Journal of Business and Social Science Review
Issue: Vol. 3; No.8; August 2022 (pp.1-26)
ISSN 2690-0866(Print) 2690-0874 (Online)
Website: www.jbssrnet.com
E-mail: editor@jbssrnet.com
Doi: 10.48150/jbssr.v3no8.2022.a1

An Example of Early Quantitative Fundamental Analysis: Forecasting Insured Losses Due to Catastrophes

Joel Rentzler
Professor of Finance
Baruch College, CUNY
E-mail: Joel.Rentzler@baruch.cuny.edu

Robert Ferguson
President
AnswersToGo

Abstract

Lacking computers, investment analysis prior to the sixties was seldom quantitative. Toward the end of the sixties, time-sharing computers were available. However, memory was tiny by today's standards. Only rudimentary statistical analysis was feasible. A few investment analysts with quantitative training tried to gain an advantage by extracting private information from public data. This paper presents one example, an analysis of catastrophic insurance losses. In 1965 Hurricane Betsy's destructive path made insurance analysts fearful that such catastrophic losses were a new norm. This motivated the analysis presented here to put this loss in perspective.

Lacking computers, financial analysis of investments prior to the sixties was seldom quantitative. Toward the end of the sixties, time-sharing computers (using teletype machines as terminals and paper tape as backup media) were available. However, usable memory was tiny by today's standards (64 kilobytes was a lot). Basic statistical analysis was feasible, and a few financial analysts with quantitative training began to try to gain an advantage over their colleagues by extracting what they viewed as private information from public data. This paper presents one example, an analysis of catastrophic insurance losses in the United States. Hurricane Betsy was an intense tropical cyclone that devastated Florida in 1965. Its destructive path, which caused $500M in losses in then current dollars, made many professional investors fearful that such catastrophic losses were a new norm that would adversely impact insurance stocks. This motivated the analysis presented here, which was written as an internal memorandum in 1969 at one buy-side boutique research firm to put the loss from Hurricane Betsy in perspective.

Introduction

Insurance is purchased as protection against certain kinds of random events. Examples include losses due to fire, disease, crime, and accident. Policyholders find it desirable to insure against most events that can have a serious detrimental impact on them. When these events occur in large enough numbers, a reasonably accurate prediction can be made of the average per capita loss. This fact enables an insurance company to cover its losses with a high degree of certainty with a premium only moderately above the average loss. These kinds of events seldom produce wide swings in a company's earnings.

In the case of infrequent large-scale events such as hurricanes, tornados, and other major catastrophes, an insurance company does not have the protection of the law of averages. For example, in 1965 there were only 4 hurricanes with just one considered a major hurricane (Cat 3+), Hurricane Betsy. Hence the average per capita loss can fluctuate enough to produce extremely erratic earnings. Because of this, it is important to know what the catastrophic losses experienced during a particular period might amount to. Also, to analyze earnings trends, it is necessary to know to what extent historical catastrophic losses were abnormal.

The purpose of this memorandum is to outline a method for determining to what extent historical catastrophic losses were abnormal and for predicting what they might be in the future.

Definitions and assumptions

A catastrophe is defined as a single event, which produces a total insured loss of at least $1.0 million.[i] The kinds of events, which produce catastrophic losses, include hurricanes, tornados, disorders, hailstorms, and windstorms. In this memorandum, it is assumed that the number of catastrophes, which occur during a period, is independent of the number, which occurred prior to that period. It is also assumed that the size of a catastrophe is independent of the number and size of those, which occurred previously.

The size of a catastrophe

Estimating the probability distribution of the size of a catastrophe was accomplished by fitting a probability function to a sample of catastrophic losses. Other things equal, the larger the sample the more accurate the fitting process. In this case, a large sample could be obtained only by combining data from several different years. This is not proper unless the probability distribution is the same for these years. Because it is plausible to expect the size of a catastrophe to have a trend over time, it was necessary to test the reasonableness of assuming that the probability distribution did not change significantly from year to year before combining the data. The entire process is outlined below.

A record of all catastrophic losses from 1953 to 1966 was obtained from the Insurance Information Institute. This record is shown in Appendix A. The reasonableness of treating all or a major part of this data as coming from the same probability distribution was checked in two ways. First, the data for each year was used to estimate the appropriate fractiles of that year’s distribution. These estimates were plotted on chart paper and a smooth curve was fit to them by eye. These curves represent an initial estimate of the cumulative probability distribution. The charts for 1954, 1964 and 1965 are typical and are shown in Appendix B, along with a more detailed description of their derivation. (A complete set of charts for all 14 years is available from the authors on request.)

A visual inspection of the charts in Appendix B reveals little evidence for rejecting the hypothesis that the probability distribution of the size of a catastrophe is relatively constant over time. If there were a trend to the size of a catastrophe, the shapes of the curves would be expected to vary from year to year. However, some variation would be expected due to random effects even if there is no trend. The key is to see if there is more variation between widely separated years than between adjacent years. As an example, consider the variation between the curves for 1964 and 1965 as compared with that between the curves for 1954 and 1965. This comparison reveals little evidence of a long-term trend. The same holds true when other one-year and eleven-year gap graphs are considered. An examination of the shapes of the curves for each year is another useful exercise. Most of the data is consistent with the kind of shape exhibited by the curve for 1964. A priori, a change in the probability distribution of the size of a catastrophe would be expected to take the form of a long-term trend. There are only a few isolated years of seemingly different data and no impression of a long term-trend, so perhaps the differences in shape over the years can be attributed to randomness.

In view of the result of the visual examination, a formal test of the hypotheis that the probability distribution of the size of a catastrophe is constant over time was made using a non-parametric runs test(described in Appendix C). The concept is that, if the observations from two samples from two unknown distributions are combined and arranged in order of increasing size, the number of runs of observations from each sample conveys information about the likelihood that the two distributions are the same. The number of runs tends to be larger when the distributions are the same than when they are different. In this case, the procedure consisted of combining the catastrophic loss figure for two selected years, arranging them in order of increasing size, and determining whether the number of runs observed was consistent with the hypothesis that the two distributions are the same. This was done for all 91 possible combinations of years taken two at a time. The results are shown in Appendix D. Appendix E contains a detailed analysis of these results which shows that they are entirely consistent with the hypothesis that the probability distribution of the size of a catastrophe is constant over time. Therefore it is reasonable to combine all the data from 1953 to 1966.

Next, an intitial estimate of the cumulative distribution was obtained by applying the method outlined in Appendix B to the combined data. The resulting scatter diagram is shown in Appendix F. Several kinds of probability distributions were fit to the data, to find an analytic representation. A gamma distribution was found to be adequate. The procedure is described more fully in Appendix F.

The final result is that the size of a catastrophe is distributed approximately as follows.

$$f(y) = \frac{y^{\alpha} e^{-\left(\frac{y}{\beta}\right)}}{\Gamma(\alpha+1)\,\beta^{(\alpha+1)}}$$

(1)

$y \equiv$ $ln(x)$.

$x \equiv$ Insured loss (millions of dollars).

$\alpha \equiv$ 1.3765

$\beta \equiv$ 0.6796

A plot of this cumulative distribution is shown in Appendix F.

The number of catastrophes

The number of catastrophes occurring during a year was assumed to be independent of the number occurring during any prior year. Specifically, the number was conjectured to have a Poisson distribution with the parameter growing at a constant rate from year to year.[ii] Appendix G contains a more elaborate description of this model, together with a detailed account of the process used to fit it to the data.

The final result is that the number of catastrophes occurring during a year has approximately the probability distribution shown below.

$$f(x_i) = e^{-m_i}\left(\frac{m_i^{x_i}}{x_i!}\right)$$

(2)

$$m_i = m_n \left(1+R\right)^{(i-n)}$$
(3)

$i \equiv$ The year the distribution applies to.

$x_i \equiv$ The number of catastrophes in year i.

$n \equiv$ 1966

$m_n \equiv$ 14.46

$R \equiv$ 0.02984

A table of this probability distribution for each of the years 1967 to 1976 is shown in Appendix G.

The total loss

Knowing the probability distribution of the size of a catastrophe and the probability distribution of the number of catastrophes, it is possible to calculate the expected total loss for a particular year. If this is done for the years from 1953 to 1966, these theoretical losses then can be compared with the actual losses to determine the probable error, which will be experienced if these distributions are used to forecast future losses. Appendix H contains a detailed analysis of this kind using two related techniques. The result is two probabilistic forecasts of future total losses. The forecast contained in Table 12was obtained by combining the two empirical probability distributions discussed above. The one shown in Tables 13 was obtained by applying two variable linear regression to the log-linear model of total losses suggested by these distributions. Because the mathematical technique used to obtain the figures shown in Table 13 tends to be more efficient than the one used to obtain the figures shown in Table 12, Table 13 is probably a better guide to the future than Table 12.

Appendix A

Ranked observations of yearly catastrophic losses in millions from 1953 to 1966.

TABLE 1

	Ranked Observations of Yearly Catastrophic Losses (millions)													
Size														
Rank	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966
1	1.050	1.100	3.300	1.300	1.400	2.500	1.300	1.100	2.000	1.300	1.300	1.100	1.500	1.000
2	1.700	1.900	4.500	1.600	1.400	2.500	4.700	2.000	2.250	1.800	1.700	1.150	1.500	1.300
3	1.750	2.200	5.100	1.700	2.100	2.500	7.200	2.700	2.250	2.000	1.800	1.400	2.000	1.700
4	1.950	2.750	6.500	3.000	2.600	4.000	7.900	4.000	3.250	2.400	2.400	2.000	2.500	2.500
5	2.350	4.750	6.600	3.700	2.750	4.000	13.000	5.300	4.250	2.600	3.500	2.500	3.000	2.600
6	2.400	7.150	9.500	4.000	2.800	5.000	13.100	5.600	4.500	4.500	5.000	2.700	4.000	2.600
7	3.000	9.250	11.700	4.500	3.700			8.500	6.000	4.500	6.000	2.710	5.000	2.800
8	4.200	12.500	19.000	5.700	4.500			9.800	6.250	6.000	11.000	3.500	6.000	3.800
9	5.400	122.050	25.200	16.900	8.400			91.000	7.000	6.200		3.500	14.000	3.900
10	6.650	129.700		20.000	11.700				7.500	6.300		3.600	30.000	4.200
11	9.350				32.200				11.000	6.500		5.000	38.000	5.000
12	9.400								13.000	8.100		5.000	70.000	5.400
13	11.900								100.000	8.500		7.000	500.000	5.500
14	12.250									9.800		9.500		7.500
15	14.300									17.500		12.000		57.000
16										23.300		15.000		
17										81.000		23.000		
18												30.000		
19												67.200		

Appendix B

The accompanying charts were obtained by using the K’th order statistic as an estimate of the K/(N+1) fractile of each year’s cumulative probability distribution (N being the number of catastrophes which occurred during the year). After plotting these estimates, a smooth curve was drawn through the points, fitted by eye using french curves, to obtain an estimate of the cumulative probability distribution. These are the curves shown in the charts (A complete set of all fourteen charts is available from the authors on request)

TABLE 2

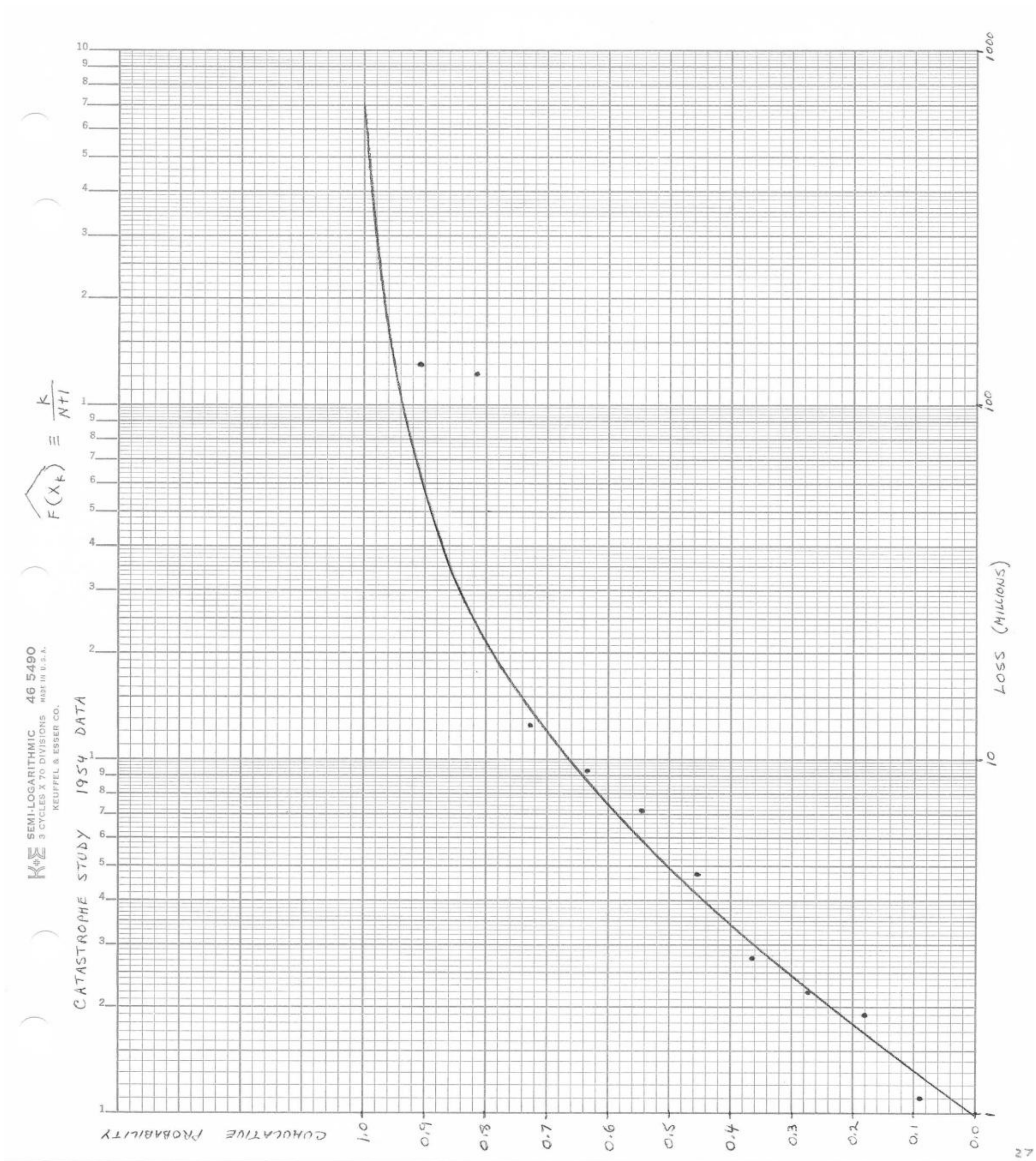
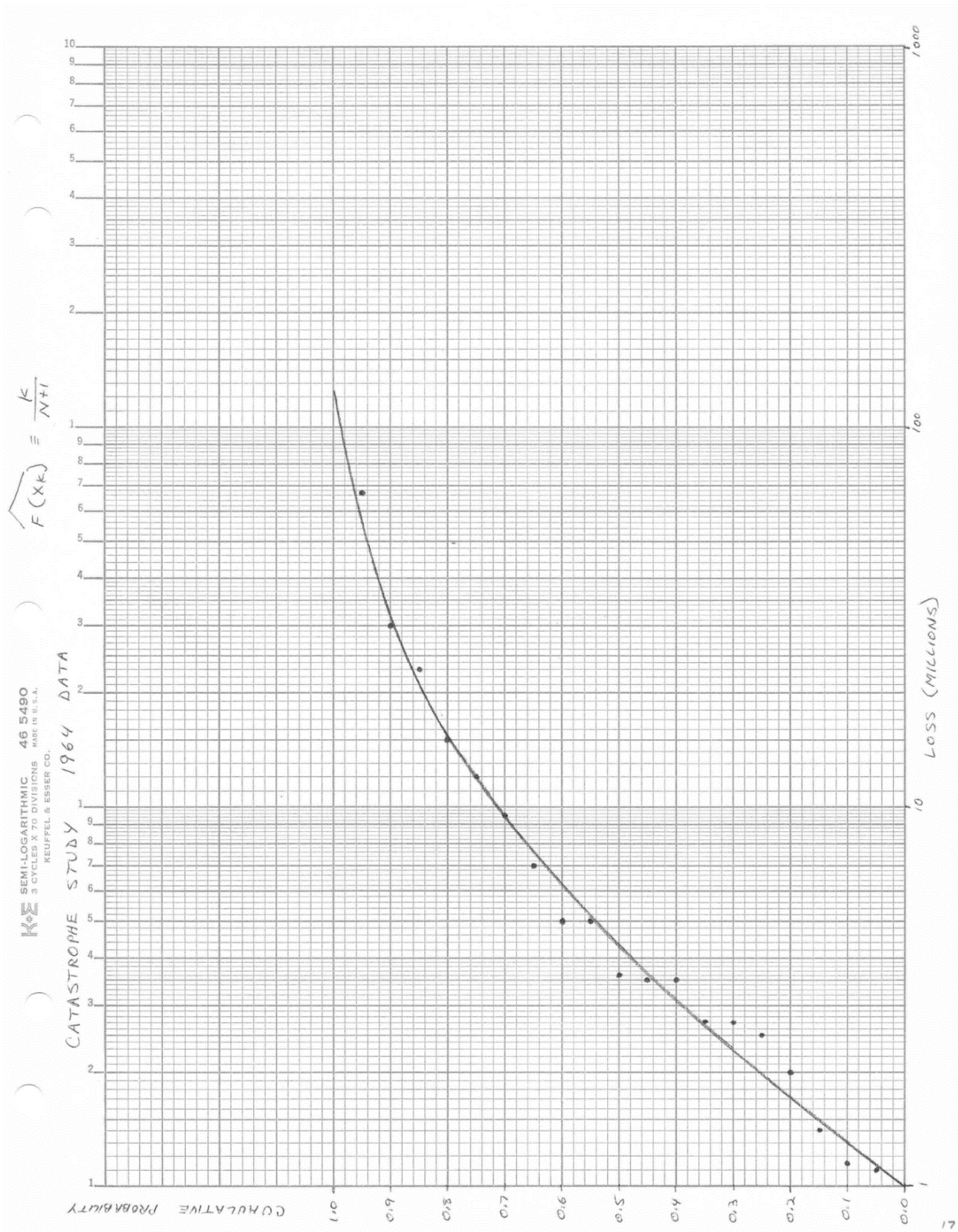
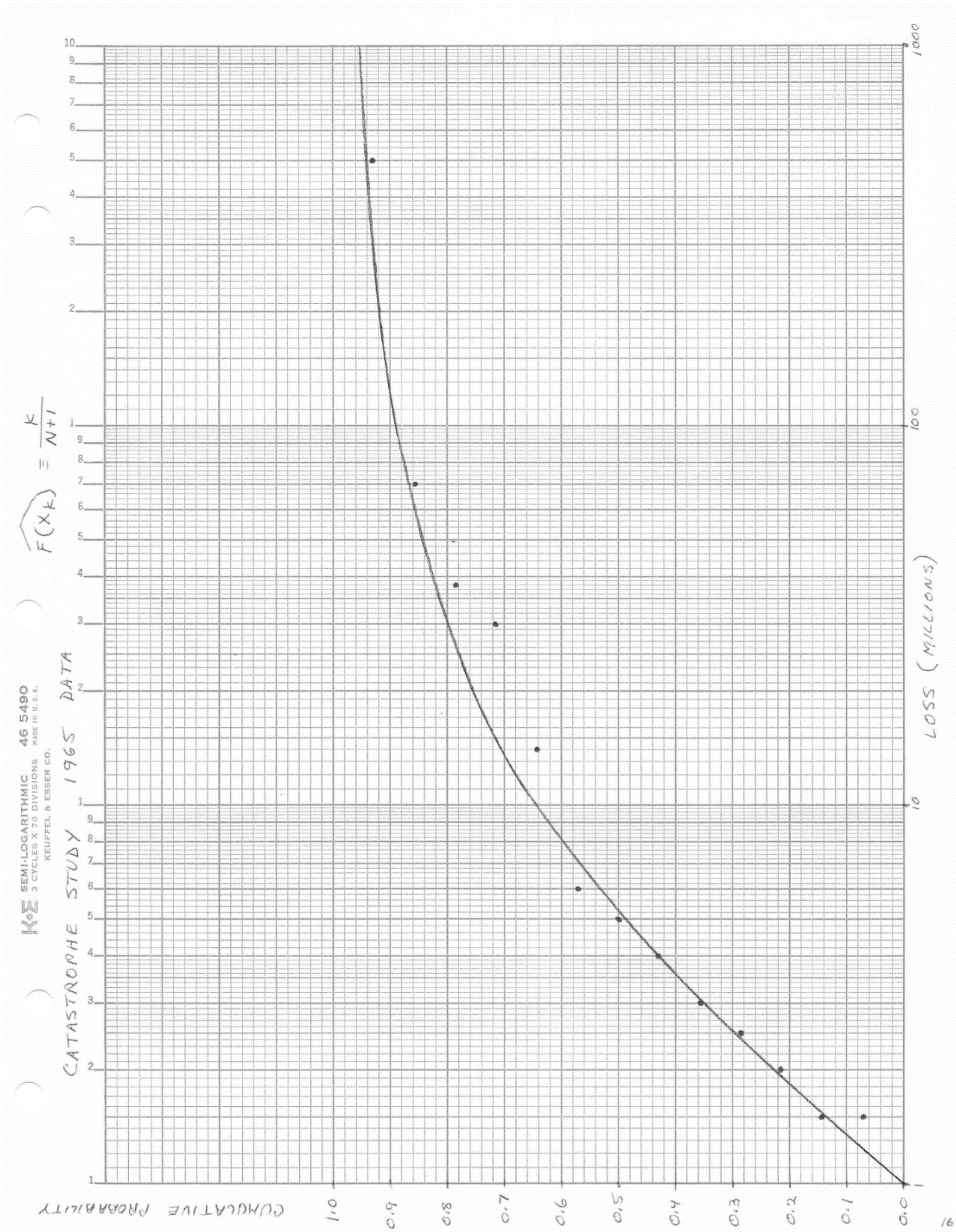


TABLE 3

TABLE 4




K+E SEMI-LOGARITHMIC 46 5490
3 CYCLES X 70 DIVISIONS
KEUFFEL & ESSER CO.
CATASTROPHE STUDY 1965 DATA
F(X_k) = K/(N+1)
LOSS (MILLIONS)
CUMULATIVE PROBABILITY
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1000
100
10

Appendix C

The procedure outlined below is used to test the null hypothesis that two samples come from the same distribution. It is described in detail in section 16.4, page 409 of Introduction to the Theory of Statistics by Mood and Graybill.[iii]

Let $x_i, i = 1, \cdots, n_x$, be a sample from a density $f_x(x)$. Let $y_i, i = 1, \cdots, n_y$, be a sample from density $f_y(y)$. Let the two samples be combined and arranged in order of magnitude. This will result in a sequence of x's and y's. Define a run as a sequence of letters of one kind bounded by letters of the other kind. Let the number of runs be d. Then the probability density of the number of runs, $h(d)$, is as follows.

For d even and $k = \frac{d}{2}$.

$$h(d) = 2\frac{\binom{n_x - 1}{k-1}\binom{n_y - 1}{k-1}}{\binom{n_x + n_y}{n_x}}$$

(4)

For d odd and $k = \frac{d-1}{2}$.

$$h(d) = \frac{\binom{n_x - 1}{k}\binom{n_y - 1}{k-1} + \binom{n_x - 1}{k-1}\binom{n_y - 1}{k}}{\binom{n_x + n_y}{n_x}}$$

(5)

To test the null hypothesis in question with a probability α for the Type I error, find the smallest integer, d_α, such that

$$\sum_{d=0}^{d_\alpha} h(d) \geq \alpha$$

(6)

and reject the null hypothesis if $d \leq d_\alpha$.

If n_x and n_y exceed 10, the distribution of d is approximately normal. The mean and variance of are

$$\mu_d = 2\frac{n_x n_y}{n_x + n_y} + 1$$

(7)

$$\sigma_d^2 = 2\frac{n_x n_y \left(2n_x n_y - n_x - n_y\right)}{\left(n_x + n_y\right)^2 \left(n_x + n_y - 1\right)}$$

(8)

Appendix D

The upper right portion (above the diagonal) of Table 5 shows the number of runs observed for each pair of years. The lower left portion (below the diagonal) of the table shows the number of runs expected under the null hypothesis, rounded to the nearest integer. An asterisk denotes rejection of the null hypothesis that the two samples come from the same probability distribution at the stated level of significance.

TABLE 5

Pairwise Runs Test

Sample	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966
1953	diag	16	10	12	16	7 *	9	18	16	20	15	18	16	17
1954	13	diag	9	9	15	7	9	16	15	17	11	16	14	14
1955	12	10	diag	12	9	6	8	11	13	17	10	13	11	11
1956	13	11	10	diag	12	7	6 *	11	12	11	10	13	14	17
1957	14	11	11	11	diag	7	6 *	13	16	17	12	17	14	15
1958	10	9	8	9	9	diag	5	5 *	7	7 *	7	5 *	5 *	7 *
1959	10	9	8	9	9	7	diag	9	12	9	8	9	6 *	9
1960	12	10	10	10	11	8	8	diag	12	13	13	16	14	16
1961	15	12	12	12	13	9	9	12	diag	14	10	18	13	12
1962	17	14	13	14	14	10	10	13	16	diag	16	22	16	14
1963	11	10	9	10	10	8	8	9	11	12	diag	15	14	15
1964	18	14	13	14	15	10	10	13	16	19	12	diag	20	18
1965	15	12	12	12	13	9	9	12	14	16	11	16	diag	18
1966	16	13	12	13	14	10	10	12	15	17	11	18	15	diag

Notes

1	Null hypothesis: Probability distribution of the size of a catastrophe is identical across years.
2	The upper diagonal entries are the observed number of runs.
3	The lower diagonal entries are the expected number of runs.
4	Asterisks denote significance at the 5% level.

Appendix E

Because the runs test described in Appendix C was applied repeatedly, the occurrence of some asterisks in the table of Appendix D is to be expected. Under the null hypothesis, the probability of obtaining exactly k asterisks in n applications of this test at a significance level of α is:

$$\binom{n}{k}\alpha^k \left(1-\alpha\right)^{n-k}$$

(9)

Thus, the probability of obtaining at least m asterisks in n such applications is:

$$\sum_{k=m}^{n} \binom{n}{k} \alpha^{k} (1-\alpha)^{n-k}$$
(10)

The figures in Table 6A were obtained by applying Equation (10) to the data in Appendix D and suggests that the observed number of asterisks observed there is unusually high. However, an examination of the table in Appendix D reveals that a substantial proportion of the asterisks observed are related to the year 1958. (This was done for five other significance levels and a similar result was observed. These tables are available from the authors on request.) To check the premise that only the 1958 data are abnormal, Equation (10) was applied to the remainder of the data. The results are shown in Table 6B. They are entirely consistent with the hypothesis that the probability distribution of the size of a catastrophe is constant over time.

Ordinarily, it is not proper to eliminate data selectively when conducting a test of this kind. In this case, it is justified on the basis of a priori knowledge. Because of the nature of catastrophic events, any change in the probability distribution of the size of a catastrophe would be expected to take the form of a relatively smooth long-term trend. If this were true, the density of the asterisks in Table 5 of Appendix D would be greatest in the upper right hand corner, where samples from widely separated years are compared. This is not the case. Furthermore, a mechanism for changing the probability distribution of the size of a catastrophe suddenly and for only one year is diffcult to imagine. For these reasons, it is fair to interpret the existence of one apparently abnormal year in the middle of the time period as coincidental.

TABLE 6

Table 6A

Significance Level of Each Test (alpha)	Probability of Obtaining at Least the Observed Number of Asterisks	Observed Number of Asterisks	Number of Applications of the Test	Percentage of Applications With Asterisks
0.05	0.038	9	91	0.0989

Table 6B

Significance Level of Each Test (alpha)	Probability of Obtaining at Least the Observed Number of Asterisks	Observed Number of Asterisks	Number of Applications of the Test	Percentage of Applications With Asterisks
0.05	0.754	3	78	0.0385

Appendix F

Table 7 contains a record of the catastrophic losses experienced from 1953 to 1966 arranged in order of increasing size. An estimate of the corresponding fractile of the cumulative probability distribution, obtained by using the k'th order statistic as a proxy for the k/(n+1) fractile, also is shown. This data is plotted in Table 8.

To see if the data could be represented adequately by a normal or log-normal probability distribution, it also was plotted on both arithmetic and logarithmic probability paper. This showed that the descriptive ability of these distributions was poor. Next, the gamma distribution was selected for investigation. The reasons for this choice were that the range of definition for the size of a catastrophe easily could be made to correspond to that of the distribution and the variety of shapes which the function can assume. The gamma distribution was fit to the logarithms of the loss figures, expressed in millions.

Initial estimates of the parameters were obtained by using a modified version of the method of moments. Denoting the distribution by

$$f(y) = \frac{y^{\alpha} e^{-\left(\frac{y}{\beta}\right)}}{\Gamma(\alpha+1)\beta^{(\alpha+1)}}$$

(11)

it can be shown that

$$\mu = \beta(\alpha+1)$$

(12)

$$\sigma^2 = \beta^2(\alpha+1)$$

(13)

The mean of the sample was used as a proxy for μ and an unbiased estimate of the variance was used in place of σ^2. The resulting estimates were:

$$\alpha = 1.292656$$
$$\beta = 0.7243593$$

(14)

As a check of the goodness of fit, the Pearson χ^2 test was applied.[iv] The range of definition of the estimated gamma distribution was broken down into 9 intervals, each with an expected number of occurrences of at least 15. Then the χ^2 quantity

$$\chi^2 = \sum_{i=1}^{9} \frac{(v_i - np_i)^2}{np_i}$$

(15)

was calculated, where v_i is the actual number of occurrences in the i'th interval and np_i is the expected number of occurrences in the i'th interval. This quantity has approximately a χ^2 distribution with 6 degrees of freedom when the parameters are fit by minimizing χ^2. The test is conservative in other cases. For the parameters shown above, the χ^2 value is approximately 7.927. This is significant at the 25% level. Clearly, the fit is good.

The parameters were then varied in a trial and error fashion in order to find values which minimized χ^2. Within the limits of attainable precision, the following estimates seemed best.

$$\alpha = 1.3765$$
$$\beta = 0.6796$$

(16)

The χ^2 value for these parameters is approximately 7.650. This is a significant at the 27% level. The corresponding cumulative distribution is plotted in Table 9. Its slope is slightly more in tune with Table 8.

The final result is that the size of a catastrophe is distributed approximately as follows.

$$f(y) = \frac{y^{\alpha} e^{-\left(\frac{y}{\beta}\right)}}{\Gamma(\alpha+1)\beta^{(\alpha+1)}} \qquad y \ge 0$$

(17)

$y \equiv$ $Ln(x)$

$x \equiv$ Insured loss (millions)

$\alpha =$ 1.3765

$\beta =$ 0.6796

TABLE 7

Ranked Loss and Estimated Fractile of the Probability Distribution of a Catastrophe								
Size Rank	Loss (millions)	Estimated Fractile	Size Rank	Loss (millions)	Estimated Fractile	Size Rank	Loss (millions)	Estimated Fractile
1	1.00	0.006173	55	2.80	0.339506	109	6.65	0.672840
2	1.05	0.012346	56	2.80	0.345679	110	7.00	0.679012
3	1.10	0.018519	57	3.00	0.351852	111	7.00	0.685185
4	1.10	0.024691	58	3.00	0.358025	112	7.15	0.691358
5	1.10	0.030864	59	3.00	0.364198	113	7.20	0.697531
6	1.15	0.037037	60	3.25	0.370370	114	7.50	0.703704
7	1.30	0.043210	61	3.30	0.376543	115	7.50	0.709877
8	1.30	0.049383	62	3.50	0.382716	116	7.90	0.716049
9	1.30	0.055556	63	3.50	0.388889	117	8.10	0.722222
10	1.30	0.061728	64	3.50	0.395062	118	8.40	0.728395
11	1.30	0.067901	65	3.60	0.401235	119	8.50	0.734568
12	1.40	0.074074	66	3.70	0.407407	120	8.50	0.740741
13	1.40	0.080247	67	3.70	0.413580	121	9.25	0.746914
14	1.40	0.086420	68	3.80	0.419753	122	9.35	0.753086
15	1.50	0.092593	69	3.90	0.425926	123	9.40	0.759259
16	1.50	0.098765	70	4.00	0.432099	124	9.50	0.765432
17	1.60	0.104938	71	4.00	0.438272	125	9.50	0.771605
18	1.70	0.111111	72	4.00	0.444444	126	9.80	0.777778
19	1.70	0.117284	73	4.00	0.450617	127	9.80	0.783951
20	1.70	0.123457	74	4.00	0.456790	128	11.00	0.790123
21	1.70	0.129630	75	4.20	0.462963	129	11.00	0.796296
22	1.75	0.135802	76	4.20	0.469136	130	11.70	0.802469
23	1.80	0.141975	77	4.25	0.475309	131	11.70	0.808642
24	1.80	0.148148	78	4.50	0.481481	132	11.90	0.814815
25	1.90	0.154321	79	4.50	0.487654	133	12.00	0.820988
26	1.95	0.160494	80	4.50	0.493827	134	12.25	0.827160
27	2.00	0.166667	81	4.50	0.500000	135	12.50	0.833333
28	2.00	0.172840	82	4.50	0.506173	136	13.00	0.839506
29	2.00	0.179012	83	4.50	0.512346	137	13.00	0.845679
30	2.00	0.185185	84	4.70	0.518519	138	13.10	0.851852
31	2.00	0.191358	85	4.75	0.524691	139	14.00	0.858025
32	2.10	0.197531	86	5.00	0.530864	140	14.30	0.864198
33	2.20	0.203704	87	5.00	0.537037	141	15.00	0.870370
34	2.25	0.209877	88	5.00	0.543210	142	16.90	0.876543
35	2.25	0.216049	89	5.00	0.549383	143	17.50	0.882716
36	2.35	0.222222	90	5.00	0.555556	144	19.00	0.888889
37	2.40	0.228395	91	5.00	0.561728	145	20.00	0.895062
38	2.40	0.234568	92	5.10	0.567901	146	23.00	0.901235
39	2.40	0.240741	93	5.30	0.574074	147	23.30	0.907407
40	2.50	0.246914	94	5.40	0.580247	148	25.20	0.913580
41	2.50	0.253086	95	5.40	0.586420	149	30.00	0.919753
42	2.50	0.259259	96	5.50	0.592593	150	30.00	0.925926
43	2.50	0.265432	97	5.60	0.598765	151	32.20	0.932099
44	2.50	0.271605	98	5.70	0.604938	152	38.00	0.938272
45	2.50	0.277778	99	6.00	0.611111	153	57.00	0.944444
46	2.60	0.283951	100	6.00	0.617284	154	67.20	0.950617
47	2.60	0.290123	101	6.00	0.623457	155	70.00	0.956790
48	2.60	0.296296	102	6.00	0.629630	156	81.00	0.962963
49	2.60	0.302469	103	6.20	0.635802	157	91.00	0.969136
50	2.70	0.308642	104	6.25	0.641975	158	100.00	0.975309
51	2.70	0.314815	105	6.30	0.648148	159	122.05	0.981481
52	2.71	0.320988	106	6.50	0.654321	160	129.70	0.987654
53	2.75	0.327160	107	6.50	0.660494	161	500.00	0.993827
54	2.75	0.333333	108	6.60	0.666667			

TABLE 8

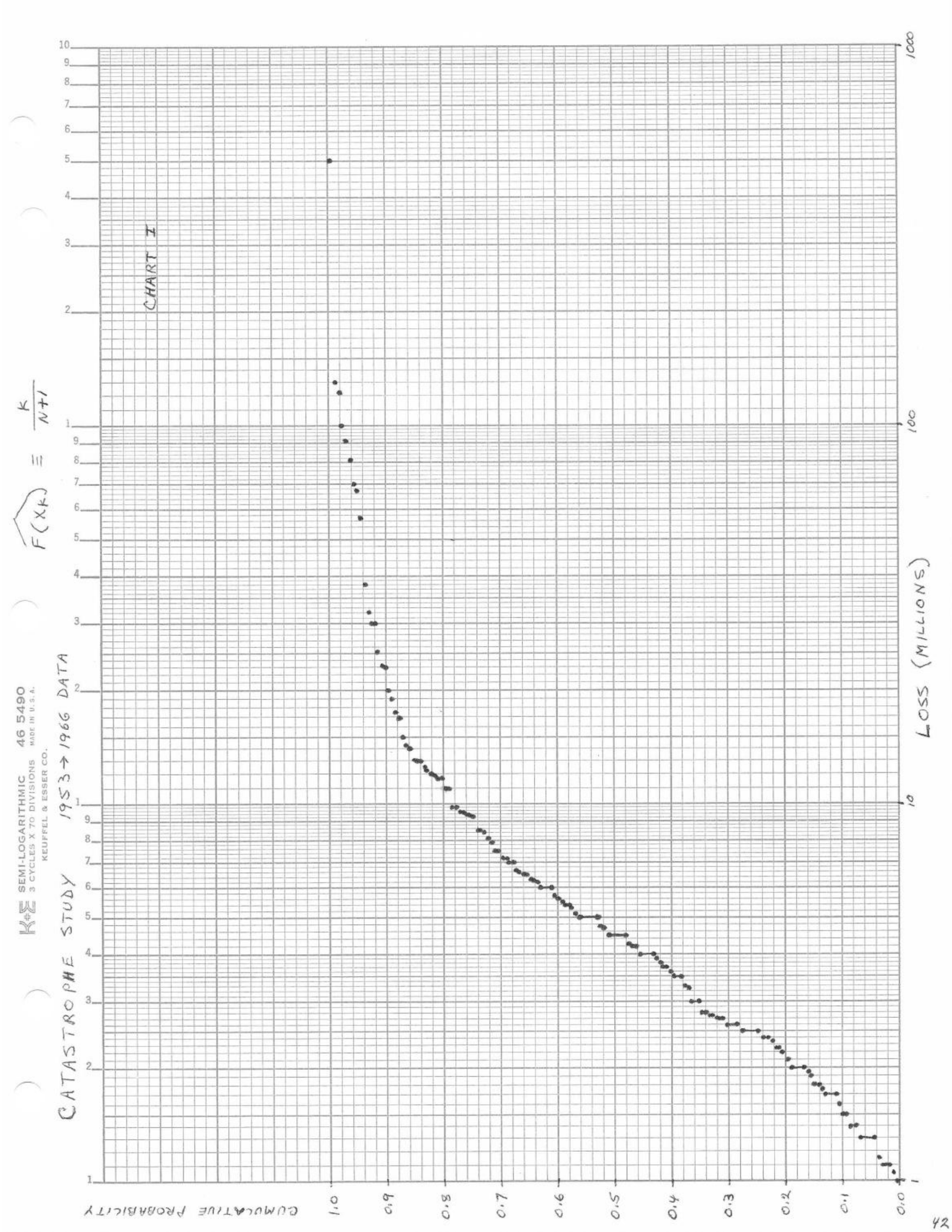
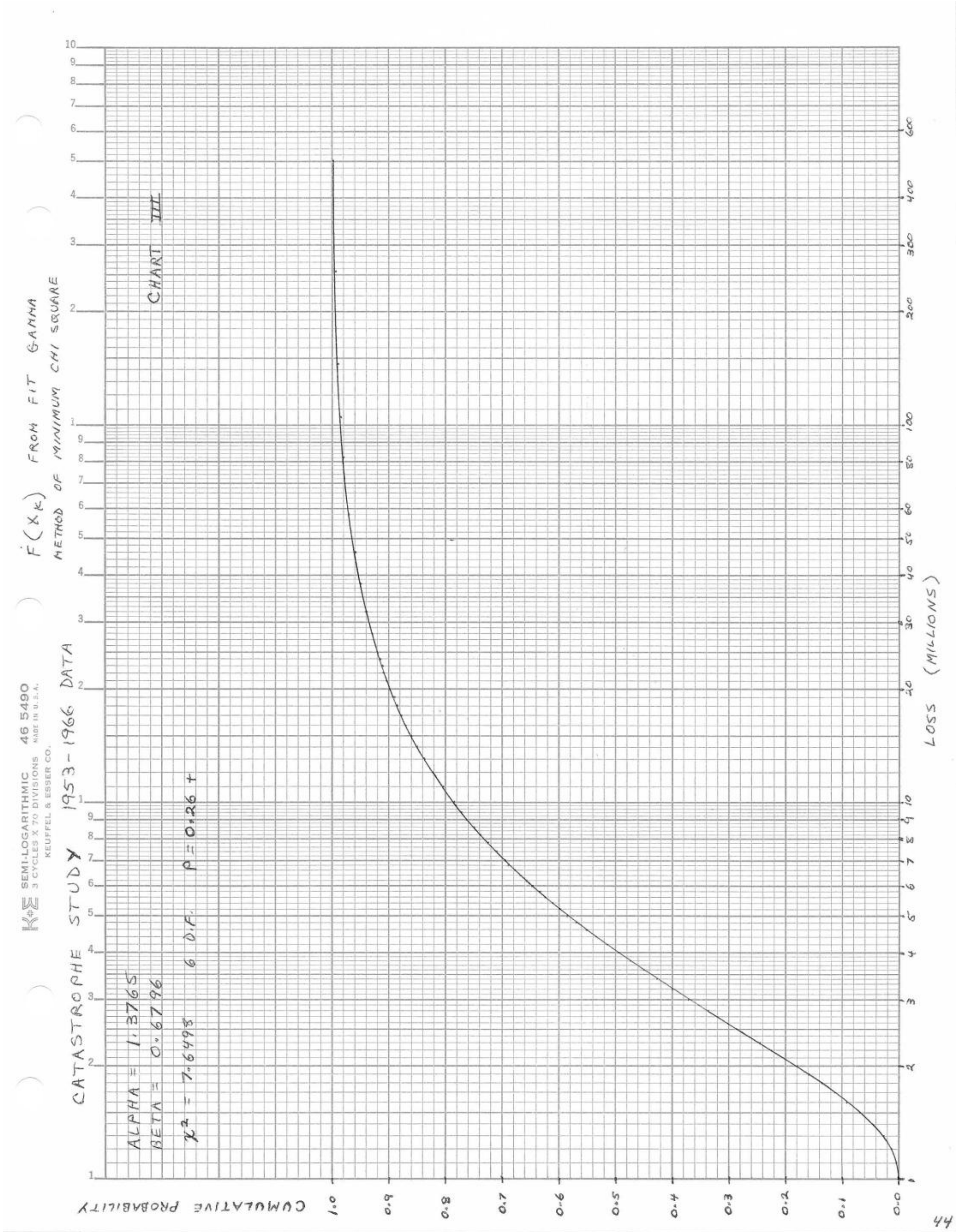


CHART I
CATASTROPHE STUDY
1953 → 1966 DATA
LOSS (MILLIONS)
CUMULATIVE PROBABILITY
42

Doi: 10.48150/jbssr.v3no8.2022.a1

TABLE 9



Appendix G

The number of catastrophes which occur during a year was assumed to have a Poisson distribution with the parameter growing at a constant rate over time. Denoting the number of catastrophes in the i'th year by x_i, the probability distribution of x_i by $f_i(x_i)$, and the Poisson parameter by m_i, this can be written as:

$$f_i(x_i) = \left(\frac{m_i^{x_i}}{x_i!}\right) e^{-m_i}$$

(18)

$x_i \equiv$ 0, 1, 2, 3, …

$\boldsymbol{i} \equiv$ 1953, 1954, . . . , 1966, where i=1 represents the year 1953 and i=14 represents the year 1966.

$m_i \equiv m_{1966}(1+R)^{(i-1966)}$

$m_{1966} =$ 14.46

$\boldsymbol{R} =$ 0.02984

The above model was fit to the data in Appendix A using the method of minimum χ^2.[v] This method consists of choosing parameters which minimize the quantity

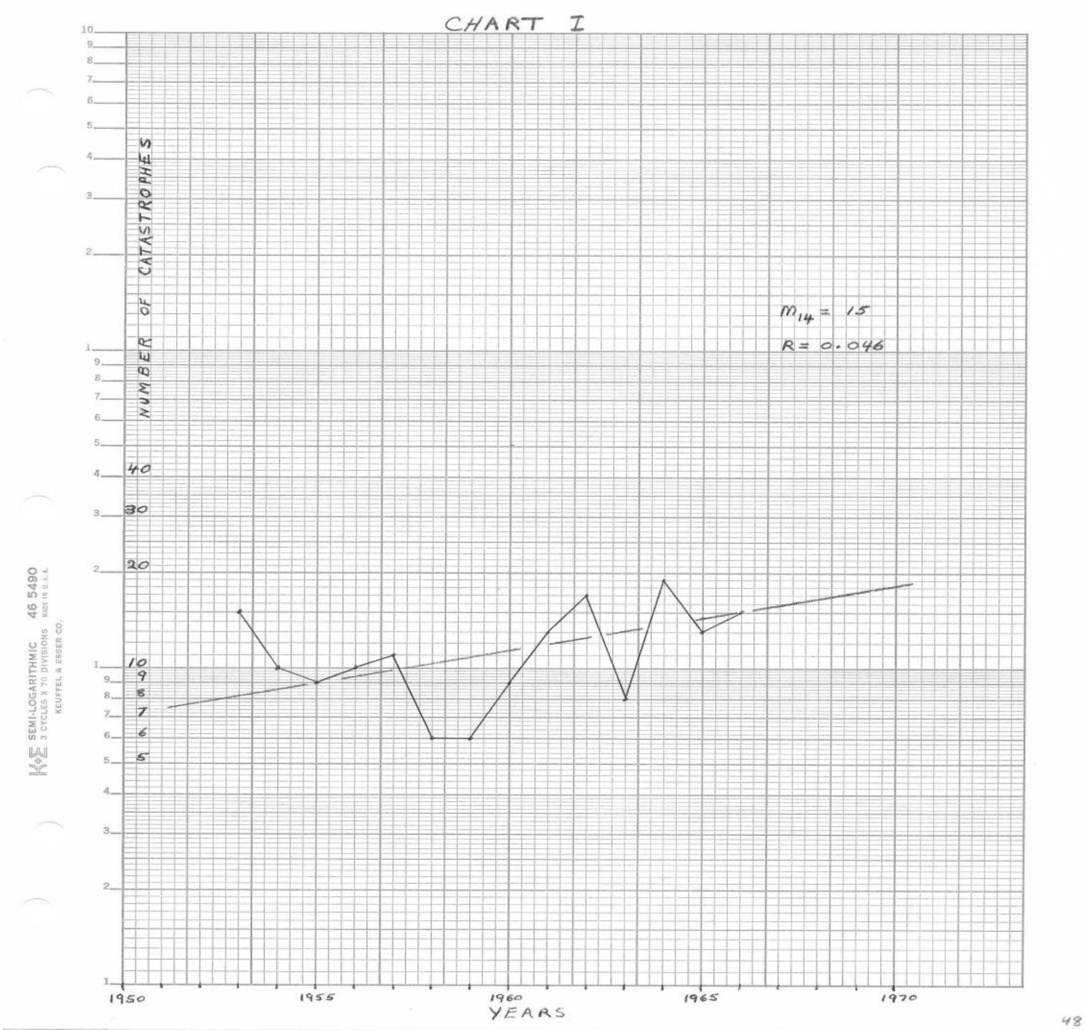
$$\chi^2 = \sum_{i=1}^{L} \frac{(v_i - np_i)^2}{np_i}$$

where v_i is the actual number of occurrences in the i'th class and np_i is the expected number of occurrences in the i'th class. This quantity has approximately a χ^2 distribution with (L-1-Q) degrees of freedom when Q parameters are fit from the data by minimizing χ^2. In this case, L=14, because each year was treated as a class.[vi] With this scheme, np_i is the expected number of events in the i'th year. This is the expected value of the i'th year's Poisson distribution, which is m_i. Substituting these values into the formula results in

$$\chi^2 = \sum_{i=1}^{14} \frac{\left(v_i - m_{1966}(1+R)^{(i-1966)}\right)^2}{m_{1966}(1+R)^{(i-1966)}}$$

(19)

Because both m_{14} and R are being varied in order to minimize χ^2, Q=2 and χ^2 will have



approximately a χ^2 distribution with 11 degrees of freedom.

Initial estimates for the minimizing values of the parameters (m_{1966} and R) were obtained by plotting the number of catastrophes on semi-logarithmic paper and fitting a trend line by eye. Table 10 shows the plotted data and the subjectively drawn trend line. The parameter values corresponding to this line are $m_{1966} = 15$ and $R = 0.046$. For these parameters, the χ^2 value is approximately 16.08. This corresponds to a significance level of about 13.8%.
TABLE 10

The parameters weres then varied in a trial and error fashion in order to find values that minimized χ^2. Within the limits of attainable precision, the following estimates seemed best.

$m_{1966} = 14.46$
(20)
$R = 0.02984$
(21)

The χ^2 value for these parameters is approximately 14.79. This is significant at the 19.2% level.

The final result is that the number of catastrophes that occur during a year has approximately a Poisson distribution with the parameter growing at a constant rate. In algebraic terms:

$$f_i(x_i) = \left(\frac{m_i^{x_i}}{x_i!} \right) e^{-m_i}$$
(22)

$x_i \equiv$ $0, 1, 2, 3, \cdots$

$i \equiv$ 1953, 1954, . . ., 1966, where i=1 represents the year 1953 and i=14 represents the year 1966.

$m_i \equiv$ $m_{1966} (1+R)^{(i-1966)}$

$m_{1966} =$ 14.46

$R =$ 0.02984

Table 11 contains a tabulation of this estimating function for the years 1967 and 1976.

(The results for the years 1967 through 1976 is available from the authors on request.)

TABLE 11

Year:		**1967**		**Year:**		**1976**
Expected Number of Catastrophes:		**14.89**		**Expected Number of Catastrophes:**		**19.40**
Number		**Estimated**		**Number**		**Estimated**
Of	**Estimated**	**Cumulative**		**Of**	**Estimated**	**Cumulative**
Catastrophes	**Probability**	**Probability**		**Catastrophes**	**Probability**	**Probability**
0	0.00000	0.00000		0	0.00000	0.00000
1	0.00001	0.00001		1	0.00000	0.00000
2	0.00004	0.00004		2	0.00000	0.00000
3	0.00019	0.00023		3	0.00000	0.00001
4	0.00070	0.00093		4	0.00002	0.00003
5	0.00208	0.00301		5	0.00009	0.00011
6	0.00516	0.00817		6	0.00028	0.00039
7	0.01099	0.01916		7	0.00077	0.00116
8	0.02045	0.03961		8	0.00187	0.00303
9	0.03384	0.07345		9	0.00402	0.00705
10	0.05039	0.12384		10	0.00780	0.01485
11	0.06821	0.19205		11	0.01377	0.02862
12	0.08465	0.27670		12	0.02226	0.05088
13	0.09697	0.37367		13	0.03322	0.08410
14	0.10314	0.47681		14	0.04604	0.13014
15	0.10240	0.57920		15	0.05956	0.18970
16	0.09530	0.67451		16	0.07222	0.26192
17	0.08348	0.75799		17	0.08243	0.34435
18	0.06906	0.82705		18	0.08886	0.43321
19	0.05413	0.88118		19	0.09074	0.52395
20	0.04030	0.92148		20	0.08803	0.61198
21	0.02858	0.95006		21	0.08134	0.69332
22	0.01935	0.96941		22	0.07173	0.76505
23	0.01253	0.98194		23	0.06052	0.82556
24	0.00777	0.98971		24	0.04892	0.87449
25	0.00463	0.99434		25	0.03797	0.91246
26	0.00265	0.99699		26	0.02834	0.94079
27	0.00146	0.99845		27	0.02036	0.96116
28	0.00078	0.99923		28	0.01411	0.97527
29	0.00040	0.99963		29	0.00944	0.98471
30	0.00020	0.99983		30	0.00611	0.99081

Appendix H

The probability distribution of the size of a catastrophe and the probability distribution of the number of catastrophes determines the probability distribution of total catastrophic loss. In this case, however, only estimates of the first two distributions are available. If these estimates are used in place of the true distributions, the result will not reflect errors arising from the approximate nature of the empirical distributions. Because these errors are important, it seems worthwhile to allow for them when making a prediction. In order to accomplish this, the empirical distributions were used to obtain an estimate of the expected total loss for each of the years from 1953 to 1966.

These estimates then were compared to the actual figures to obtain an estimate of the probable error that will be experienced if the empirical distributions are used to predict future losses. The method of comparison consisted of treating the estimated expected total loss as if it were the result of a two-variable linear regression calculation with time as the independent variable.[vii]

Given the estimated distribution of the size of a catastrophe, $h(y)$, and the estimated distribution of the number of catastrophes, $g(x)$, for a particular year, the estimated expected total loss, L, is found from

$$L=\sum_{x=0}^{\infty} g(x)\left\{\left[\prod_{i=1}^{x}\int_{1}^{\infty}\right]\left[\sum_{i=1}^{x} y_i\right]\left[\prod_{i=1}^{x} h(y_i)\right]\left[\prod_{i=1}^{x} dy_i\right]\right\}$$
(23)

In this expression, g(x) is a Poisson distribution and h(y) is derived from a Gamma distribution. The expression can be simplified by noting that

$$\left\{\left[\prod_{i=1}^{x}\int_{1}^{\infty}\right]\left[\sum_{i=1}^{x} y_i\right]\left[\prod_{i=1}^{x} h(y_i)\right]\left[\prod_{i=1}^{x} dy_i\right]\right\}=xE(y)$$
(24)

This is because

$$\int_{1}^{\infty} y_i h(y_j)\,dy_j=(1-\delta_{ij})\,y_i+\delta_{ij}E(y)$$
(25)

Thus,

$$L=\sum_{x=0}^{\infty} xE(y)\,g(x)=E(x)E(y)$$
(26)

All that remains is to evaluate E(x) and E(y). E(x) is the expected value of the particular year's Poisson distribution. This is simply the Poisson parameter, m_i. E(y) is the estimated expected value of the size of a catastrophe and can be found as outlined below.

In Appendix F, it is concluded that the logarithm of the size of a catastrophe has approximately a gamma distribution. Denoting this quantity by y,

$$f(y)=\frac{y^{\alpha}e^{-\left(\frac{y}{\beta}\right)}}{\Gamma(\alpha+1)\beta^{(\alpha+1)}} \qquad ,y\geq 0$$
(27)

In order to find the distribution of the size of a catastrophe, the following transformation must be made.

$$y = \ln(x)$$
(28)

Denoting the distribution of x by g(x), it is clear that

$$g(x) = f\left[y(x)\left(\frac{dy}{dx}\right)\right]$$
(29)

since this transformation has a single valued inverse. The calculations are carried out below.

$$g(x) = \frac{[\ln(x)]^{\alpha} e^{-\left(\frac{\ln(x)}{\beta}\right)}}{\Gamma(\alpha+1)\beta^{(\alpha+1)}x} \qquad , x \geq 1$$
(30)

It is the expected value of this distribution that is required.

$$E(x) = \int_{1}^{\infty} xg(x)\,dx$$
(31)

$$E(x) = \int_{1}^{\infty} \frac{[\ln(x)]^{\alpha} e^{-\left(\frac{\ln(x)}{\beta}\right)}}{\Gamma(\alpha+1)\beta^{(\alpha+1)}}\,dx$$
(32)

To evaluate this integral, use the transformation

$$y = \ln(x)$$
(33)
Then,

$$\int_{1}^{\infty} xg(x)\,dx = \int_{0}^{\infty} x(y)\,g(x(y))\left(\frac{dx}{dy}\right)dy$$
(34)

$$E(x) = \int_{0}^{\infty} \frac{y^{\alpha} e^{-\frac{y}{\beta}}}{\Gamma(\alpha+1)\beta^{\alpha+1}} e^{y}\,dy$$
(35)

$$E(x)=\int_0^\infty \frac{y^\alpha e^{-\frac{y}{\left(\frac{\beta}{1-\beta}\right)}}}{\Gamma(\alpha+1)\beta^{\alpha+1}}dy$$

(36)

$$E(x)=\frac{1}{(1-\beta)^{\alpha+1}}\int_0^\infty \frac{y^\alpha e^{-\frac{y}{\left(\frac{\beta}{1-\beta}\right)}}}{\Gamma(\alpha+1)\left(\frac{\beta}{1-\beta}\right)^{\alpha+1}}dy$$

(37)

The integrand in Equation (37) is a gamma distribution, so

$$E(x)=\frac{1}{(1-\beta)^{\alpha+1}}$$

(38)

This is the estimated expected value of the size of a catastrophe. Substituting this result in Equation (26) provides the estimated expected total loss for a particular year.

$$L_i=\frac{m_i}{(1-\beta)^{\alpha+1}}$$

(39)

$$L_i=\frac{m_N(1+R)^{-(N-i)}}{(1-\beta)^{\alpha+1}}$$

(40)

$$L_i=\left[\frac{m_N(1+R)^{-N}}{(1-\beta)^{\alpha+1}}\right](1+R)^i$$

(41)

$$L_i=L_0(1+R)^i$$

(42)

In making probabilistic predictions, the logarithms of the above estimates, L_i, were treated as if they were the result of a two variable linear regression calculation with time as the independent variable. This treatment was suggested by the linear form of the logarithmic relationship and the least squares character of the method of minimum chi square. All the predictions for a particular year consist of three numbers. One represents the best estimate and is simply L_k for the k'th year. The other two numbers define an estimated confidence interval for the actual loss. They are obtained from the following equations.

 Doi: 10.48150/jbssr.v3no8.2022.a1

The upper and lower limits of the confidence interval are

$$e^{\left[\ln\left(L_k\right)\pm t_{\frac{\varepsilon}{2}}\hat{\sigma}\sqrt{1+\frac{1}{N}+\frac{\left(y_k-\bar{y}\right)^2}{\sum_{i=1}^{N}\left(y_i-\bar{y}\right)^2}}\right]}$$

(43)

$$\hat{\sigma}=\sqrt{\left(\frac{1}{N-3}\right)\sum_{i=1}^{N}\left[\ln\left(L_i\right)-\ln\left(L_{Ai}\right)\right]^2}$$

(44)

$$\bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i$$

(45)

$y_i \equiv$ Year i.

$t_{\frac{\varepsilon}{2}} \equiv$ The t value that cuts off $\frac{\varepsilon}{2}$ of the right tail of a t distribution with (N-3) degrees of freedom.

$L_i \equiv$ The estimated total loss in year i.

$L_{Ai} \equiv$ The actual total loss in year i.

Tables I contains the results of applying these equations to the estimated and actual losses for the years 1953 to 1966.

Equation (42) implies that the logarithmic form for estimating total loss is linear.

$$\ln\left(L_i\right)=\ln\left(L_0\right)+\left[\ln\left(1+R\right)\right]i$$

(46)

This suggests using a two variable linear regression as an alternative method of estimating future losses. Tables II reflects this approach.

The difference between Tables I and II is striking. The estimated total losses in Table IIare considerably more consistent with actual total losses than those in the Table I. In addition, the confidence intervals in Table II are much narrower. The reason for these differences is that Table II reflects a more efficient estimation procedure for the parameters of the final log-linear model for total losses. Thus, Table IIis more reliable than Table I.

		Table I		
	90% Confidence Interval			
	Actual	Estimated		
	Total	Total	Lower	Upper
Year	Loss	Loss	Limit	Limit
1953	87.65	147.53	15.19	1432.60
1954	293.35	151.93	16.42	1405.92
1955	91.40	156.47	17.62	1389.74
1956	62.40	161.14	18.76	1384.27
1957	73.55	165.94	19.81	1389.89
1958	20.50	170.90	20.76	1407.14
1959	47.20	176.00	21.56	1436.73
1960	130.00	181.25	22.20	1479.61
1961	169.25	186.66	22.67	1536.90
1962	192.30	192.23	22.95	1610.01
1963	32.70	197.96	23.04	1700.63
1964	197.86	203.87	22.95	1810.76
1965	677.50	209.95	22.69	1942.80
1966	106.80	216.22	22.27	2099.57
1967	NA	222.67	21.70	2284.22
1968	NA	229.31	21.02	2501.25
1969	NA	236.16	20.25	2754.70
1970	NA	243.20	19.39	3050.15
1971	NA	250.46	18.48	3393.96
1972	NA	257.93	17.54	3793.53
1973	NA	265.63	16.57	4257.57
1974	NA	273.56	15.60	4796.21
1975	NA	281.72	14.64	5421.34

		Table II		
	90% Confidence Interval			
	Actual	Estimated		
	Total	Total	Lower	Upper
Year	Loss	Loss	Limit	Limit
1953	87.65	72.28	11.24	464.80
1954	293.35	76.59	12.39	473.50
1955	91.40	81.16	13.58	485.20
1956	62.40	86.00	14.78	500.30
1957	73.55	91.13	15.99	519.20
1958	20.50	96.56	17.19	542.50
1959	47.20	102.30	18.34	570.80
1960	130.00	108.40	19.43	604.90
1961	169.25	114.90	20.45	645.50
1962	192.30	121.70	21.37	693.60
1963	32.70	129.00	22.18	750.40
1964	197.86	136.70	22.86	817.10
1965	677.50	144.80	23.43	895.40
1966	106.80	153.50	23.87	987.00
1967	NA	162.60	24.18	1094.00
1968	NA	172.30	24.36	1219.00
1969	NA	182.60	24.44	1365.00
1970	NA	193.50	24.40	1534.00
1971	NA	205.00	24.27	1732.00
1972	NA	217.30	24.05	1963.00
1973	NA	230.20	23.75	2231.00
1974	NA	243.90	23.38	2545.00
1975	NA	258.50	22.96	2910.00

Endnotes

[i] This was the definition the insurance industry used at the time. The industry supplied the data used in the analysis. (Note that the current definition of a catastrophe has increased to $25.0 million.)

[ii] Assuming a growing Poisson parameter for the number of catastrophes and a constant probability for the size of a catastrophe seems inconsistent. This was done to simplify the computations so that they could be run more easily on the then available computers.

[iii] Mood, Alexander and Franklin Graybill. Introduction to the Theory of Statistics. McGraw-Hill Book Company. Second Edition, 1963.

[iv] This technique is described in sections 30.1 – 30.3 of Cramer, Harald. Mathematical Methods of Statistics. Princeton University Press. 1946.

[v] Cramer describes this technique in sections 30.1 to 40.3 of Mathematical Methods of Statistics.

[vi] Professor John Rolph of Columbia University suggested this method of classification. It is not in accordance with the classification scheme outlined in Cramer, but does reflect the intent of that material. Clearly, a low χ^2 value will be achieved only if the model is consistent with the data.

[vii] Professor John Rolph of Columbia University suggested this method of analysis. It is an approximate treatment which should give reasonable results.