Arab Journal of Administrative Sciences

2000, Vol. 7, No. 2, 327-338

MODELLING FIRE LOSSES: A SAUDI EXPERIENCE

Ahmed A. Kamhawey Abaza

King Suad University

Kingdom of Saudi Arabia

Key Words

Stochastic Process; Model fitting; Fire loss; MPY; Simulation.

Abstract

The totel amount of fire in a given period can be modeled as a risk process. In this paper a practical application is presented in which the probability distribution of a fire loss experience for a Saudi large estate company is modeled. Also, the article is concerned with the choice of the best fitting.

Introduction

At a certain point in time, results of theoretical work have to be applied to practical situations. For example, there are two main aspects of general insurance in which knowledge of the structure of the risk elements are needed (Beard,1958) first, in the rate making process; and second, in dealing with the question of financial stability (monetary risk). Traditional methods of rate making are based only on an estimate of the mean expected loss. Financial stability studies are usually based on an estimate of the variance of possible losses. Total loss distribution, F(x,t) is of great importance to insurers in their task of determining appropriate premiums, reinsurance policies, credibility measures and contingency retention levels. It is also important when property owners decide to purchase insurance coverage or invest in fire protection, or when an insurer designs deductible schedules. It is a basic assumption in all actuarial work and risk theory that there is a probability distribution of loss amount underlying the risk process. In other words, if a loss occurs, there is a probability than S(x) that the loss will be for an amount equal to or less than x. In theoretical studies this distribu-

Submitted January 1999, Accepted May 1999.

tion is often presented as continuous function, having a derivative $S'(x) = s(x)$, $s(x)$ which is called the probability density function of fire loss amount.

In this paper a practical application is presented by modeling the probability distribution of a Saudi fire loss amount. It is also concerned with the choice of an appropriate model for this distribution.

Fire Loss as a Stochastic Process

The total amount of fire in a given period can be modeled as a risk process characterized by two stochastic variables (Shpilberg,1977):

The number of fires and the amounts of the losses. If,

$P_r(t)$ = Probability of r losses in the observed period, t

$S(x)$ = Probability that, given a fire loss, has an amount $\leqslant x$

S^{*r} = r^{th} convolution of the distribution function of fire loss amount, $S(x)$,

Then the probability that total loss in a period of length t, is $\leqslant x$ can be expressed as:

$$F(x, t) = \sum_{r=0}^{\infty} P_r(t) S^{*r}(x) \quad (1)$$

a compound distribution whose mean and variance are given by:

$$m_F = m_P m_S$$

$$\sigma_F^2 = m_P \sigma_S^2 + m_S^2 \sigma_P^2 \quad (2)$$

where m_P and m_S are the mean number of losses and the mean loss per fire and σ_P^2 and σ_S^2 are the corresponding variances (Panjer and Willmot,1992).

Model Selection

In choosing a model, one must consider the extent and format of the data. In general, the more extensive the data, the more complicated the model one can fit. However, this is not necessarily desirable. Even with extensive data, a simpler model, which gives an adequate fit is to be preferred to a more complicated model due to ease of interpretation of the results and ease of model fitting. It is often desirable to fit more than one model in a given situation. Results obtained by previous researchers indicated that the poisson and negative binomial distributions are likely candidates to describe claim frequency(Shpilberg and de Neufvill, 1975). Also, the results of previous research suggested that the gamma (including the exponential), lognormal and Pareto distributions are promising candidates to be applied to severity case(Panjer and Willmot,1983).

The Data and General Procedures

The data subjected to analysis are the fire losses to buildings owned by a large estate company in Riyadh, Saudi

Arabia. The general manager of this company provided the researcher, on a personal basis, with unpublished fire experiences from 1974 through 1997. An average of 303 buildings were exposed to loss during each of these years, and a total of 80 fire losses were recorded. Loss records were maintained in sufficient detail so that both frequency and severity distribution could be tabulated. Although the exposed units clearly are not homogeneous, the available information is not sufficient to permit any subdivision of the loss experience. Consequently, the data are treated as if all the exposure units are identical. The approach, although not theoretically precise, should produce reasonably accurate results as long as the loss distributions are stable over the observation period and the proportionate representation of each type of building in the exposure base does not change appreciably. Also the suggested inflation adjustments (Shpilberg and de Neufvill, 1975). are applied throughout the analysis. The data base is summarized in Table 1, which shows frequency, average severity and total annual Riyal losses (3.75 Riyal = 1 U.S Dollar) for each year of the sample period. Frequency appears reasonably stable over time, although evidence shows that average frequency increased toward the end of the sample period (Stephens,1974). Both average severity and total losses fluctuated markedly during the sample period. The largest single loss, amounting to more than S.R. 600,000 in current prices, occurred in 1980. Total losses exceeded S.R. 100,000 in only four of the 24 years.

Prior Work

Several probability distributions have been suggested by researchers to model the fire loss amount distribution, S(x). All of them are relatively successful in modeling the mid range of the distribution, but their adequacy to model the region of the larger values of X(tail) cannot be determined with confidence from observed data. Paradoxically, the least known part of S(x) has the greatest effect on the numerical results of many practical problems such as optimal investment decisions in insurance and the optimal reinsurance coverage. A more than adequate review of the types of distributions used in the past to model S(x) is contained in a paper by Benktander (Gerber *et al*,1987).

The Distributional Approach

In order to utilize this approach, best fitting frequency and severity distributions must be identified. It is often desirable to fit more than one model in a given situation. It is possible that more than one model ade-

Table 1
Fire Loss Experience of A Saudi Large Estate Company

Year	Frequency	Average Severity (S.R)	Total Losses (S.R)	No-Of Exposure Units
1974	2	4,392	8,784	279
1975	4	17,820	71,280	270
1976	1	3,671	3,671	273
1977	4	4,666	18,664	476
1978	2	15,446	30,892	287
1979	2	1,983	3,966	282
1980	3	210,966	631,626	292
1981	4	2,966	11,464	297
1982	3	42,398	127,194	302
1983	2	2,475	4,950	308
1984	1	8,028	8,028	312
1985	4	7,613	30,452	314
1986	1	8,028	8,28	312
1987	2	4,740	9,480	308
1988	1	14,790	14,790	310
1989	4	2,169	8,676	306
1990	2	57,099	114,498	305
1991	2	2,575	5,450	310
1992	6	5,469	32,814	320
1993	6	17,644	105,864	315
1994	6	6,890	41,340	325
1995	5	2,446	12,230	330
1996	7	2,774	19,418	322
1997	7	6,612	46,284	329
Average	3,375	18,884	57,052	303
Standard	Deviation	1.910	42,003	124,971

quately describes the data and one would hope for good agreement between these models. Results obtained by previous researchers indicated that the poisson and negative binomial distributions are likely candidates to describe claim frequency. Utilizing parameters estimated from the sample data and expected frequencies were generated under the poisson and negative binomial hypotheses by using the SPSS 7.5 for Windows software. The fitted poisson frequencies were obtained by assuming a theoretical q-value equal to the mean 0.0114665. Poisson probabilities, which denoted by P_0, P_1, P_2, were then calculated recursively. Also the expected numbers of exposure units making 0,1,2,3 and 4 claims then calculated by multiplying these probabilities by 7062 units.

To fit the negative binomial distribution, the theoretical and observed means and variance were equated:

$$\text{mean} = k\,(1\text{-}P)\,/\,P = 0.0114665;$$

$$\text{variance} = k\,(1\text{-}P)\,/\,P^2 = 0.0119047.$$

Dividing the first of these two equations by the second, we obtain P=0.963191 whence k = 0.299648.

The fitted negative binomial probability P_0, P_1, P_2,......., , were then calculated recursively. Again, the fitted frequencies were obtained by multiplying these probabilities by 7062 units.

These values were compared with the observed frequency distribution, and the results of this comparison are presented in Table 2.

Table 2
Goodness of Fit Tests: Fire Loss Frequency

Number of Losses	Number of Exposure Units		
	Observed	Poisson	Negative Binomial
0	6983.0	6981.485	6983.08
1	77.0	80.052	77.021
2	2.0	0.459	1.842
3	0.0	0.002	0.052
4	0.0	0.000	0.002
5	0.0	0.000	
chi-square Value		5.291	0.068

The estimated Poisson parameter is 0.011467. The negative binomial parameters are 0.299648 and 26.1673 . In order to obtain and compute the chi-square values, the "five units per category" rule of thumb had to be violated. Combining all multiple loss categories yielded chi-square statistics with one degree of freedom. The value of $\chi^2_{.05,1} = 3.841$ and $\chi^2_{.01,1} = 6.635$.

Visual inspection indicates that both distributions fit the Saudi fire loss moderately well. Thus, if the violation of the rule does not substantially distort the results, the hypothesis that frequency is Poisson distributed can be rejected at the 5 percent. The negative binomial hypothesis is not rejected at either level of significance.

These findings are in agreement with prior expectations; provide additional evidence of the superiority of the negative binomial and of its excellent fit to the observed frequencies (See Table 2).

On the basis of these results, the negative binomial distribution was adopted as a frequency distribution.

It is more difficult for severity than for frequency to find the best-fitting theoretical distribution. The first step in the process of finding a severity distribution was to obtain an overall picture of the shape of the empirical distribution. The results of previous researches suggested that the gamma (including the exponential), lognormal and pareto distributions are promising candidates (Teugels,1985).

The hypothesis that the fire loss severity distribution was exponential was tested using an empirical density function test and the Anderson-Darling statistic, discussed in Stephens (6), and could be rejected at better than the 1 percent level of significance. Rejection of the exponential distribution and shape effectively eliminated the pareto distribution as a serious candidate but failed to discriminate between the lognormal and gamma distributions. Since their densities are characterized by "light" tails where as lognormal distributions are characterized by "heavy" tails, and the search for a fire loss must be narrowed to those densities characterized by heavy tails, so the researcher must eliminate from consideration the gamma family (Embrechts and Goldie,1982).

The Severity Distribution

Hence, among the distributions traditionally applied to property- liability loss data, only the lognormal remained to be tested.

As we know, if a random variable Y is lognormally distributed, then the random variable $X = \ln Y$ is normally distributed. Thus, the normal distribution could be utilized to make a judgment on the applicability of the lognormal distribution (Stephens, 1974).

In view of the foregoing test results, the assumption was made that the loss distribution for each building would be described by a lognormal distribution with the parameters varying from building to building. If the variation in the lognormal distribution parameters is itself described by a probability distribution, an unconditional severity distribution can be derived and expressed as follows (Cummins and Friefelder, 1978):

Let Y be a random variable equal to the severity of loss. Suppose that Y follows a lognormal distribution with parameters m and h and let X = ln Y. The random variable X is normally distributed with a mean m and precision h. Assume that m and h differ among buildings, that the conditional distribution of m is normal with mean m and variance $(1/\tau h)$, and h is gamma distributed with parameters α and β . The unconditional distribution of X can then be obtained by performing the following integration:

The parameters of the non-central t were estimated by using the method of moments. Since there are three parameters, three moments were needed and the mean, variance and fourth moment about the origin were used. The third moment could not be used due to the symmetry of the t-distribution. The estimated value of the non-centrality parameter m is 8.3214; the degree of freedom parameter is 4, and the scale parameter w is 1. By using the estimated parameters a theoretical severity distribution was generated depending on the SPSS 7.5 for Windows Software. This distribution and its empirical counterpart are presented in Table 3.

The t distribution utilized above has a non-centrality parameter of 8.19903, 4 degrees of freedom and a scale parameter of 1. To describe the continuous severity distribution, intervals with equal probability for the non-central t distribution were selected. The last three categories were

$$f(x) = \int_h \int_\mu f(x/\tau h) g_1 (\mu/m, \tau h) g_2 (h/\alpha, \beta) d\mu \, dh \qquad (3)$$

where $g_1 \; \mu/m, \tau h$ and $g_2 \; (h/\alpha,\beta)$ are distributions of μ and h, respectively.

Raiff and Schlaifer (Raiffa and Scglaifer,1961:304) give the result that f(x) is non-central t- distributional, i.e.,

$$f(x) = \left(\frac{w}{\gamma}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{\gamma+1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} \left[1 + \frac{w(x-m)^2}{\gamma}\right]^{-\left(\frac{\gamma+1}{2}\right)} \qquad (4)$$

Where $W = \frac{\gamma\tau}{2\beta(1+\tau)}$ and $\gamma = 2\alpha$

Table 3
Goodness of Fit Tests: A Saudi Fire Loss Severity

Individual Loss Amounts	Observed No. of Losses	Percent	Non Central No. of Losses	Percent
less than S.R 1000	7	8.75	8	10.00
S.R 1000-	14	17.50	8	10.00
S.R 1500-	11	13.80	8	10.00
S.R 2500-	6	7.50	8	10.00
S.R 3500-	8	10.00	8	10.00
S.R4500-	7	8.75	8	10.00
S.R 6500-	7	8.75	8	10.00
S.R 9500-	3	3.75	8	10.00
S.R 17,500-	7	8.75	8	10.00
S.R 30,000-	5	6.25	4	5.00
S.R 40,000-	0	0.00	2	2.50
S.R50,000- and over	5	6.25	2	2.50
	80			

combined for chi-square testing, (Kendall and Stuart, 1961).

The value of chi-square for the data in Table. 3 is 12.75 . The value falls between the critical values for α = 0.10 and 6 degrees of freedom and ($x^2_{0.0,6}$ = 0.10) and ل=0.10 and 9 degrees of freedom ($x^2_{0.1,9} = 14.684$).

In the first case, the hypothesis that the data come from a non-central t distribution is rejected; in the second case the hypothesis can be accepted. In view of the statistical results and the theoretical support for the non-central t, this distribution was adopted for use in the simulation phase of the analysis.

Calculating the Saudi MPY for Fire Experience

The traditional focus of risk management on loss identification, reduction and transfer gradually being supplanted by a more balanced approach, with retention playing a prominent role. This evolutionary process has directed attention to techniques for making decisions about retention plans. Such decisions should be made by evaluating a company's loss experience according to theoretically sound business decision rules, the maximum probable yearly aggregate loss (MPY) which is a fractile point in the right tail of total annual loss cost-distribution. The theoretically preferred method for calculating the MPY is to develop the distribution of total annual loss costs and to find the value L* such that F(L*) = 0.99, where F(*) is the relevant cumulative distribution function. L* then becomes the MPY estimate. Convolutions of the severity distribution then will be probability weighted to yield F(*)(Hachemeister,1975). The formula for this calculation is shown in equation (1), which can be rewritten as follows.

$$F(L) = \sum_{q=0}^{\infty} g(q).\, H^{q*}(L) \qquad (5)$$

where g(q): is the frequency density function.

H(L): is the severity distribution function.

and $H^{q*}(L)$: is the q th convolution of the severity distribution function, i.e., the distribution function of the sum of q severity random variables.

As this method requires the data to be kept on a claim-by-claim basis, it can not be used when only yearly aggregated data are available. In many cases a closed form solution to equation (5) is difficult or impossible to obtain. An alternate approach that can be used to obtain the cumulative distribution function is simulation; It can be a very useful and powerful tool for solving problems in general insurance. There are two main situations in which it may well be the best approach (Hossak *et al*,1983):

1. When the problem can not be solved exactly mathematically.

2. When an exact mathematical solution is possible but extremely difficult or tedious, and adequate approximate results can be obtained quickly and simply by simulation.

Once the frequency and severity distributions have been established, the simulation can proceed. For each simulation, the following is done:

1. Randomly select the number of claims, n, from the frequency distribution.

2. Do the following, n times: Randomly select a claim amount from the severity distribution.

3. The "simulated" loss amount equals the sum of all claims amounts generated by step 2.

These three steps are repeated, until the desired number of simulations are completed. Of course, a computer is used to perform the simulations and to keep track of the results.

A table summarizing the results of a simulation analysis is presented (Table 4). The following four types of statistics are presented in this table:

Table 4
Total Unlimited Losses

Average	S.R	59,142
STD DEV	S.R	121,794
MAX SIM	S.R	749,682
ITH-%ILE	S.R	2,546
5TH-%ILE	S.R	12,235
10TH-%ILE	S.R	33,271
25TH-%ILE	S.R	146,235
50TH-%ILE	S.R	298,346
75TH-%ILE	S.R	427,874
90TH-%ILE	S.R	613,246
95TH-%ILE	S.R	640,259
99TH-%ILE	S.R	667,587

1. The average;
2. The standard deviation;
3. The maximum loss simulated; and
4. The percentiles of the loss distribution.

Results and Discussion

At a certain point in time, the results of the theoretical work have to be applied to practical situations. In this article the researcher has adopted, modified and applied the distributional method to model a Saudi fire loss amount. The data subjected to analysis are the fire losses to buildings owned by a large estate company in Riyadh during the period of 1974 through 1997. In view of the statistical results and theoretical support, the simulation approach could be used to obtain the total loss cumulative distribution. Monte Carlo simulations utilizing the negative binomial distribution for frequency and the non-central t distribution for severity produced 0.99 MPY value of S.R. 667,587 for fire losses.

A table summarizing some results of the simulation analysis is presented (Table 4). The first figure in the table show Average Annual Loss Amount which is the arithmetic mean of annual losses generated by all simulated years.

The second figure shows the Standard Deviation. This measure of the variation is useful in assessing the uncertainty or risk with each loss.

The maximum simulated loss is shown in figure labeled MAX SIM. This amount, S.R. 749,682, represents the largest of the losses simulated.

Finally, it is very important to say that, precise results can be obtained by the distributional method only by selecting the right theoretical distributions.

References

Beard, R.E.1958. Analytical Expressions of the Risks Involved in General Insurance. *Transactions of The 15th International Congress of Actuaries, 2.*

Cummins, J. and Friefelder, L.1978. A Comparative Analysis of Alternative Maximum Probable Yearly Aggregate Loss Estimators. *Journal of Risk and Insurance*, 45(2): 27-52.

Embrechts, P. and Goldie, C.1982. On Convolution Tails. *Stochastic Processes and Their Applications,* 13: 263-278.

Gerber, H., Goovaerts, M., and Kaas, R.1987. On the Probability and Severity of Ruin. *Astin Bulletin*, 17: 151-163.

Hachemeister, C.1975. *Credibility for Regression Models with Application to Trend.* Proceeding of the Berkeley Actuarial Research Conference on Credibility, New York: Academic Press.

Kendall, M. and Stuart, A.1961. *The Advanced Theory of Statistics*, New York: Hafner Publishing Co.,2: 430-433.

Hossak, I., Pollard, J and Zehnwirth, B.1983. *Introductory Statistics with Applications in General Insurance*, Sydney: Cambridge University Press,195-197.

Panjer, H. and Willmot, G. 1992. *Insurance Risk Models.* Society of Actuaries, Schaumburg, U.S.A, 295-298.

Panjer, H. and Willmot, R.1983. Compound Poisson Models in Actuarial Risk Theory. *Journal of Econometric*, 23: 63-76.

Raiffa, H. and Schlaifer, R.1961. *Applied Statistical Decision Theory*, Cambridge, Mass: The M.I.T. Press.

Shpilberg, D.1977. The Probability Distribution of Fire Loss Amount. *Journal of Risk and Insurance*, 44(1): 103-115.

Shpilberg, D. and de Neufvill, R.1975. Use of Decision Analysis for Optimizing Choice of Fire Protection and Insurance: An Airport Study. *Journal of Risk and Insurance,* 42(1):133-149.

Stephens, M.A.1974. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of The American Statistical Association,* 69: 730-737.

Teugels, J.1985. Approximation and Estimation of Some Compound Distributions. *Insurance: Mathematics and Economics*, 4:143-153.

الملخص

بناء نموذج خسائر الحريق: من واقع الخبرة السعودية

أحمد عبدالله قمحاوي أباظه
جامعة الملك سعود

إجمالي خسائر الحريق في فترة معينة يمكن نمذجتها باعتبارها عملية تعامل مع الخطر، وفي هذه الدراسة تم اختيار واختبار أفضل التوزيعات الاحتمالية التي يمكنها تمثيل وتقدير تكرار حدوث الحريق، وشدة الوطئة حين الحدوث لخبرة سعودية، ثم استخدام أسلوب المحاكاة للحصول على توزيع الخسارة الإجمالية مع تقدير الخسارة القصوى السنوية المحتملة.

Ahmed A. Kamhawey (Ph.D., City University, London, 1981), Professor of Actuarial Science, King Saud University. Areas of Interest Include: Risk Management, Application of Statistics in General Insurance and Credibility Theory.