Z_{(m)n}$ has a distribution with density function

$$\psi_m(y_m) = \frac{m^m}{(m-1)!} e^{-my_m - me^{-y_m}} dy_m \qquad (1)$$

if $F(Z)$ is of the "exponential type" [1]. This type has been chosen to explain the theory since it includes well known distributions like gamma, normal and log normal apart from the simple exponential function. The "reduced" mth largest value y_m is defined as

$$y_m = a_{mn}(Z_{(m)n} - b_{mn}) \qquad (2)$$

19

where a_{mn} and b_{mn} are the solutions of

$$F_n(b_{mn}) = 1 - \frac{m}{n} \tag{3}$$

and

$$a_{mn} = \frac{n}{m} f_n(b_{mn}) \tag{4}$$

If the precise structure of $F(Z)$ and hence the density function $f(Z)$ is known, the parameters a_{mn} and b_{mn} can be estimated from (3) and (4). However, if only the values of $Z_{(m)n}$ over a succession of periods are available, estimates could be obtained from

$$b_{mn} = \bar{Z}_{(m)n} - \frac{\bar{y}_m}{a_{mn}} \tag{5}$$

and

$$a_{mn} = \frac{\sigma_m}{\sigma_{mZ}} \tag{6}$$

where $\bar{Z}_{(m)n}$ and σ_{mZ} are the mean and standard error of $Z_{(m)n}$. Gumbel [1] has tabulated the mean $\bar{y}_m$ and standard error σ_m of y_m for $m = 1$ to 10; I have extended this table up to $m = 40$ [7]. These limiting or asymptotic values are true only in the case when the number of periods available for $Z_{(m)n}$ is large.

The parametric values of the parent distribution $F(Z)$ could be expected to change over a number of years. Hence it might be considered desirable to use only a small number of successive values of $Z_{(m)n}$. In this case, the following procedure is suggested. If the number of periods is N, the values y_{mj} $(j = 1 \ldots .. N)$ of y_m are given by

$$u_{mj} = me^{-y_{mj}} \tag{7}$$

where u_{mj} is obtained from the cumulative frequencies

$$\phi_m(u_{mj}) = \int_{u_{mj}}^{\infty} u^{m-1} e^{-u} \frac{du}{\Gamma(m)}$$

$$= \frac{j}{N+1} \tag{8}$$

with the aid of tables of incomplete gamma functions. (The variable u_m has a gamma distribution). The mean and standard error of y_m for the sample size N could then be calculated and used in (5) and (6) for estimating a_{mn} and b_{mn}.

If the N values of $Z_{(m)n}$ are arranged in increasing order of magnitude the jth observation $Z_{(m)n_j}$ in that arrangement would correspond to the reduced extreme y_{mj}. By fitting the straight line

$$Z_{(m)n_j} = b_{mn} + \frac{y_{mj}}{a_{mn}} \qquad (9)$$

either graphically or by the method of least squares estimates of the parameters could be obtained. There are other methods of estimation eg maximum likelihood, but these are beyond the scope of the present discussions. Practical results based on extreme value theory would involve errors due to the particular method adopted for estimating the parameters. Hence the relative efficiency of different methods is an important statistical problem which needs to be studied critically.

VARIATION IN SAMPLE SIZE

In the previous section it has been assumed that the extremes $Z_{(m)n}$ are from samples (periods) with constant sizes n. In the real world, however, the number of fires or claims would vary from period to period. Jung [6] suggested that, if the sample sizes differ but represent "equally exposed" intervals of time, the methods described for constant n are still applicable but with certain modification. Following Franckx [8], he introduced the "operational time" an estimate of which would be given by the average of the sample sizes. The samples might be deemed to have this average size.

I have considered a different approach as described in Appendix 1.

I have shown that,

$$Z_{(m)n_j} = b_{mn} + \frac{y_{mj} + \log_e (n_j/n)}{a_{mn}} \qquad (10)$$

approximately, where $Z_{(m)n_j}$ pertains to a sample of size n_j in the jth period. The sample size n could refer to the base period or the

average sample size. As in the case of constant n, the parameters may be estimated by fitting the linear relationship (10) or from the equations

$$b_{mn} = \bar{Z}_{(m)n_j} - \frac{\bar{y}_m + \bar{p}_j}{a_{mn}} \tag{11}$$

and

$$a_{mn}^2 = \frac{\sigma_m^2 + \sigma_p^2 + 2\sigma_{yp}}{\sigma_{mz}^2} \tag{12}$$

where $\bar{p}_j$ and σ_p^2 are the mean and variance of $p_j [= \log_e (n_j/n)]$. The covariance σ_{yp} of y_{mj} and p_j could be included in the calculation though it is theoretically equal to zero
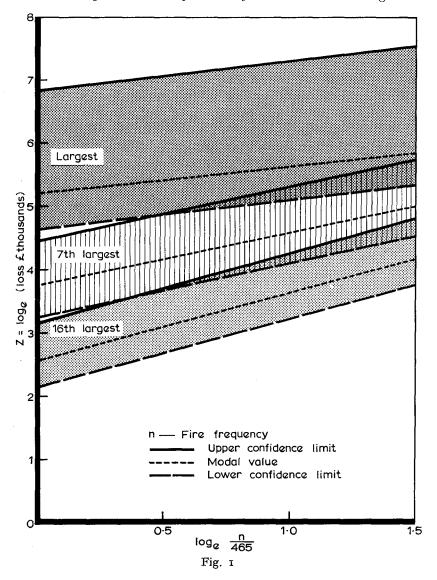
Expression (8) given in Appendix 1 is exactly true for a parent of simple exponential form. I have studied the errors in this approximation numerically for gamma and standard normal distributions for sample sizes from 450 to 1000 with $n = 450$. The errors were not serious for this range. However, it is known that for a normal distribution, the asymptotic form (1) does not furnish a satisfactory approximation unless n is extremely large.

LARGE LOSSES IN AN INDUSTRY

Pareto [9, 10] and logarithmic normal [11] are the forms usually suggested for the parent probability distribution $F(Z)$ of fire loss. The latter is slightly less "dangerous" than the former [12]. For physical reasons I would prefer a distribution of the "exponential type" for the logarithm of loss [13]. Of course, if the loss has a Pareto distribution its logarithm follows a simple exponential form.

With the assumption mentioned above, the theoretical results discussed in the previous sections were applied to large losses that occurred in the textile industry in the United Kingdom. The data related to the top 17 losses in the industry during the 21-year period from 1947 to 1967. With the help of the index numbers for retail prices the observed losses were corrected for inflation and the logarithms of corrected losses used in the calculations. The straight line (10) was fitted in order to take into consideration the increasing number of fires during the period. The results are reproduced in

Cols 2 and 3 of the table in Appendix 2. (Fire Research Note No. 910 [7] contains details of the analytical steps). The parametric values pertain to a sample size of $n_1 = 465$ fires in the base year 1947; these were estimated from the extreme observations only under the broad assumption that the probability distribution of the logarithm

Fig. 1

of loss belongs to the exponential family. This assumption appears to be justified since in all the 17 cases a high correlation was obtained between the observations Z and the theoretical values y.

Equation (8) in Appendix 1 measures the effect of an increase in the fire frequency n on the modal large loss b_{mn_j}. This line has been shown in Fig. 1 for the largest, 7th largest and 16th largest losses in the textile industry. Also shown in this figure are the upper and lower confidence lines for the extremes obtained by inserting in (10) the corresponding limits for the reduced extremes. The limits for y_m were obtained directly from their distributions shown in (1). The probability of exceeding the upper line or falling short of the lower one for a given (n_j/n_1) is 0.025.

The confidence lines represent a control chart based on the current trend. The increase in the frequency n_j of fires may be partly due to the inadequacy of fire prevention measures. In addition, if some or all of the actual large losses corrected for inflation exceeded the corresponding upper limits it may be concluded that general changes in fire-fighting and fire protection methods, or in the industrial processes are taking place to alter the picture for worse. If the losses are less than the lower limits, then the changes are for the better.

As in the case of human life, fires have a high rate of "infant mortality". In 1967, for example, out of a total number of 982 fires attended by fire brigades in buildings concerned with textile manufacture 524 fires were confined to exterior components, appliances and common service spaces [14]. If these fires are disregarded the "duration of burning" t of the remaining fires would have a probability distribution with "increasing failure rate" [15]. Hence Z (log loss), which is proportional to t as a first approximation, can be assumed to have a failure rate $h(u)$ increasing exponentially so that

$$h(u) = e^{\alpha + \beta u} \tag{13}$$

For fires "fought" expression (13) would be expected to be true except for small values of Z.

The extreme value parameter a_{mn} as defined in (4) denotes the value of the failure rate function at the characteristic large value b_{mn_j} the jth year i.e.

$$a_{mn} = e^{\alpha + \beta b_{mn_j}} \tag{14}$$

Using the 17 pairs of values for the two parameters, values of α and β were estimated as -4.0825 and 0.3839 respectively for the year 1967 ($n_j = 982$). Since the density function of Z is

$$f(Z) = h(Z)e^{-\int_0^z h(u)du}\,dz \qquad (15)$$

the transformed variable

$$\xi = \frac{e^{\alpha+\beta z}}{\beta} \qquad (16)$$

has the density function

$$f(\xi) = Ke^{-\xi}\,d\xi \qquad (17)$$

true for the domain $\log K \leqslant \xi \leqslant \infty$. In view of the fact that Z is the logarithm of loss x

$$x = (\beta\xi\, e^{-\alpha})^{1/\beta} \qquad (18)$$

The expected value of x for a desired range of loss could be calculated from (17) and (18) using tables of incomplete gamma functions. In the example considered the average loss in the textile industry in 1967 was estimated as £2200 in the range £55 to £10,000 (at 1967 values). At present data are not available for individual losses below £10,000. As illustrated in the example, by analysing the extreme values, it is possible to estimate the total loss in an industry or in a group of buildings with or without fire protection measures like sprinklers. Individual totals for different groups or geographical areas are required for economic studies. The British Insurance Association publish only national totals for each month and details of their method of estimation are not known.

PARAMETERS OF THE PARENT DISTRIBUTION

Suppose the parent distribution $F(Z)$ has a location parameter μ and a scale parameter σ. It is emphasized that μ is not necessarily the expectation nor σ the standard deviation. The problem is to estimate μ and σ from samples of large values from $F(Z)$.

Consider the standardized variable

$$t = \frac{Z-\mu}{\sigma} \qquad (19)$$

If n observations of t in a sample are arranged in decreasing order the mth observation in that arrangement is

$$t_{(m)n} = \frac{Z_{(m)n} - \mu}{\sigma} \qquad (20)$$

If the form of $F(Z)$ is known the distribution $G(t)$ of t is also known, so that the expectation, variance and covariance of $t_{(m)n}$ could be calculated by the method described by Ogawa [16]. The moments of order statistics in small samples from well known distributions have already been discussed and tabulated [17].

For large n the precise calculations of moments of order statistics by an exact approach are time-consuming and impracticable. In this case, the following approximate method based on extreme value theory may be adopted. If $A_{(m)n}$ and $B_{(m)n}$ are the extreme value parameters of $t_{(m)n}$ then the reduced variable

$$y_m = A_{(m)n}(t_{(m)n} - B_{(m)n}) \qquad (21)$$

also has the density function shown in (1) if $F(Z)$ and hence $G(t)$ is of the exponential type. Since $G(t)$ is a known distribution, for a given n, the parameters can be calculated from the equations

$$G_n(B_{(m)n}) = 1 - (m/n) \qquad (22)$$

and

$$A_{(m)n} = (n/m)\, g_n\, (B_{(m)n}) \qquad (23)$$

where $g(t)\, [= G'(t)]$ is the density function of t. From (21), the expected value and variance of $t_{(m)n}$ are

$$E(t_{(m)n}) = \bar{t}_{(m)n}$$

$$= B_{(m)n} + \frac{\bar{y}_m}{A_{(m)n}} \qquad (24)$$

and

$$V(t_{(m)n}) = \frac{\sigma_m^2}{A_{(m)n}^2} = \sigma_{mt}^2 \qquad (25)$$

As mentioned earlier $\bar{y}_m$ and σ_m^2 are the expected value and variance of y_m already tabulated for different values of m. I have proved that the covariance of extremes y_m and y_l where $m > l$ is the same as the variance of y_m [7].

Hence

$$Cov\left[t_{(m)n}, t_{(l)n}\right] = \frac{\sigma_m^2}{A_{(m)n} \cdot A_{(l)n}}$$

$$= \sigma_{mlt}^2 \qquad (26)$$

Now, from (2), (20) and (21)

$$\sigma_{mt}^2 = \frac{\sigma_{mZ}^2}{\sigma^2}$$

$$= \frac{\sigma_m^2}{a_{mn}^2 \sigma^2} \qquad (27)$$

Therefore from (25) and (27)

$$A_{(m)n} = \sigma a_{mn} \qquad (28)$$

Similarly

$$t_{(m)n} = \frac{\bar{Z}_{(m)n} - \mu}{\sigma}$$

$$= \frac{b_{mn} + (\bar{y}_m/a_{mn}) - \mu}{\sigma} \qquad (29)$$

so that from (24), (28) and (29)

$$B_{(m)n} = \frac{b_{mn} - \mu}{\sigma} \qquad (30)$$

In (28) and (30), $A_{(m)n}$ and $B_{(m)n}$ are known quantities while a_{mn} and b_{mn} can be estimated from the observations $Z_{(m)n_j}$ by one of the methods already discussed. Thus these two equations yield estimates of μ and σ. This method is slightly different from the one suggested by Jung [6] in the case of the largest value, i.e. $m = 1$.

However, as pointed out by Jung, it is difficult to draw reliable conclusions from estimates based on just one extreme, viz the mth. It is possible to overcome this difficulty by considering a number of extremes, say, $m = 1$ to r provided of course such data are available for analysis. In this case we may proceed as follows:

from (20) and (21)

$$\bar{Z}_{(m)n} = \mu + \sigma \bar{t}_{(m)n} \qquad (31)$$

$$\sigma_{mZ}^2 = \sigma^2 . \sigma_{mt}^2 \qquad (32)$$

where $\bar{t}_{(m)n}$ and σ^2_{mt} are given by (24) and (25). Also, from (26),

$$Cov\,[Z_{(m)n},\,Z_{(l)n}] = \sigma^2_. \cdot \sigma^2_{mlt} \qquad (33)$$

with $m > l$ and σ^2_{mlt} given by (26). Then following Lloyd [18] we could obtain least squares estimates of μ and σ by minimizing the quadratic (matrix)form

$$(Z - C\Theta)'\,V^{-1}\,(Z - C\Theta) \qquad (34)$$

where

$$Z = \begin{vmatrix} \tilde{Z}_{(1)n} \\ \tilde{Z}_{(2)n} \\ \text{---------} \\ \text{---------} \\ \tilde{Z}_{(r)n} \end{vmatrix},\; C = \begin{vmatrix} 1 & \bar{t}_{(1)n} \\ 1 & \bar{t}_{(2)n} \\ \text{---------} \\ \text{---------} \\ 1 & \bar{t}_{(r)n} \end{vmatrix},\; \Theta = \begin{vmatrix} \mu \\ \sigma \end{vmatrix}$$

and V is of the form

$$V = \begin{vmatrix} \dfrac{\sigma^2_1}{A^2_{(1)n}} & \dfrac{\sigma^2_2}{A_{(1)n} \cdot A_{(2)n}} & \dfrac{\sigma^2_3}{A_{(1)n} \cdot A_{(3)n}} & \dfrac{\sigma^2_r}{A_{(1)n} \cdot A_{(r)n}} \\[2.5ex] \dfrac{\sigma^2_2}{A_{(2)n} \cdot A_{(1)n}} & \dfrac{\sigma^2_2}{A^2_{(2)n}} & \dfrac{\sigma^2_3}{A_{(2)n} \cdot A_{(3)n}} & \dfrac{\sigma^2_r}{A_{(2)n} \cdot A_{(r)n}} \\[2.5ex] \dfrac{\sigma^2_3}{A_{(3)n} \cdot A_{(1)n}} & \dfrac{\sigma^2_3}{A_{(3)n} \cdot A_{(2)n}} & \dfrac{\sigma^2_3}{A^2_{(3)n}} & \dfrac{\sigma^2_r}{A_{(3)n} \cdot A_{(r)n}} \\[2.5ex] \text{---------} & & & \\[2.5ex] \dfrac{\sigma^2_r}{A_{(r)n} \cdot A_{(1)n}} & \dfrac{\sigma^2_r}{A_{(r)n} \cdot A_{(2)n}} & \dfrac{\sigma^2_r}{A_{(r)n} \cdot A_{(3)n}} & \dfrac{\sigma^2_r}{A^2_{(r)n}} \end{vmatrix}$$

It is not necessary here to reproduce the equations giving the estimates of μ and σ and discuss other connected statistical problems.

EXCESS OF LOSS REINSURANCE

Suppose that the claims in a given category of risk could be regarded as independent random variables with a distribution function $V(x)$. The net premium per claim for an excess loss cover

above an amount L is, in accordance with current insurance practice

$$P(L) = \int_L^\infty (x - L)\, v(x)\, dx$$

$$= \int_L^\infty xv(x)\, dx - L \int_L^\infty v(x)\, dx \qquad (35)$$

where $v(x) \left[= \dfrac{d}{dx}\, V(x) \right]$ is the density function. This expression can also be written as

$$P(L) = \int_L^\omega [1 - V(x)]\, dx \qquad (36)$$

Suppose the transformed variable $Z(= \log_e x)$ has a distribution function $F(Z)$ and density function $f(Z)$ of the exponential type. It may also be assumed that Z has a location parameter μ and a scale parameter σ. If the standardized variable t shown in (19) has a known distribution function $G(t)$ and density function $g(t)$, the values of μ and σ could be estimated from extreme observations by the generalised method described in the previous section. Since

$$V(x) = F(\log x) = G \left[\frac{\log x - \mu}{\sigma} \right] \qquad (37)$$

and

$$dx = e^Z\, dZ = e^{\mu + t\sigma}\, \sigma dt \qquad (38)$$

it is easy to verify that

$$P(L) = e^\mu \sigma p \left(\frac{\log L - \mu}{\sigma} \right) \qquad (39)$$

where the function

$$p(k) = \int_k^\infty [1 - G(t)]\, e^{t\sigma}\, dt \qquad (40)$$

Expressions (39) and (40) are similar to the results due to Jung [6] but applicable to the specific case in which the logarithm of the claim amount x, rather than x, has an exponential type distribution.

For example, consider the Pareto distribution

$$V(x) = 1 - x^{-\lambda}\ (1 \leqslant x \leqslant \infty) \qquad (41)$$

In this case

$$F(Z) = 1 - e^{-\lambda z}$$
$$G(t) = 1 - e^{-t} \tag{42}$$

with $\mu = 0$ and $\sigma = (1/\lambda)$. From (40) and (42)

$$p(k) = \frac{\lambda}{\lambda - 1} \, e^{-(\lambda - 1/\lambda)k}$$

so that for $k = \dfrac{\log L}{\sigma}$

$$P(L) = \frac{L^{1-\lambda}}{\lambda - 1} \tag{43}$$

in agreement with the result obtained directly from (36) and (41).

In the case of a log normal distribution

$$1 - G(t) = \Omega_G(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-u^2/2} \, du \tag{44}$$

Hence

$$p(k) = \int_k^\infty \Omega_G(t) \, e^{t\sigma} \, dt$$

$$= \left[\frac{\Omega_G(t) \, e^{t\sigma}}{\sigma} \right]_k^\infty + \frac{1}{\sigma} \int_k^\infty g(t) \, e^{t\sigma}$$

where

$$g(t) = -\frac{d}{dt} \, \Omega_G(t) = \frac{1}{\sqrt{2\pi}} \, e^{-t^2/2}$$

After simplification

$$p(k) = (1/\sigma) \left[e^{\sigma^2/2} \, \Omega_G(k - \sigma) - e^{k\sigma} \Omega_G(k) \right] \tag{45}$$

For a given $k = \dfrac{(\log L - \mu)}{\sigma}$ the values of $\Omega_G(k - \sigma)$ and $\Omega_G(k)$ could be obtained from tables of the (standard) normal probability function. (It must be kept in mind that μ and σ are the mean and standard deviation of $Z = \log x$).

In order to apply the method mentioned above, $G(t)$ has to be a known distribution. If the structure of $G(t)$ is not known but can be assumed to be of the exponential type the following procedure is feasible. Consider r claims x_i ($i = 1$ to r) above L in decreasing order starting with the largest. Let $x_{(m)}$ be the mth claim in that arrangement with $m = 1$ as the largest. It is assumed that $Z = \log x$ has an exponential type parent distribution $F(Z)$. Since the $Z_{(m)} = \log x_{(m)}$ has the probability density function shown in (1) the expected value of $x_{(m)}$ is given by

$$Ex_{(m)} = \frac{m^m}{(m-1)!} \int_{-\infty}^{\infty} e^{-my_m - me^{-y_m}} e^{b_m + (y_m/a_m)} \qquad (46)$$

By integrating (46) it may be seen that

$$\bar{x}_m = Ex_{(m)} = \frac{e^{b_m} m^{\Theta_m} \Gamma(m - \Theta_m)}{\Gamma(m)} \qquad (47)$$

where $\Theta_m = (1/a_m)$. As discussed earlier estimates of a_m and b_m could be obtained from observations on $x_{(m)}$ during successive periods.

The aggregate net premium above L is

$$S(L) = \sum_m x_{(m)} - rL \qquad (48)$$

Over a number of periods $S(L)$ has a probability distribution.

The expected value of $S(L)$ is given by

$$\bar{S}(L) = \sum_m Ex_{(m)} - rL$$

$$= \sum_m \bar{x}_{(m)} - rL \qquad (49)$$

where $\bar{x}_{(m)}$ is given by (47). It is hoped to study the higher moments of $S(L)$ later.

DISCUSSION AND CONCLUSIONS

Large losses play a vital role in the economics of fire protection measures which are designed to prevent a fire from becoming large. Similarly large clams exercise a critical effect on the performance

of an insurer whose top risks are not cut away by reinsurance. A reinsurer on the other hand is worried about the fluctuations in the portfolio of large risks he accepts. Hence large losses or claims merit special investigation.

Large losses fall on the extreme (upper) tail of the probability distribution of fire loss. The behaviour of this tail is quite distinct from that of the remaining major portion of the probability curve. The nature of the tail could be studied by applying statistical techniques concerned with truncated distributions. But this is possible only if loss figures are available for all fires. In many practical situations this is not the case. At the national level, for example, figures for individual losses are available only for fires costing £ 10,000 or more in direct damage. In these circumstances the asymptotic theory of extreme values can prove useful. In this paper generalized techniques have been presented which enable the maximum use to be made of extreme observations for practical purposes.

From physical considerations I conclude that the logarithm of loss has a probability distribution belonging to the exponential family. This family includes well known distributions like log normal, log (simple) exponential, ie Pareto etc. It also includes the distribution with increasing failure rate for log loss as described in (13); this assumption would lead to the following distribution for x ie loss

$$f(x) = K\lambda\beta x^{\beta-1} e^{-\lambda x^\beta} \tag{50}$$

where

$$\lambda = \frac{e^\alpha}{\beta} \tag{51}$$

It may be recognised that (50) is one of the forms of Weibull density. The assumption of an exponential type parent would enable one to use the generalized form of Gumbel's Type 1 asymptotic distribution for the mth largest order statistics.

For using the asymptotic form the number of fires n in a category of risk has to be large. For small n the general theory of order statistics [17] would be applicable, but this would require a knowledge of the exact nature of the parent distribution. This is beyond the scope of this paper.

I will conclude by agreeing with Jung [6] that the use of extreme values has to be approached with caution. But extreme value techniques are useful in situations where only large losses are available for analysis. Of course better solutions would be possible from data on all fires or claims.

ACKNOWLEDGMENT

This paper is Crown Copyright, reproduced by permission of the Controller of Her Majesty's Stationery Office. It is contributed by permission of the Director, Building Research Establishment. (The Fire Research Station is the Joint Fire Research Organization of the Department of the Environment and Fire Offices' Committee).

REFERENCES

[1] GUMBEL, E. J. (1958). Statistics of extremes. Columbia University Press, New York.
[2] THEPAUT, A. (1950). Le traite d'excédent du coût moyen relatif (ECOMOR). Bull. trimest. Inst. Actu. Fr., *49*, 273-343.
[3] BEARD, R. E. (1955). Some statistical aspects of non-life insurance. J. Inst. Actu. Students' Soc., *13*, Part 3, 139-57.
[4] BEARD, R. E. (1963). Some notes on the statistical theory of extreme values. Astin Bull., Vol. III, Pt. I, 6-12.
[5] D'HOOGE, L. (1965). Theorie des valeurs extrêmes et la tarification de "P" excess of loss. Astin Bull., Vol. III, Pt. II, 163-177.
[6] JUNG, J. (1965). On the use of extreme values to estimate the premium for excess of loss reinsurance. Astin Bull., Vol. III, Pt. II, 178-184.
[7] RAMACHANDRAN, G. (1972). Extreme value theory and fire losses—further results. Department of the Environment and Fire Offices' Committee Joint Fire Research Organisation Fire Research Note No. 910.
[8] FRANCKX, E. (1963). Sur la fonction de distribution du sinistre le plus élevé. Astin Bull., Vol. II, Pt. III, 415.
[9] BENKERT, L. G. and STERNBERG, I. (1957). An attempt to find an expression for the distribution of fire damage amount. Trans. 15th International Congress of Actuaries, *2*, 288-294.
[10] MANDELBROT, B. (1964). Random walks, fire damage amount and other Paretian risk phenomena. Ops Res., *12*, 582-585.
[11] BENKERT, L. G. (1963). The log normal model for the distribution of one claim. Astin Bull., Vol. II, Pt. I, 9-23.
[12] BENKTANDER, G. (1963). A note on the most 'dangerous' and skewest class of distributions. Astin Bull., Vol. II, Pt. III, 387.
[13] RAMACHANDRAN, G. (1970). Some possible applications of the theory of extreme values for the analysis of fire loss data. Ministry of Technology and Fire Offices' Committee Joint Fire Research Organisation Fire Research Note 837.
[14] United Kingdom fire statistics 1967. London 1969. Her Majesty's Stationery Office.

[15] RAMACHANDRAN, G. (1969). The Poisson process and fire loss distribution. Thirty-seventh session of the International Statistical Institute, London.
[16] OGAWA, J. (1951). Contributions to the theory of systematic statistics, 1. Osaka Math. J., 3, 175-213.
[17] SARHAN, A. E. and GREENBERG, B. G. (Eds) (1962). Contributions to Order Statistics. John Wiley, New York.
[18] LLOYD, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics. Biometrika, 39, 88-95.

APPENDIX I

By definition

$$F_n(b_{mn}) = 1 - (m/n) \tag{52}$$

and

$$F(Z) = 1 - e^{-\int_0^z h(u)\,du} \tag{53}$$

where $h(u)$ is the failure rate function.

Hence

$$H(b_{mn}) = \int_0^{b_{mn}} h(u)\,du$$
$$= \log (n/m) \tag{54}$$

For exponential type parents the critical quotient $Q(Z)$ given by

$$Q(Z) = \frac{h(Z)}{-f'(Z)/f(Z)} \tag{55}$$

tends to unity for large Z [1]. For large Z the density of probability $f(Z)$ becomes very small and the same holds for the probability $\{1 - F(Z)\}$ of a value exceeding Z. If the variate is unlimited the derivative $f'(Z)$ also converges to zero. From (55) we may write

$$h(Z) = \frac{f(Z)}{1 - F(Z)} \sim \frac{-f'(Z)}{f(Z)} \tag{56}$$

By taking derivatives of $h(Z)$ we may extend (56) to write that, for large Z

$$h(Z) \sim \frac{-f'(Z)}{f(Z)} \sim \frac{-f''(Z)}{f'(Z)} \sim \frac{-f'''(Z)}{f''(Z)} \tag{57}$$

From (56), using (57) it can be shown that $h'(Z)$ tends to zero for large Z. By taking further derivatives it is easy to verify that all the derivatives of $h(Z)$ tend to zero asymptotically. In fact this property is implied in the derivation of the asymptotic distribution of extreme values for exponential type parents.

Let b_{mn_j} be the characteristic mth large value in samples of size n_j from $F(Z)$. In the neighbourhood of b_{mn}

$$H(b_{mn_j}) - H(b_{mn})$$

$$= (b_{mn_j} - b_{mn})\, H'(b_{mn}) + \frac{(b_{mn_j} - b_{mn})^2}{2}\, H''(b_{mn}) + \text{-----}$$

$$= (b_{mn_j} - b_{mn})\, h(b_{mn}) + \frac{(b_{mn_j} - b_{mn})^2}{2}\, h'(b_{mn}) + \text{-----}$$

$$= (b_{mn_j} - b_{mn})\, h(b_{mn}) \tag{58}$$

Since the derivatives of $h(b_{mn})$ vanish for extreme values. From (54), the left-hand side of (58) is equal to $\log \left(\dfrac{n_j}{n} \right)$.

Hence

$$b_{mn_j} = b_{mn} + \frac{1}{a_{mn}} \log \frac{n_j}{n} \tag{59}$$

approximately since $h(b_{mn}) = a_{mn}$.

We have

$$Z_{mn_j} = b_{mn_j} + \frac{y_{mn_j}}{a_{mn_j}} \tag{60}$$

where y is the reduced variable which is independent of the sample size. From (59) and (60)

$$Z_{mn_j} = b_{mn} + \frac{y_{mn_j} + \log \left(\dfrac{n_j}{n} \right)}{a_{mn}} \tag{61}$$

Since a_{mn_j} is equal to the constant value a_{mn} for values of b_{mn_j} in the vicinity of b_{mn}.

20

APPENDIX 2

Results for the textile industry, *UK*

Extremes (m) (1)	a_{mn_1} (2)	b_{mn_1} (3)
1	2.247	5.214
2	1.785	4.829
3	1.626	4.534
4	1.460	4.327
5	1.387	4.113
6	1.424	3.988
7	1.239	3.749
8	1.163	3.564
9	1.212	3.448
10	1.034	3.259
11	0.973	3.137
12	0.925	2.972
13	0.886	2.832
14	0.924	2.749
15	0.937	2.680
16	0.950	2.583
17	1.002	2.537

The parameter b_{mn_1} is based on the logarithms of losses in units of £ thousands at 1947 values. $[n_1 = 465]$.