Fire Research Note No 991

FACTORS AFFECTING FIRE LOSS - MULTIPLE REGRESSION MODEL WITH EXTREME VALUES

by

G RAMACHANDRAN

October 1973

FIRE RESEARCH STATION

50217

Fire Research Station
BOREHAMWOOD
Hertfordshire
WD6 2BL

Tel 01.953.6177

F.R. NOTE No. 991
October, 1973.

FACTORS AFFECTING FIRE LOSS - MULTIPLE REGRESSION MODEL WITH EXTREME VALUES

by

G. Ramachandran

SUMMARY

In this paper, a multiple regression model with extreme observations is developed and illustrated with an example. The classical model needs to be modified to take into consideration the biases due to the use of large values rather than values covering the entire range of the fire loss variable. The presence or absence of sprinklers and the height of the building, i.e. single storey or multi-storey, are the two qualitative factors studied in this note. The total floor area is the third independent variable included in the analysis which is of a quantitative character. Judged from extreme losses, sprinklers appear to reduce the expected damage considerably.

The model uses extreme observations individually; regression parameters are estimated from two sets of extremes, viz. the largest and the second largest and their replicated values over six years. The parameters have different values depending upon the rank of the extremes. In a later study it is hoped to estimate a single (constant) value for each regression parameter by carrying out a more complicated analysis combining the information on all large losses.

KEY WORDS: Large fires, loss, factors, multiple regression.

DEPARTMENT OF THE ENVIRONMENT AND FIRE OFFICES' COMMITTEE
JOINT FIRE RESEARCH ORGANIZATION

LIST OF SYMBOLS

x	Financial loss (in units of £'000)
z	Logarithm of x (dependent variable)
v_1	Independent variable; takes the value + 1 if the building is sprinklered, or - 1 if the building is not sprinklered
v_2	Independent variable; takes the value + 1 if the building is multi-storeyed, or - 1 if the building is single storeyed
x_3	Total floor area in units of 100 sq ft
v_3	Logarithm of total floor area (in units of 100 sq ft); independent variable
μ	Expected value (average) of z over its entire range; an average for all sizes of buildings
σ	Standard error of z over its entire range
μ_v	Expected value of z for a given set of values for v_1, v_2 and v_3
σ_v	Standard error of z for a given set of values for v_1, v_2 and v_3 (residual)
t	The standardised loss
$z_{(m)}$	The mth largest observed loss from top ($m = 1$ is the largest)
$t_{(m)}$	The standardised m^{th} largest loss from top
$B_{(m)}$	'Characteristic' (modal) value of $t_{(m)}$
$A_{(m)}$	Value of the 'intensity function' of the parent distribution at B_m
$G(t)$	(Cumulative) distribution function of t
$g(t)$	Density function of t
n	Sample size; number of fires per year (excluding small ones)
y_m	The reduced m^{th} extreme
$\bar{y}_m$	The expected value (average) of y_m
σ_m	The standard error of y_m
$\bar{z}_m$	The expected value of $z_{(m)}$
σ_{mz}	The standard error of $z_{(m)}$
β_0	The constant term in the regression model
β_1	The regression parameter pertaining to sprinklers
β_2	The regression parameter pertaining to storeys
β_3	The regression parameter pertaining to the total floor area

LIST OF SYMBOLS (cont'd)

β_{om}, β_{1m}, β_{2m} and β_{3m} are the same as β_o, β_1, β_2 and β_3 respectively but pertain to the regression with the mth extreme observations.

β'_{om}	The constant term in the actual regression with the mth extreme observations
e_{mjk}	Residual error in the regression
$z_{(m)jk}$	The mth largest loss in the jth year ($j = 1....6$) for the kth sub population ($k = 1....4$)

v_{1mjk}, v_{2mjk} and v_{3mjk} are the same as v_1, v_2 and v_3 respectively but pertain to the mth largest loss in the jth year for the kth sub population.

μ'_{vm}	Expected value of z for a given set of values for v_1, v_2 and v_3 as estimated by the regression with the mth extreme observations
μ'_{vmjk}	Value of μ'_{vm} estimated by substituting v_{1mjk}, v_{2mjk} and v_{3mjk} in the regression equation
R_{mw}	The weighted residual standard error in the regression with the mth extremes
μ_{vm}	Estimate of μ_v based on the mth extremes
σ_{vm}	Estimate of σ_v based on the mth extremes
μ_x	The expected value of x
α_m	A composite constant term for the mth regression
E	The process of taking the expected value
Var	The process of taking the variance, viz. the square of the standard deviation

FACTORS AFFECTING FIRE LOSS - MULTIPLE REGRESSION MODEL WITH EXTREME VALUES

by

G. Ramachandran

1. INTRODUCTION

1.1 The problem

In previous papers[1,2] the author discussed the application of the statistical theory of extreme values for analysing large fire losses. In this note the problem of multiple regression with extreme values is investigated with a view to assessing the relative contributions of various factors to fire losses. In this preliminary study only three factors are included, but it is hoped to extend these techniques to perform a combined analysis and also to include more factors.

1.2 Data

Data on large losses and the hazards in which these fires occurred are available for a number of years. However, information on fire protection devices and other particulars of buildings involved in large fires is available only for 1965 and later years. For this reason the methods developed are applied to large losses during the period 1965 to 1970. The textile industry has been chosen for purposes of illustration. All the values and conclusions given in this paper refer only to this industry.

2. FACTORS

2.1 Qualitative factors

Certain factors associated with the building are qualitative in character, eg. the presence or absence of sprinklers. Sprinklered buildings may be expected to differ from those without this protection in regard to the extent of spread of fire. In the same way, single storey and multi-storey buildings are deemed to be two distinct populations. The two factors mentioned above subdivide the major industrial group into four categories of buildings. For an application of the extreme value theory the number of fires in a year in any category should be large and this requirement restricts the number of categories.

The top two losses in each of the four sub groups of the textile industry during 1965 to 1970 were corrected for inflation with 1965 as the base year. The corrected figures are given in Tables 1 and 2 in Appendix 1 together with their logarithms (to base 10). In the case of sprinklered buildings the figures refer to fires in which sprinklers operated. The probability of non operation will be taken into account in a subsequent study of costs and benefits of sprinklers.

The presence or absence of sprinklers would be denoted by the variable v_1. If the building is sprinklered, v_1 has been assigned the value +1 and the value −1 if the building was not provided with sprinklers. Similarly the value +1 has been assigned to the variable v_2 if the building was multi-storeyed and the value −1 if it was single storeyed. In Appendix 1 the values of v_1 and v_2 are also shown for the four sub groups. The interaction between the two factors is not included in this study.

2.2 Quantitative factor

It has been established that the fire loss depends upon the size of the building or value at risk[3,4]. The loss figures need adjustment taking into consideration the differences in the sizes of buildings. Previous studies indicate that fire loss has a power relationship with the size of the building[3,4]. Hence the logarithm of loss, viz. z has a linear relationship with the variable v_3, the logarithm of the total floor area of the building. The variable v_3 is quantitative in character. The values of v_3 are also shown in the tables in Appendix 1.

3. NUMBER OF FIRES

3.1 Sprinklered and non-sprinklered buildings

Every year fire brigades in the United Kingdom attend about 1100 fires in buildings engaged in the textile trade[5]. About 45 per cent of these fires are in sprinklered buildings. But, according to a survey conducted by the Station some years ago, about one third of fires in sprinklered buildings are neither attended by fire brigades nor reported to the Organisation. Hence about 750 fires occur in sprinklered buildings in the textile industry against 600 fires in buildings without sprinklers.

3.2 Single storey and multi-storey buildings

According to a survey conducted by Building Research Station[6], about 43 per cent of industrial buildings are single storeyed. Also if the size of a building is doubled the frequency of fires could be expected to increase by a factor of $\sqrt{2}$ [3]. If these factors are taken into consideration the number of fires in multi-storeyed buildings would be about twice the number in single storey buildings. Hence the estimated figures for the number of fires in a year are those shown in column 2 of the table in Appendix 2.

4. EXTREME VALUE PARAMETERS

4.1 The mth extreme of the standardised variable

In a previous note it has been shown that if x is the fire loss the transformed variable $z = \log x$ follows a probability distribution of the 'exponential type'[1]. It may be assumed specifically that z has a log normal distribution with mean μ and standard deviation σ. Consider now the standardised variable

$$t = \frac{z - \mu}{\sigma} \qquad \ldots\ldots (1)$$

which has a normal distribution with zero mean and unit standard deviation. The fire losses in a particular year constitute a sample and if they are arranged in decreasing order of magnitude the mth largest value of t from top is given by

$$t_{(m)} = \frac{z_{(m)} - \mu}{\sigma} \qquad \ldots\ldots (2)$$

where $z_{(m)}$ is the logarithm of the mth largest loss. The probability density of $t_{(m)}$ is[2]

$$\chi_m(t_{(m)}) = \frac{m^m}{(m-1)!} A_m e^{-my_m - me^{-y_m}} dt_{(m)} \qquad \ldots\ldots (3)$$

where

$$y_m = A_m (t_{(m)} - B_m) \qquad \ldots\ldots (4)$$

The parameters A_m and B_m are solutions of

$$G_n(B_m) = 1 - \frac{m}{n} \quad \text{and} \qquad \ldots\ldots (5)$$

$$A_m = \frac{n}{m} g_n(B_m) \qquad \ldots\ldots (6)$$

where $G(t)$ and $g(t)$ are the (cumulative) distribution function and density function of the standard normal variable t and n denotes the sample size, i.e. number of fires in a year.

4.2 Values of the parameters

It may be assumed that during a short period of six years there was no appreciable increase in the number of fires and hence an average value of n can be used in the analysis. About 50 per cent of the fires were small ones which did not spread beyond the appliance of origin[5]. Disregarding these as cases of 'infant mortality' the large losses have been deemed to come from samples of sizes n shown in the third column of the table in Appendix 2. The values of A_m and B_m for $m = 1$ and 2 are also shown in this table; these values were obtained from (5) and (6) using tables of the normal probability integral[7].

4.3 Variance and expected value

From (2) and (4)

$$\begin{aligned} \text{var}(t_{(m)}) &= \text{var}(z_{(m)}) / \sigma^2 \\ &= \sigma^2_{mz} / \sigma^2 \\ &= \sigma_m^2 / A_m^2 \qquad \ldots\ldots (7) \end{aligned}$$

where σ^2_{mz} is the variance of $z_{(m)}$ and σ^2_m the variance of the reduced variable y_m. Hence an estimate of the standard deviation σ of the parent distribution is given by

$$\sigma = \frac{\sigma_{mz} A_m}{\sigma_m} \qquad \ldots\ldots (8)$$

Also the expected value is

$$E(t_{(m)}) = \frac{E(z_{(m)}) - \mu}{\sigma}$$

$$= \frac{\bar{z}_m - \mu}{\sigma}$$

$$= B_m + \frac{\bar{y}_m}{A_m} \qquad \ldots\ldots (9)$$

where $\bar{z}_m$ and $\bar{y}_m$ are the expected mean values of z_m and y_m respectively. From (9)

$$\mu = \bar{z}_m - \sigma\left[B_m + \frac{\bar{y}_m}{A_m}\right] \qquad \ldots\ldots (10)$$

5. REGRESSION MODEL

5.1 The problem

For a given v_1, v_2 and v_3 the dependent variable z has an expected value μ_v and (residual) standard error σ_v. Also,

$$\mu_v = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3 \qquad \ldots\ldots (11)$$

The location parameter μ and scale parameter σ mentioned in the previous section take the values μ_v and σ_v in the regression model. The problem is to estimate the regression parameters β_0, β_1, β_2, β_3 and σ_v using the mth large values $z_{(m)}$ and the associated values of v_1, v_2 and v_3.

5.2 The regression parameters

Consider the regression model

$$z_{(m)jk} = \beta'_{0m} + \beta_{1m} v_{1mjk} + \beta_{2m} v_{2mjk} + \beta_{3m} v_{3mjk} + e_{mjk} \qquad \ldots\ldots (12)$$

where $z_{(m)jk}$ is the mth largest value in the jth year (j = 1-6) for the kth sub population (k = 1....4) and v_{1mjk}, v_{2mjk} and v_{3mjk} are the associated values of v_1, v_2 and v_3. The expected value of the residual error e_{mjk} is zero. If the residual variance $E(e^2_{mjk})$ is denoted by R^2_m

it is known that R^2_m is proportional to σ^2_{mz}, the variance of $z_{(m)}$ as defined in the previous section. But from (7), para. 4.3,

$$\sigma^2_{mz} = \sigma^2_v \, \sigma^2_m / A^2_m \qquad \ldots\ldots (13)$$

Since the values of A_m (Appendix 2) differ from one sub population to another, a weighted regression needs to be performed minimising

$$Q_m = \sum_{k=1}^{4} A^2_{mk} \sum_{j=1}^{6} \left(z_{(m)jk} - \beta'_{om} - \beta_{1m} v_{1mjk} - \beta_{2m} v_{2mjk} - \beta_{3m} v_{3mjk} \right)^2 \qquad \ldots\ldots (14)$$

where A_{mk} refers to the kth sub population. The normal equations are shown in Appendix 3. By solving these equations estimated values of β'_{om}, β_{1m}, β_{2m} and β_{3m} are obtained. For the example considered in this note the following estimates were obtained.

Table 1

Regression parameters

Extremes (m)	β_{om}	β_{1m}	β_{2m}	β_{3m}	R^2_{mw}
1	0.9813	-0.3262	0.0617	0.4262	0.7629
2	1.5664	-0.3094	0.1972	0.1556	0.4987

5.3 <u>The weighted residual variance due to regression</u>

For a given set of values v_1, v_2 and v_3 the mth largest loss (from the top) usually written as $E(z_{(m)}/v_1, v_2, v_3)$ would be given by

$$\mu'_{vm} = \beta'_{om} + \beta_{1m} v_1 + \beta_{2m} v_2 + \beta_{3m} v_3 \qquad \ldots\ldots (15)$$

If μ'_{vmjk} is the value estimated by substituting in (15), the observed values v_{1mjk}, v_{2mjk} and v_{3mjk} (of v_1, v_2 and v_3) corresponding to the observed value $z_{(m)jk}$ of the dependent variable, i.e. log loss, the weighted residual variance is given by

$$R^2_{mw} = \frac{1}{20} \sum_{k=1}^{4} A^2_{mk} \sum_{j=1}^{6} \left(z_{(m)jk} - \mu'_{vmjk} \right)^2 \qquad \ldots\ldots (16)$$

For the example considered, the values shown in the last column of Table 1 were obtained.

5.4 The variance of the parent regression

Following the derivation of (8) in para. 4.3 it can be easily seen that

$$\sigma_v^2 = R_{mw}^2 / \sigma_m^2 \qquad \ldots\ldots (17)$$

Since R_{mw}^2 is based on a large number, i.e. 20 degrees of freedom, the following asymptotic values[2] may be used for the variance σ_m^2 of the reduced variable y_m

$$\sigma_1^2 = 1.6449$$

$$\sigma_2^2 = 0.6449$$

Expression (17) gives an estimate of the variance σ_v^2 of the parent distribution and may be denoted by σ_{vm}^2 since it is based on the mth largest observations of the dependent variable. The following estimates were obtained.

$$\left.\begin{aligned} \sigma_{v1}^2 &= 0.4638 \\ \sigma_{v2}^2 &= 0.7733 \end{aligned}\right\} \ldots\ldots (18)$$

5.5 The location parameter of the parent regression

From (9), para. 4.3,

$$\frac{E\left[z_{(m)} / v_1, v_2, v_3\right] - \mu_v}{\sigma_{vm}}$$

$$= (\mu'_{vm} - \mu_v) / \sigma_{vm}$$

$$= B_m + \frac{\bar{y}_m}{A_m}$$

so that

$$\mu_v = \mu'_{vm} - \sigma_{vm}\left(B_m + \frac{\bar{y}_m}{A_m}\right) \qquad \ldots\ldots (19)$$

as in (10), para. 4.3. Using (15) in para. 5.3, expression (19) can be

rewritten as

$$\mu_V = \beta_{om} + \beta_{1m} v_1 + \beta_{2m} v_2 + \beta_{3m} V_3 \qquad \ldots\ldots (20)$$

where

$$\beta_{om} = \beta'_{om} - \sigma_{vm} \left(B_m + \frac{\bar{y}_m}{A_m} \right) \qquad \ldots\ldots (21)$$

The expected values $\bar{y}_m$ of the reduced variable y_m are as follows[2]

$$\bar{y}_1 = 0.5772$$

$$\bar{y}_2 = 0.2704$$

The values of β_{om} are given in Table 2 for the four sub populations. (The parameter β_{om} would have a constant value in the case where the average number of fires per year was the same for all the sub populations)

Table 2

Values of β_{om}

Sub population		m	
		1	2
Sprinklered	- single storey	-0.8027	-0.4145
	- multi-storey	-0.9569	-0.6389
Non sprinklered	- single storey	-0.7502	-0.3380
	- multi-storey	-0.9094	-0.5685

5.6 Simplified form

In view of the differences in the annual frequencies of fires in the four populations, there are four regression equations for each extreme. Since v_1 and v_2 take the values +1 or -1 the equations can be reduced to the following simple form with just v_3 as the independent variable.

$$\mu_{vm} = E\left[z / v_3 \right] = \alpha_m + \beta_{3m} V_3 \qquad \ldots\ldots (22)$$

where

$$\alpha_m = \beta_{om} + \beta_{1m} V_1 + \beta_{2m} V_2 \qquad \ldots\ldots (23)$$

The values of α_m are given in Table 3.

Table 3

Values of $_m$

Sub population		m	
		1	2
Sprinklered	single storey	-1.1906	-0.9211
	multi-storey	-1.2214	-0.7511
Non sprinklered	single storey	-0.4857	-0.2258
	multi-storey	-0.5215	-0.0619

5.7 Conversion of results to original units

The set of parameter values β_{om}, β_{1m}, β_{2m} and β_{3m} for any m (m = 1 or 2 in this case) are estimates of the regression parameters β_o, β_1, β_2 and β_3 shown in (11), para. 5.1. Hence there are two estimates of the regression parameters. For a given total floor area (x_3) in units of 100 sq ft the expected value of μ_{vm} can be obtained from (22) by choosing any appropriate α_m and β_{3m} and putting $v_3 = \log_{10} x_3$. For a log normal distribution the expected value μ_x in the original units is[8]

$$\mu_x = e^{\mu + \frac{\sigma^2}{2}} \qquad \ldots\ldots (24)$$

where μ and σ are the mean and standard deviation of $z = \log_e x$. In the calculations, 10 has been used as base for z ; the logarithm of loss in units of £'000. Hence the expected loss in the original scale as estimated by the mth extreme observations is

$$\mu_x = 1000 \times e^{k\mu_{vm} + \frac{k^2}{2} \sigma^2_{vm}} \qquad \ldots\ldots (25)$$

where $k = \log_e 10 = 2.3026$

Figures 1 and 2 depict the relationship (25) between the expected loss and total floor area for m = 1 and 2. The expected (or mean) loss is at 1965 values.

5.8 Interpretation of results

For a given total floor area, the expected loss in a single storey building does not appear to differ very much from the expected loss in a multi-storey building. Perhaps, in a multi-storey building the horizontal spread of fire is restricted by better compartmentation but fire spreads vertically upwards. It is apparent that sprinklers reduce the expected loss to a considerable extent. From Fig. 1, for example, the expected loss in a building of total floor area 100,000 sq ft would be about £20,000 if the building were not sprinklered but sprinklers would reduce the loss to £4,000. The difference between the effect of sprinklers shown by Fig. 1 and Fig. 2 is due to random fluctuations.

5.9 Cost-Benefit of sprinklers

The problem considered in this paper is the expected reduction in loss due to sprinklers in a fire. This expected value is one of three ingredients in an assessment of the economic value of sprinklers. Probability of fire starting and probability of sprinkler heads not operating are the other two components. All these factors would be included in a cost benefit study of sprinklers (at the national level) which is beyond the scope of this paper. When all these factors are evaluated and the cost of installing and maintaining sprinklers is taken into account, it will be possible to determine a critical size for each major group of industrial and commercial buildings, above which it would be economically justifiable to provide sprinklers. In buildings smaller than the critical size the costs would be expected to exceed the benefits.

6. DISCUSSION

6.1 The location of the extreme

Like the average, median or mode the mth extreme loss in a risk category reflects the relative damage in this category. The observation with $m = 1$ is the largest and $m = n$ the smallest in a sample of n fires. The observation with the rank $m = \frac{n}{2}$ is the median. For a normal distribution the average, median and mode coincide; the mth extreme is situated at a distance from these central values.

6.2 The need for a modified model

For a multiple regression analysis assessing the contributions from various factors to the expected damage, only large losses are available at

present. Hence the problem studied in this paper is to estimate the regression parameters by using extreme observations. Repeated observations (over years) of an extreme with any chosen value of the rank m could be used for this purpose. But such estimates would be biased since the entire range of the fire loss variable has not been covered. In the modified model presented in this paper adjustments have been made to correct these biases.

6.3 Reasons for a single multiple regression

The main population has been divided into sub populations and extreme losses from each category have been considered in the analysis. The years provided replicated observations on the extremes. The model requires information on the number of fires per year in each sub population. This number has to be large and hence restricts the number of sub populations and the parameters that could be included. It is possible to perform a separate regression analysis for each sub population but this would also restrict the number of parameters unless data over a large number of years are used. Otherwise the number of degrees of freedom for the residual error would be small. For these reasons a single multiple regression analysis was carried out for each extreme (m).

6.4 The parent and the extreme

It was assumed that the parent probability distributions of the sub populations are log normal with a constant standard error σ_v. Expression (17) shows the relationship between σ_v and the (weighted) residual error R_{mw} obtained in the regression. The formula also involves the standard error σ_m of the reduced or standardised m^{th} extreme. The expected value μ_v of the parent and the expected value μ'_{vm} estimated by the regression are related through expression (19). This expression , apart from the mean $\bar{y}_m$ of the reduced extreme and σ_{vm} , includes parameters A_m and B_m the values of which depend upon the annual frequency (n) of fires. Thus the model takes into consideration the differences between sub populations in regard to the frequency of fires. The problem of confidence limits for the expected value and regression parameters is being investigated separately.

6.5 Combined regression

The values of the regression parameters vary from one large loss to another, i.e. m = 1 to 2 , as shown in Tables 1 to 3 and Figs 1 and 2. This variation is due to random fluctuations in the observations. It

would be better to estimate an overall mean value for each parameter β so that this mean and hence constant value could be used to assess the contribution of the concerned factor. For this purpose a combined regression analysis would have to be carried out using a number of extremes, say, $m = 1$ to r jointly and taking into consideration the variances as well as co-variances of the residual errors. This involves complicated computations which it is hoped to attempt in the near future. The model could also be generalised to include more factors like source of ignition, age of the building etc.

6.6 Similar study

Nelson and Hahn[9] have discussed the linear estimation of a regression relationship from censored data using order statistics. In this paper similar estimation procedures are considered using extreme order statistics.

7. CONCLUSIONS

As illustrated in this paper it is possible to modify the classical multiple regression model in order to assess the contributions of various factors as revealed by extreme observations. The model takes into consideration the biases due to the use of extremes and the variation in the frequency of fires from one sub population to another.

For a given total floor area, the expected loss in a single storey building does not appear to differ significantly from the expected loss in a multi-storey building. On the other hand, sprinklers reduce the expected loss to a considerable extent. For example, in a building with a total floor area of 100,000 ft^2 the "gain" would be £16,000 per fire. Figures 1 and 2 show the expected gain due to sprinklers for buildings of different sizes (total floor area). These qualitative and quantitative conclusions are based on the top two extremes only and it is hoped to improve the estimates by performing a comprehensive regression analysis combining the information on a number of extremes.

8. REFERENCES

1. RAMACHANDRAN, G. 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 Organization F.R. Note No.837.

2. RAMACHANDRAN, G. Extreme value theory and fire losses - further results. Department of the Environment and Fire Offices' Committee Joint Fire Research Organization F.R. Note No. 910.

3. BLANDIN, A. Bases techniques de l'assurance centre 1 incendie A. Martel, 1956.

4. RAMACHANDRAN, G. Fire loss indexes. Ministry of Technology and Fire Offices' Committee Joint Fire Research Organization F.R. Note No.839.

5. United Kingdom Fire and Loss Statistics (An annual publication). Her Majesty's Stationery Office.

6. STONE, P. A. The economics of factory buildings. Factory Building Studies No. 12. Her Majesty's Stationery Office.

7. PEARSON, E. S. and HARTLEY, H. O. Biometrika tables for statisticians, Vol. 1. University Press, Cambridge, 1962.

8. BENKERT, L. G. (1963). The log normal model for the distribution of one claim. ASTIN Bull. Vol. 11, Pt. 1, 9-23.

9. NELSON, W. and HAHN, G. J. (1973) Linear estimation of a regression relationship from censored data - Part II. Best Linear unbiased estimation and theory. Technometrics, 15, 133-150.

APPENDIX 1

Table 1

m = 1 (largest)

Sub population	Year	Loss (£'000) (x)	$z = \log_{10}x$	v_1	v_2	v_3
Sprinklered - single storey	1965	11.0	1.0414	+1	-1	2.1761
	1966	38.0	1.5798	+1	-1	1.7782
	1967	54.5	1.7364	+1	-1	2.4914
	1968	52.1	1.7168	+1	-1	2.0212
	1969	12.5	1.0969	+1	-1	2.2430
	1970	40.0	1.6021	+1	-1	2.4472
Sprinklered - multi-storey	1965	333.0	2.5224	+1	+1	3.2765
	1966	85.7	1.9330	+1	+1	3.6532
	1967	260.0	2.4150	+1	+1	3.3802
	1968	168.6	2.2269	+1	+1	3.6233
	1969	83.3	1.9207	+1	+1	3.1847
	1970	265.6	2.4242	+1	+1	3.5441
Non sprinklered - single storey	1965	1900.0	3.2788	-1	-1	2.9154
	1966	136.1	2.1339	-1	-1	1.6812
	1967	243.6	2.3867	-1	-1	2.7404
	1968	400.0	2.6021	-1	-1	3.3181
	1969	333.3	2.5228	-1	-1	3.4314
	1970	148.0	2.1703	-1	-1	2.8173
Non sprinklered - multi-storey	1965	1000.0	3.0000	-1	+1	3.3381
	1966	380.9	2.5808	-1	+1	3.3125
	1967	939.0	2.9727	-1	+1	4.0614
	1968	434.7	2.6382	-1	+1	3.2095
	1969	916.6	2.9622	-1	+1	3.3512
	1970	149.6	2.1750	-1	+1	2.6021

Table 2

m = 2

Sub population	Year	Loss (£'000) x	$z = \log_{10} x$	v_1	v_2	v_3
Sprinklered - single storey	1965	10.0	1.0000	+1	-1	2.2923
	1966	11.4	1.0569	+1	-1	1.7482
	1967	43.6	1.6395	+1	-1	2.6532
	1968	36.5	1.5623	+1	-1	2.0212
	1969	21.7*	1.3369*	+1	-1	2.1236
	1970	26.4	1.4216	+1	-1	1.9031
Sprinklered - multi-storey	1965	275.0	2.4393	+1	+1	3.1225
	1966	73.3	1.8651	+1	+1	3.8573
	1967	50.0	1.6990	+1	+1	3.6355
	1968	79.1	1.8982	+1	+1	3.0792
	1969	41.6	1.6191	+1	+1	1.5563
	1970	209.6	2.3214	+1	+1	3.0614
Non sprinklered - single storey	1965	141.0	2.1492	-1	-1	2.0969
	1966	135.2	2.1310	-1	-1	2.0792
	1967	99.0	1.9956	-1	-1	2.3979
	1968	347.8	2.5413	-1	-1	3.0334
	1969	166.6	2.2217	-1	-1	2.4771
	1970	46.4	1.6665	-1	-1	2.6628
Non sprinklered - multi-storey	1965	445.0	2.6484	-1	+1	3.1538
	1966	294.2	2.4686	-1	+1	3.1126
	1967	272.7	2.4357	-1	+1	2.7782
	1968	347.8	2.5413	-1	+1	2.6990
	1969	750.0	2.8751	-1	+1	4.1329
	1970	140.0	2.1461	-1	+1	3.1206

* Estimated (median value of 5 years)

APPENDIX 2

Extreme value parameters

Sub population (1)		Estimated no. of fires (2)	n (3)	m = 1		m = 2	
				A_1 (4)	B_1 (5)	A_2 (6)	B_2 (7)
Sprinklered	- single storey	250	125	2.7375	2.4089	2.5000	2.1444
	- multi-storey	500	250	2.9750	2.6521	2.7375	2.4089
Non sprinklered	- single storey	200	100	2.6700	2.3264	2.4200	2.0538
	- multi-storey	400	200	2.8800	2.5759	2.6700	2.3264

APPENDIX 3

The problem is to minimise

$$Q_m = \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} \left(Z_{(m)jk} - \beta'_{om} - \beta_1 V_{1mjk} - \beta_{2m} V_{2mjk} - \beta_{3m} V_{3mjk} \right)^2$$

Differentiating Q_m successively with respect to β'_{om} , β_{1m} , β_{2m} and $_{3m}$ and equating each derivative to zero the normal equations are

$$\sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} Z_{(m)jk} = 6\beta'_{om} \sum_{k=1}^{4} A_{mk}^2 + \beta_{1m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{1mjk}$$
$$+ \beta_{2m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{2mjk} + \beta_{3m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{3mjk}$$

$$\sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} Z_{(m)jk} \cdot V_{1mjk} = \beta'_{om} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{1mjk} + \beta_{1m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1} V_{1mjk}^2$$
$$+ \beta_{2m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{2mjk} \cdot V_{1mjk}$$
$$+ \beta_{3m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{3mjk} \cdot V_{1mjk}$$

$$\sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} Z_{(m)jk} V_{2mjk} = \beta'_{om} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{2mjk} + \beta_{1m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{1mjk} \cdot V_{2mjk}$$
$$+ \beta_{2m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{2mjk}^2 + \beta_{3m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{3mjk} \cdot V_{2mjk}$$

$$\sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} Z_{(m)jk} \cdot V_{3mjk} = \beta'_{om} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{3mjk} + \beta_{1m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{1mjk} \cdot V_{3mjk}$$
$$+ \beta_{2m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{2mjk} V_{3mjk} + \beta_{3m} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{3mjk}^2$$

Inserting the values of v_1 and v_2 the four equations reduce to:

(1)
$$\sum_{k=1}^{4} A_{mk}^2 \bar{z}_{mk} = \beta'_{om} \sum_{k=1}^{4} A_{mk}^2 + \beta_{1m}\left[A_{m1}^2 + A_{m2}^2 - A_{m3}^2 - A_{m4}^2\right]$$
$$+ \beta_{2m}\left[-A_{m1}^2 + A_{m2}^2 - A_{m3}^2 + A_{m4}^2\right]$$
$$+ \beta_{3m} \sum_{k=1}^{4} A_{mk}^2 \bar{V}_{3mk}$$

(2)
$$\left[A_{m1}^2 \bar{z}_{m1} + A_{m2}^2 \bar{z}_{m2} - A_{m3}^2 \bar{z}_{m3} - A_{m4}^2 \bar{z}_{m4}\right]$$
$$= \beta'_{om}\left[A_{m1}^2 + A_{m2}^2 - A_{m3}^2 - A_{m4}^2\right] + \beta_{1m} \sum_{k=1}^{4} A_{mk}^2$$
$$+ \beta_{2m}\left[-A_{m1}^2 + A_{m2}^2 + A_{m3}^2 - A_{m4}^2\right]$$
$$+ \beta_{3m}\left[A_{m1}^2 \bar{V}_{3m1} + A_{m2}^2 \bar{V}_{3m2} - A_{m3}^2 \bar{V}_{3m3} - A_{m4}^2 \bar{V}_{3m4}\right]$$

(3)
$$\left[A_{m1}^2 \bar{z}_{m1} + A_{m2}^2 \bar{z}_{m2} - A_{m3}^2 \bar{z}_{m3} + A_{m4}^2 \bar{z}_{m4}\right]$$
$$= \beta'_{om}\left[-A_{m1}^2 + A_{m2}^2 - A_{m3}^2 + A_{m4}^2\right]$$
$$+ \beta_{1m}\left[-A_{m1}^2 + A_{m2}^2 + A_{m3}^2 - A_{m4}^2\right] + \beta_{2m} \sum_{k=1}^{4} A_{mk}^2$$
$$+ \beta_{3m}\left[-A_{m1}^2 \bar{V}_{3m1} + A_{m2}^2 \bar{V}_{3m2} - A_{m3}^2 \bar{V}_{3m3} + A_{m4}^2 \bar{V}_{3m4}\right]$$

(4)
$$\frac{1}{6} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} z_{(m)jk} \cdot V_{3mjk}$$
$$= \beta'_{om} \sum_{k=1}^{4} A_{mk}^2 \bar{V}_{3mk}$$
$$+ \beta_{1m}\left[A_{m1}^2 \bar{V}_{3m1} + A_{m2}^2 \bar{V}_{3m2} - A_{m3}^2 \bar{V}_{3m3} - A_{m4}^2 \bar{V}_{3m4}\right]$$
$$+ \beta_{2m}\left[-A_{m1}^2 \bar{V}_{3m1} + A_{m2}^2 \bar{V}_{3m2} - A_{m3}^2 \bar{V}_{3m3} + A_{m4}^2 \bar{V}_{3m4}\right]$$
$$+ \frac{\beta_{3m}}{6} \sum_{k=1}^{4} A_{mk}^2 \sum_{j=1}^{6} V_{3mjk}^2$$

The four sub populations are denoted by the subscript k as follows:

$k = 1$ Sprinklered, single storey

$k = 2$ Sprinklered, multi-storey

$k = 3$ Non sprinklered, single storey

$k = 4$ Non sprinklered, multi-storey

The terms $\bar{z}_{mk}$ and $\bar{v}_{3mk}$ are the averages for the sub population k, i.e.

$$\bar{z}_{mk} = \frac{1}{6} \sum_{j=1}^{6} z_{mjk}$$

$$\bar{v}_{3mk} = \frac{1}{6} \sum_{j=1}^{6} v_{3mjk}$$

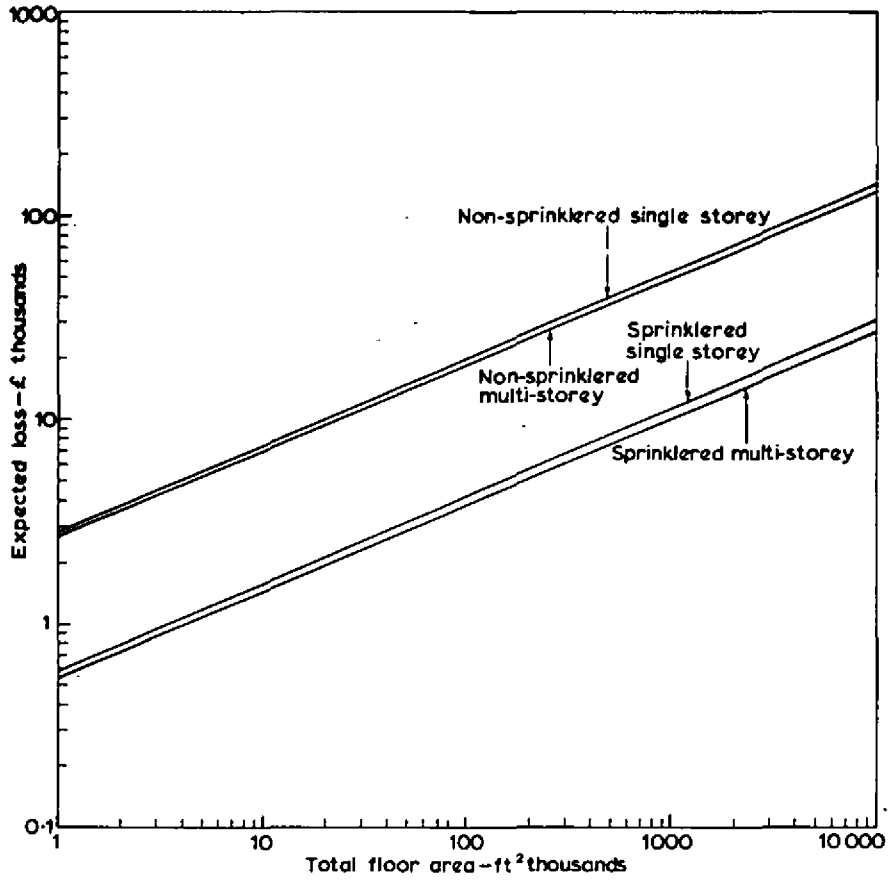

Figure 1 Total floor area and loss (m=1)

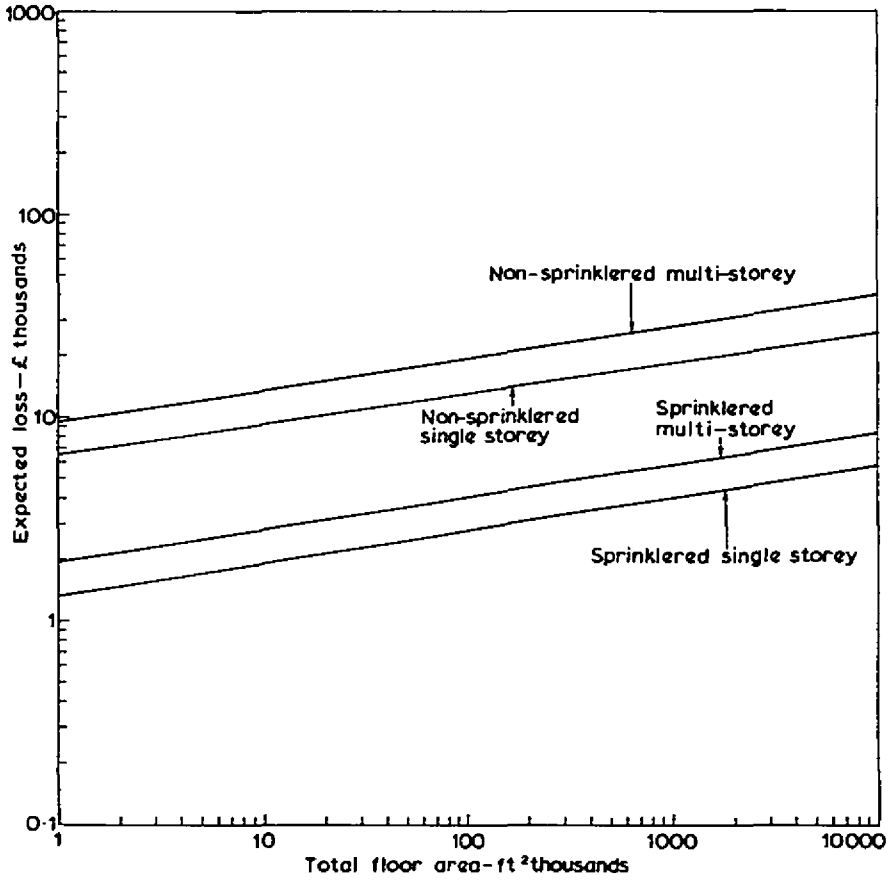

Figure 2 Total floor area and loss (m=2)

Printed in England for Her Majesty's Stationary Office
by Roma Printers London W13 8PR from copy supplied