Volume LXXXI, Part 1 **No. 154**

PROCEEDINGS
May 15, 16, 17, 18, 1994

AN ACTUARIAL APPROACH TO PROPERTY CATASTROPHE COVER RATING

DANIEL F. GOGOL

Abstract

Forty-one years of catastrophe loss data by state are used in this study to produce a model for rating catastrophe covers for insurers in any region of the continental United States. Smooth surfaces are fitted to the data by region, and experience rating is applied in an attempt to give appropriate weight to regional departures from the smoothed results. Severity distributions and frequencies are estimated for each region, and a method for applying them in pricing catastrophe covers is discussed. A method for using the experience of an insurer to produce an experience modification is also presented.

ACKNOWLEDGMENTS

The author would like to thank Bruce Baumgarten, who introduced him to this subject, Margaret O'Brien and Sheldon Cohen, who helped with the computer work, and Robert Bear, who made very useful comments on an earlier draft.

1. INTRODUCTION

United States catastrophe cover rating is an interesting problem from both practical and theoretical points of view.

On the practical side, it is an important untreated problem. No systematic attempt at using insurance loss data to produce catastrophe cover rates can be found in insurance literature. Discussions of methods involving weather data are in Clark [7] and Friedman [9]. Catastrophe rates fluctuate greatly in the various regions of the country depending on the supply of capacity and whether there has been a large catastrophe in the area recently. Pricing practices were not much different two decades ago when Ingrey [12] stated:

> The general yardstick is the "payback period," or, in how many years will a total loss be amortized in advance. Payback periods depend upon location, type of business written, and past experience in addition to the basic ingredients of amount of capacity required, subject premium, and rate. The adequacy of the initial retention is largely overlooked as are the incremental functions of exposure types; to wit, a company writing mobile homes has a much greater incremental exposure function than another insurer writing private dwellings.

Catastrophe rating is also a challenging theoretical problem. The number of large catastrophes in any region is small, so it is important to use the experience of surrounding areas as well. It is useful to examine the relationship between catastrophe experience and a region's longitude, latitude, and distance from the coast. Also, the size of a region affects the probability of a catastrophe destroying more than a given percentage of property value.

By fitting a smooth surface that is a function of these variables to catastrophe loss data, it is possible to base estimates of expected losses for each region on more than just its own experience. Expected losses by region should generally have a smoother pattern than the sparse data.

An attempt is made in this paper to estimate the appropriate credibility to be given to the actual experience of a region, as opposed to the weight given to the expected losses indicated by a fitted smooth surface. After the indications of smoothed surfaces and the actual experience of a region are credibility-weighted to estimate the expected number of catastrophes for the region in various loss size intervals, a loss distribution is fitted to the estimates in order to smooth them in a reasonable way and to estimate tail probabilities.

2. THE MODEL

A. *Data*

To compare the relative destructive power of two natural catastrophes hitting different states, it is useful to consider the amount of property insurance premium in each state, as well as the amount of insured property damage in each state. The insured loss in each state will depend not only on the intensity and size of the catastrophe but also on the insured property in the area.

"Catastrophe premium," defined below, will be used as the exposure base to which loss data are related. The definition is based on Ingrey [12]. It is intended that the catastrophe premiums derived from each line of business be in roughly the same proportion as expected catastrophe losses for each line. Ingrey does not present data to support the percentages used in the formula but indicates that they were developed with the cooperation of Allen Hinkelman, Excess and Casualty Reinsurance Association; Daniel Holland, Inland Marine Insurance Bureau; Donald Kifer, New York Fire Insurance Rating Organization; and Allen Royer, Multi-Line Insurance Rating Board. Data on catastrophe losses by line will be discussed in Section 3.

The definition of catastrophe premium used in this paper is a formula often used by underwriters in evaluating a company's catastrophe exposure:

$$
\begin{aligned}
\text{Catastrophe premium} = &\ (10\% \text{ of inland marine premium}) \\
&+ (10\% \text{ of commercial multiple peril}) \\
&+ (80\% \text{ of allied lines}) \\
&+ (10\% \text{ of auto physical damage}) \\
&+ (20\% \text{ of farmowners}) \\
&+ (100\% \text{ of earthquake}) \\
&+ (20\% \text{ of homeowners}) \\
&+ (15\% \text{ of ocean marine}). \qquad (2.1)
\end{aligned}
$$

An assumption, for example, that the proportion of homeowners losses caused by catastrophes is twice as high as the proportion for auto physical damage losses is implicit in the formula, since the corresponding percentages of premium are 20% and 10%.

Actually, Ingrey's formula also includes 60% of mobile home premium and 80% of difference in conditions premium, but these premiums are small and are omitted.

Additional insight is given by expressing the loss layer to be reinsured in terms of percentages of the catastrophe premium—for example, 200% excess of 20%. In this paper, layers expressed as percentages of state or regional catastrophe premium are studied. Methods of applying the study to individual company catastrophe cover rating are also discussed.

Catastrophe covers are generally for a high enough layer so that an event must cause losses to several of a company's risks in order to produce a loss to the cover. Windstorms are the most frequent causes of losses to these covers. Other frequent causes are winter freezes, hail, and flooding. Fire is a less frequent cause.

The loss data used in this study were produced by Property Claim Services (PCS) [15]. These data include each United States catastrophe having an estimated insured loss of $1 million or more from 1949 through 1981, and $5 million or more from 1982 through 1989. In order to be included, a loss must affect many insureds, although the exact number of insureds that must be affected has not been defined. (It is generally at least 1,000.) For each catastrophe, the estimated insured loss in each state is given. The PCS estimates are based on an extrapolation of estimates made by a set of insurers writing most of the property premium in the catastrophe area.

Although PCS insured loss estimates are used in the study, a loss development factor is applied in Section 3, where the method of rating catastrophe covers is described.

For each of 28 overlapping regions of the continental United States, catastrophe premium was estimated for 1949 to 1989. Gross written premium data by state from Best's [4], and for older years from *The Spectator* [6], which is no longer published, were used to compute catastrophe premiums by state for approximately every fifth year. Exponential interpolation was used for other years, based on the computed catastrophe premiums.

For each of the 28 regions, the estimated insured loss from each catastrophe from 1949 to 1989 was divided by the region's catastrophe premium for the year of the loss. The ratios, R, of individual losses to corresponding catastrophe premiums were then grouped into the somewhat arbitrarily chosen intervals:

$$8\% < R \leq 16\%,$$

$$16\% < R \leq 32\%,$$

$$32\% < R \leq 64\%, \text{ and}$$

$$R > 64\%.$$

The number of ratios falling in each interval for each region is shown in Exhibit 1. Exhibit 2, a map of the United States, may be helpful in connection with Exhibit 1, as well as later exhibits.

No evidence of a trend in the frequency of any type of catastrophe was found in the data, so no trend factor was applied. The loss trend and the premium trend are assumed to cancel each other out.

B. Smoothing the Data

The expected values of frequencies in each interval vary more smoothly as a function of regions than the data in Exhibit 1, since the data include random variation.

Most catastrophes are windstorms, and their frequency and severity are related to a region's latitude, longitude, and distance from the coast (Clark [7] and Friedman [9]). The probability distribution of the ratios of catastrophe losses to catastrophe premium is also related to the size of a region. The above facts motivate the attempt to use multiple regression for each interval of R values to fit the frequencies in Exhibit 1 to functions of the latitude, longitude, distance from the coast, and area of the 28 regions.

Multiple regression was used to relate the above variables to frequency of catastrophes in each of the intervals: $8\% < R \leq 16\%$, $16\% < R \leq 32\%$, $32\% < R \leq 64\%$, $R > 64\%$, $R > 32\%$, $R > 16\%$, and $R > 8\%$. The intervals are purposely chosen in an overlapping manner for a reason explained in Subsection 2D.

The details of the regressions are in Appendix A. Exhibit 3 shows a comparison of actual to fitted frequencies for four of the intervals.

C. Experience Rating the Regions

Weights will be selected for the actual and fitted frequencies in Exhibit 3 to produce estimates of expected frequencies by interval and region. The sum of the weights will be one. An explanation of the method of selecting them follows.

For each interval i of R values, and each region j, let the random variable $X_{i,j}$ be the frequency of catastrophes in a randomly selected 41-year period. The fitted values for interval i and region j in Exhibit 3 are estimates of the expected value of $X_{i,j}$. If each fitted value is assumed to be the mean of a probability distribution of possible expected values of $X_{i,j}$, then it can be seen that a more accurate estimate of the expected value can be produced by giving weight (credibility) to the actual frequency as well as to the fitted frequency.

The partly judgmental basis for selecting the following experience rating formula is explained in Appendix B. The number of actual catastrophes in interval i and region j is given credibility $a_{i,j}/(a_{i,j}+k_i)$ where $a_{i,j}$ is the fitted frequency for interval i and region j, $k_i = 9$ for $i = 1, 2, 5, 6$, or 7, and $k_i = 6$ for $i = 3$ or 4; where, for each interval, i is as in Table 4 of Appendix A.

D. Nested Application of Experience Rating System

For each region, experience rating is applied to estimate expected values for the frequencies in each interval of R values.

A nested process is used so that the estimates of expected frequencies for $8\% < R \leq 16\%$ and $R > 16\%$ are based not only on the separate experience for $8\% < R \leq 16\%$ and $R > 16\%$, respectively, but also on the total experience for $R > 8\%$.

By applying the experience rating formula for the interval $R > 8\%$, estimates A_j of the frequency in this interval are produced for each region j. The estimates B_j and C_j produced by applying the experience rating system to the intervals $8\% < R \leq 16\%$ and $R > 16\%$ are then multiplied by a constant D_j such that $A_j = D_j\,(B_j + C_j)$. The estimates $D_j\,B_j$ and $D_j\,C_j$ for the frequencies in region j for intervals $8\% < R \leq 16\%$ and $R > 16\%$, respectively, thus sum to the estimate for region j for the interval $R > 8\%$ and are each in proportion to the estimates B_j and C_j, respectively. It is intended that $D_j\,B_j$ and $D_j\,C_j$ approximate the expected values of the frequencies in region j for intervals $8\% < R \leq 16\%$ and $R > 16\%$, respectively, given that the total of the two expected values is A_j, and that B_j and C_j are the estimates of the two expected values based on their separate data.

The weighted frequencies by region produced by directly applying the experience rating formulas for the intervals $16\% < R \leq 32\%$ and $R > 32\%$ are then adjusted so that their sum equals the estimate for $R > 16\%$. The method is entirely similar to the method used above to adjust the estimates for $8\% < R \leq 16\%$ and $R > 16\%$ so that their sum equaled the estimate for $R > 8\%$.

This nested process is continued until estimates are produced for each of the seven intervals. The estimates for four of the intervals are in Exhibit 4.

E. Loss Distributions by Region

The estimates of expected frequency for each region produced by the above nested application of experience rating for $8\% < R \leq 16\%$, $16\% < R \leq 32\%$, $32\% < R \leq 64\%$, and $R > 64\%$ were divided by the estimate produced for $R > 8\%$; the resulting fractions f_1, f_2, f_3, and f_4 were then fitted to a probability distribution. This probability distribution was used to allocate the estimate of expected frequency for $R > 8\%$ to the above four intervals. The selected yearly frequencies are the above frequencies divided by 41, since 41 years of data were used. The yearly frequencies for $R > 8\%$ are in Table 1.

The single parameter Pareto distribution was used for all 28 regions. It generally was a good fit. A comparison of the estimates produced by the experience rating method in the previous section and by the single parameter Pareto is shown in Exhibit 4. No other tested distribution performed as well. (A study of loss distributions is in Hogg and Klugman [11].)

The single parameter Pareto was used even in regions for which another distribution fit better. This was because the generally good fit of the single parameter Pareto led to the conclusion that it was a good model for the data, and small amounts of data in particular regions were not considered credible enough to counteract this conclusion.

(See Appendix C for a discussion of the method used to fit the single parameter Pareto. The parameters of the Pareto curves used are in Table 1.)

A Pareto parameter of 1 or less implies infinite expected losses for unlimited layers. For $0 < P < 1$, the expected losses in the layer between a and b are $(b^{1-P} - a^{1-P})/(1-P)$, which approaches infinity as b approaches infinity. In reality, catastrophe losses are limited by the total insured value, so the frequency distribution falls below a

TABLE 1

FREQUENCIES (F^1) AND PARAMETERS (P)

Region	F^1	P	Region	F^1	P
1	0.213	0.96	15	0.292	0.94
2	0.335	1.21	16	0.190	1.07
3	0.727	1.26	17	0.312	1.00
4	0.682	0.95	18	0.212	1.08
5	0.419	0.60	19	0.244	1.44
6	0.431	0.86	20	0.590	0.92
7	0.749	1.61	21	0.507	1.13
8	0.184	1.24	22	0.450	1.78
9	0.235	1.27	23	0.265	1.25
10	0.566	1.49	24	0.196	0.93
11	0.788	1.54	25	0.183	1.17
12	0.453	1.59	26	0.487	1.33
13	0.254	1.16	27	0.265	1.00
14	0.282	0.98	28	0.393	1.54

Pareto at some point. Although Pareto parameters of 1 or less were selected for some regions, they are only intended to be used in estimating expected losses for limited layers of sizes that are actually reinsured. The Pareto's overestimate of frequency far out in the tail does not have a great effect in estimating expected losses for these layers. The frequency of losses above x times the truncation point is x^{-P} times the frequency above the truncation point. Since $P > 0$, this fraction x^{-P} approaches zero as x approaches infinity.

3. RATING CATASTROPHE COVERS

A. *Using the Model*

Rates for catastrophe covers include a risk charge, but this discussion is of expected losses rather than risk.

A reinsurer evaluating a catastrophe cover often receives a breakdown of the ceding company's subject property premium by state and line. The commercial multiple peril, homeowners, farmowners, and auto physical damage premiums that are considered to be subject to a catastrophe treaty are sometimes only a percentage (usually approximately 65%, 90%, 90%, and 35%, respectively) of the total premiums for those lines. It is necessary to adjust for this reduction to apply the catastrophe premium formula in this paper to the cedent.

If the cedent does not provide this information, estimates of catastrophe premium by state for a primary company can be made by using the company's major direct premium writings by state and its net written premiums by line from *Best's Insurance Reports* [3]. Based on this information and on Table 2, one of the 28 regions may be selected judgmentally as being approximately representative of the region in which the company writes.

TABLE 2

1988 CATASTROPHE PREMIUMS BY REGION (IN 000S)

Region	Premium	Region	Premium
1	$1,757,793	15	$890,083
2	473,889	16	973,760
3	881,629	17	789,209
4	521,551	18	2,231,681
5	668,967	19	546,455
6	700,932	20	1,403,180
7	478,800	21	1,848,699
8	365,904	22	1,484,958
9	180,551	23	1,793,682
10	238,494	24	2,653,051
11	273,418	25	2,778,136
12	973,046	26	3,366,938
13	1,110,098	27	5,816,632
14	683,584	28	11,961,706

For any region selected as representative of the company, the selected yearly frequency for catastrophe losses greater than 8% of catastrophe premium and the selected Pareto distribution may be found in Table 1. They may be used to compute an estimate of expected losses for any layer of a catastrophe cover by expressing the layer in terms of percentages of the company's total catastrophe premium. An example of the rating method will be given at the end of this section, but several related points are discussed first.

The method used in the example is based on historical data. However, due to the potential for an enormously damaging earthquake in California and the small number of earthquakes in the historical data used, expected losses from catastrophes in California are widely believed to be greater than the estimate that would be based on historical data. The very severe 1906 earthquake is not included in the available data.

An adjustment will be made in the rating method for catastrophe covers to reflect that the model in this paper is based on data for regions rather than for individual reinsurers. By the use of certain definitions and reasonable assumptions, the following statement could be made more precise and proven mathematically. On average, for catastrophe losses as defined by PCS, the probability distribution of ratios of catastrophe losses to catastrophe premiums has the same mean for an insurer within a region as for the region—but it has a greater variance.

The rating method, which will be applied to individual insurers, uses 0.85 times the Pareto parameter in Table 1 for the region selected as representative of the insurer. This adjustment reflects that the distributions for individual insurers have greater variance, on the average, than the distribution for the region.

The expected frequencies from Table 1 will be used, unadjusted, for individual insurers. The expected frequency of catastrophe losses, as defined by PCS, is less for an individual insurer than for the surrounding region. However, the assumption of a smaller Pareto parameter for individual insurers implies that for some percentage W, the expected frequency for $R > W\%$ is the same for the individual insurer as for the region. The estimate that $W = 8\%$ is implicit in the use of the expected frequencies from Table 1 for individual insurers.

The estimate that ultimate insured losses for catastrophes, on the average, are 1.33 times as great as the PCS estimates will be used in estimating expected losses for catastrophe covers. Since the PCS estimate is made within a few days of the catastrophe, it is natural to expect development. Also, the PCS estimate excludes allocated loss adjustment expense, all ocean marine and crop losses, and some inland marine and business interruption losses. Lastly, the model in this paper used gross losses and premiums while catastrophe reinsurance covers losses net of excess reinsurance. Studies (e.g., Ludwig [13]) have shown that net catastrophe losses are a higher percentage of net premiums than gross catastrophe losses are of gross premiums. An adjustment for this is included in the 1.33 factor.

The 0.85 factor for Pareto parameters and the 1.33 factor for losses have the combined effect of significantly raising estimated expected losses for catastrophe covers. The resulting expected losses, as a percentage of actual premiums charged, have been found to be a reasonable match to actual loss ratios for the catastrophe cover premium of two reinsurers over a 20-year and a 12-year period, respectively. (In addition, an adjustment was made to include the catastrophic year 1992.) This premium totaled almost $300 million and consisted of shares of a much greater amount of premium.

Example

Suppose that a primary insurer, in the latest year for which data are available, had writings for which region 28 is considered the best match.

Suppose that, using cp to represent the insurer's catastrophe premium, the layer to be reinsured can be expressed as $(2.00cp)$ excess of $(0.20cp)$.

The selections in Table 1 for region 28 were 0.393 catastrophe losses per year greater than 8% of catastrophe premium and a Pareto parameter of 1.54. The loss development factor of 1.33 and the adjustment factor to the Pareto parameter of 0.85, which were discussed above, are used. Therefore, 0.393 is the frequency for $R > 10.64\%$, and the Pareto parameter becomes 1.31. The expected losses to the layer in one year therefore are:

$$0.393\,(\,0.1064cp\,)\left[\frac{(0.20/0.1064)^{-0.31} - (2.20/0.1064)^{-0.31}}{0.31}\right] \tag{3.1}$$

(See Philbrick [14].) This amount equals 5.82% of catastrophe premium.

If it is not clear which region is the best match for the primary insurer, the above method may be used for more than one region and a final estimate may be judgmentally selected.

B. Underwriting Judgment

Since the above estimate is based on data from the entire region, it may be useful to judgmentally modify it if the ceding company is not believed to be typical of the region. For example, the ceding company may have a very high or low percentage of its insured property near the coast, where exposure to hurricanes is greatest. An estimate of how a ceding company compares to a region could also be made by using Clark's model [7], since that software can be applied to both regions and individual companies.

C. The Catastrophe Premium Formula

The estimated expected catastrophe losses for individual insurers were affected by the choice of percentages by line in the catastrophe premium formula defined in Section 2.

If the percentages by line that were used in the formula are multiplied by the corresponding premiums in Table 3, an approximation of the relative amounts of expected catastrophe losses by line can be derived. (Although fire premium is a portion of the property premium in Table 3, it was not included in the catastrophe premium formula; it was considered to account for only a negligible portion of catastrophe losses.)

Some data suggest that, for hurricanes, a much lower percentage of losses comes from auto physical damage than would be estimated based on the catastrophe premium formula. In [1], the All-Industry Research Advisory Council (AIRAC) estimated the following percentages of losses by line for seven hurricanes from 1983 to 1985:

Homeowners Multiple Peril	46.8%
Commercial Multiple Peril	22.2%
Auto Physical Damage	3.7%
All Others	27.3%.

TABLE 3

INDUSTRY PREMIUMS FOR SELECTED LINES —1990

	Premiums Earned (Millions)
Fire	$ 4,494
Allied Lines	2,097
Farmowners Multiple Peril	968
Homeowners Multiple Peril	18,116
Commercial Multiple Peril	17,626
Ocean Marine	1,169
Inland Marine	4,441
Earthquake	459
Auto Physical Damage	35,185

Another source of data on catastrophe losses by line was produced by the Insurance Services Office (ISO) for homeowners losses by individual catastrophe for the period 1970 to 1978 [2]. Those data indicate that homeowners and dwelling extended coverage losses are 19.6% and 2.7%, respectively, of total catastrophe losses as estimated by PCS for the same catastrophes. (The ISO estimates, like the PCS estimates, are an extrapolation of total insured losses based on data from a set of insurers in the region.) The percentage of total catastrophe losses covered under homeowners is much less in the ISO data for all catastrophes combined than in the AIRAC hurricane data. Therefore, the percentage of auto physical damage losses may well be much greater for all catastrophes combined than for hurricanes.

Hurricanes produced $6.35 billion in catastrophe losses from 1981 to 1990, compared to $9.7 billion in losses from hail and tornadoes and $3.7 billion in losses from winter storms, according to PCS.

If so desired, the catastrophe cover rating method used in this paper can be applied with a catastrophe premium formula having different percentages by line from those used. Any alternative percentages used should be chosen so that, when multiplied by the premiums in Table 3, they produce the same catastrophe premium as the percentages in this paper's formula. If this is done, then Table 1 approximates the corresponding table that would have been created if the alternative catastrophe premium formula had been used in the study. Therefore, the rating method used in this paper still gives an estimate of expected losses from catastrophes if the alternative catastrophe premium formula is used.

D. Experience Rating a Catastrophe Risk

Suppose the amount of each catastrophe loss of the ceding company for a certain time period is known. The frequency of these losses in intervals expressed in terms of ratios to the company's catastrophe premium can be compared to the experience of the region selected as being representative of the company. Exhibit 5, which shows experience from 1949 to 1969 and from 1970 to 1989 separately, may be useful for this comparison. An example of a judgmental experience rating is given below.

Example

Suppose that Insurance Company A had eight catastrophes greater than 10.64% (i.e., 8% times our selected development factor) of catastrophe premium for the period of 1970 to 1989 and that the region selected as corresponding to it had five catastrophes greater than 10.64% of catastrophe premium in the same period.

Suppose that the formula $n/(n+9)$, where n is the number of catastrophes in the region from 1970 to 1989, is the credibility assigned to the experience of Company A. (This formula is similar to one used in this paper to assign credibility to the actual frequency of catastrophes in a region.)

The credibility weighted frequency is then

$$(5/(5+9))\,(8) + (9/(5+9))\,(5),$$

which equals 6.07. The modifier produced by the experience rating is thus 6.07/5.00; that is, 1.21. This modifier is then applied to the expected losses for the reinsured layer that are estimated as in Equation 3.1.

4. CONCLUSION

A model that can be used to estimate expected losses to catastrophe covers based on insured loss data has been presented. An example of the application of the model to a specific cover was given. The obstacles to using actuarial methods in catastrophe rating are not as great as has sometimes been suggested.

The application of actuarial science gives a very useful and much needed perspective in this area.

REFERENCES

[1] *Catastrophic Losses: How the Insurance System Would Handle Two $7 Billion Hurricanes,* All-Industry Research Advisory Council, Oak Brook, Illinois, November, 1986.

[2] IRUA Actuarial Committee, Catastrophe Losses by State Assignment, Letter from Virgil H. Applequist to Charles Hewitt, Jr., June, 1980.

[3] *Best's Insurance Reports: Property-Casualty,* A. M. Best Company, Oldwick, New Jersey.

[4] *Experience By State (By Line),* formerly *Best's Executive Data Service,* A. M. Best Company, Oldwick, New Jersey.

[5] Bühlmann, Hans, "Experience Rating and Credibility," *ASTIN Bulletin* 4.3, 1967, p. 199.

[6] *The Spectator,* Chilton Co., Inc., Radnor, Pennsylvania, 1868-1975.

[7] Clark, Karen M., "A Formal Approach to Catastrophe Risk Assessment in Management," *PCAS* LXXIII, 1986, pp. 69-92.

[8] Draper, Norman R. and Harry Smith, *Applied Regression Analysis,* Second Edition, John Wiley and Sons, 1981.

[9] Friedman, D. G., "Insurance and the Natural Hazards," *ASTIN Bulletin* 7.1, 1972, p. 4.

[10] Herzog, Thomas N.,"An Introduction to Bayesian Credibility and Related Topics," in Part 4 Study Kit, Casualty Actuarial Society, 1985.

[11] Hogg, Robert V. and Stuart A. Klugman, *Loss Distributions,* John Wiley and Sons, 1984.

[12] Ingrey, Paul B., "Reinsuring the Catastrophe," Master's Thesis, The College of Insurance, 1971.

[13] Ludwig, Stephen J., "An Exposure Rating Approach to Pricing Property Excess-of-Loss Reinsurance," *PCAS* LXXVIII, 1991, pp. 110-145.

[14] Philbrick, Stephen W., "A Practical Guide to the Single Parameter Pareto Distribution," *PCAS* LXXII, 1985, pp. 44-84.

[15] *Catastrophe History of the 50 States*, Property Claim Services, Rahway, New Jersey.

EXHIBIT 1

FREQUENCIES BY REGION

Region	Interval of Ratio R			
	$8\% < R \leq 16\%$	$16\% < R \leq 32\%$	$32\% < R \leq 64\%$	$R > 64\%$
1. CA	3	1	2	0
2. AZ, NM, NV, UT, CO	10	4	1	1
3. TX	22	1	4	3
4. AL, MS, LA	14	3	5	5
5. FL	4	5	2	5
6. GA, SC, NC	8	6	4	2
7. TN, AR, OK	23	8	1	0
8. OR, WA, ID	4	1	0	1
9. ND, SD, WY, MT	4	5	1	1
10. MN, WI	13	6	5	1
11. NE, KS	22	9	4	1
12. IA, MO, IL	11	6	0	0
13. MI, IN, OH	6	2	1	1
14. KY, WV, PA	6	1	4	0
15. VA, NJ, DE, MD, DC	6	2	1	2
16. NY, VT	2	2	1	0
17. ME, NH, MA, RI, CT	7	5	0	2
18. 1, 2 (above)	3	3	1	0
19. 8, 9	8	3	0	1
20. 3, 4	8	7	2	6
21. 5, 6, 7	18	4	3	1
22. 10, 11, 12	14	4	0	0
23. 13, 14	7	3	1	0
24. 15, 16, 17	1	2	1	2
25. 1, 2, 8, 9	3	1	2	0
26. 3, 4, 7, 10, 11, 12	11	4	3	1
27. 5, 6, 13, 14, 15, 16, 17	5	2	3	1
28. Continental U.S.	9	4	2	0
	252	104	54	37

EXHIBIT 2

THE CONTINENTAL UNITED STATES

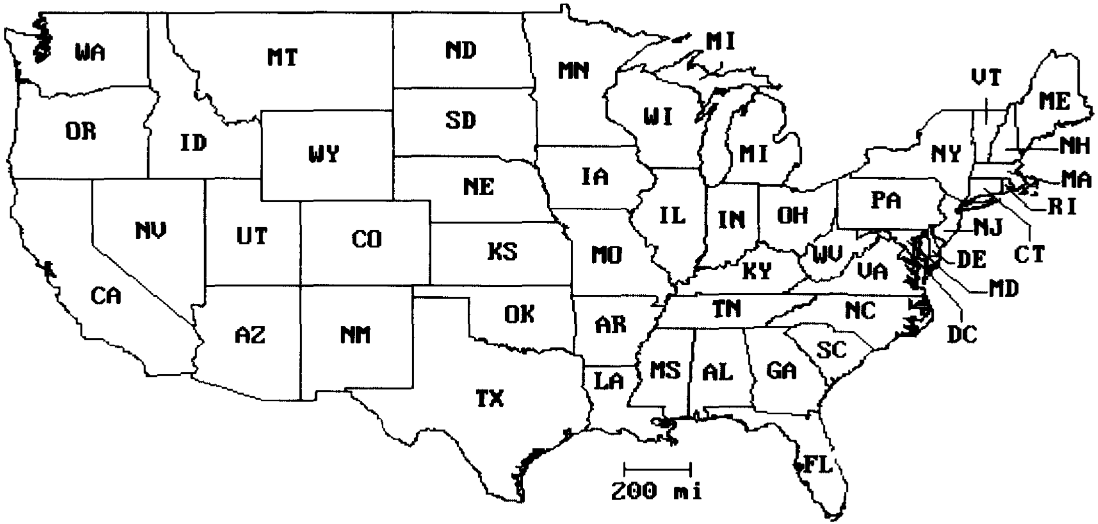

EXHIBIT 3

COMPARISON OF ACTUAL (A) TO FITTED (F) FREQUENCIES

	Interval of Ratio R							
	$8\% < R \leq 16\%$		$16\% < R \leq 32\%$		$32\% < R \leq 64\%$		$R > 64\%$	
Region	A	F	A	F	A	F	A	F
1	3	5.71	1	2.91	2	1.84	0	1.71
2	10	5.61	4	2.62	1	1.84	1	0.74
3	22	17.60	1	5.31	4	3.82	3	1.86
4	14	17.77	3	5.61	5	3.82	5	3.25
5	4	7.23	5	3.24	2	4.32	5	4.45
6	8	6.15	6	2.93	4	2.44	2	1.65
7	23	16.11	8	5.53	1	2.61	0	1.35
8	4	4.73	1	2.79	0	0.90	1	0.73
9	4	4.59	5	2.69	1	0.82	1	0.44
10	13	12.72	6	5.65	5	1.01	1	0.68
11	22	14.46	9	5.53	4	1.70	1	0.87
12	11	14.43	6	5.46	0	1.70	0	0.85
13	6	5.14	2	2.95	1	1.20	1	0.70
14	6	5.54	1	3.00	4	1.59	0	0.88
15	6	5.60	2	3.20	1	1.59	2	1.51
16	2	4.97	2	3.20	1	0.99	0	1.07
17	7	4.92	5	3.24	0	0.94	2	2.75
18	3	5.59	3	2.58	1	1.84	0	0.82
19	8	4.62	3	2.60	0	0.86	1	0.51
20	8	17.48	7	5.12	2	3.82	6	1.95
31	18	6.21	4	2.71	3	2.69	1	2.02
22	14	13.73	4	5.07	0	1.48	0	0.79
23	7	5.27	3	2.79	1	1.38	0	0.71
24	1	5.05	2	2.88	1	1.14	2	1.13
25	3	5.10	1	2.47	2	1.32	0	0.57
26	11	15.70	4	4.80	3	2.61	1	1.22
27	5	5.53	2	2.61	3	1.75	1	0.91
28	9	14.44	4	4.49	2	1.97	0	0.89
	252	252.00	104	103.98*	54	53.99*	37	37.01*

*Totals do not always match exactly, due to rounding.

EXHIBIT 4
Part 1

COMPARISON OF EXPERIENCE RATED FREQUENCIES WITH FITTED PARETO FREQUENCIES

	Experience Rated Frequencies			
Region	$8\% < R \leq 16\%$	$16\% < R \leq 32\%$	$32\% < R \leq 64\%$	$R > 64\%$
1	4.28	1.97	1.45	1.03
2	7.83	3.07	1.92	0.90
3	20.14	3.61	3.92	2.14
4	15.00	4.81	4.29	3.87
5	5.73	3.43	3.34	4.67
6	7.54	4.19	3.73	2.22
7	20.72	6.72	2.14	1.11
8	4.30	2.03	0.63	0.61
9	4.72	3.46	0.94	0.53
10	13.52	7.04	1.83	0.82
11	20.22	8.29	2.70	1.08
12	11.62	5.15	1.15	0.64
13	5.61	2.65	1.33	0.83
14	5.86	2.43	2.41	0.88
15	5.91	2.79	1.57	1.72
16	3.67	2.48	0.86	1.79
17	6.28	4.18	0.57	1.75
18	4.31	2.31	1.44	0.63
19	6.12	2.65	0.70	0.52
20	10.95	6.32	3.56	3.36
21	12.58	3.32	2.99	1.89
22	13.03	4.06	0.86	0.51
23	6.09	2.74	1.37	0.66
24	3.43	2.50	0.98	1.12
25	4.04	1.84	1.17	0.42
26	12.02	4.52	2.37	1.03
27	5.41	2.44	2.07	0.94
28	10.17	4.02	1.39	0.54
	251.10	105.04	53.68	37.23

EXHIBIT 4
Part 2

COMPARISON OF EXPERIENCE RATED FREQUENCIES WITH FITTED PARETO FREQUENCIES

	Fitted Pareto Frequencies			
Region	$8\% < R \leq 16\%$	$16\% < R \leq 32\%$	$32\% < R \leq 64\%$	$R > 64\%$
1	4.26	2.18	1.12	1.18
2	7.77	3.37	1.46	1.12
3	17.37	7.25	3.03	2.17
4	13.53	6.99	3.61	3.85
5	5.85	3.86	2.54	4.92
6	7.91	4.37	2.42	2.99
7	20.63	6.76	2.21	1.08
8	4.36	1.85	0.78	0.57
9	5.64	2.34	0.97	0.69
10	14.96	5.32	1.89	1.04
11	21.21	7.28	2.50	1.31
12	12.38	4.13	1.38	0.69
13	5.76	2.57	1.15	0.93
14	5.70	2.89	1.47	1.51
15	5.74	2.99	1.56	1.69
16	4.09	1.94	0.92	0.84
17	6.39	3.20	1.60	1.60
18	4.58	2.17	1.03	0.93
19	6.31	2.33	0.86	0.50
20	11.39	6.03	3.19	3.58
21	11.31	5.16	2.35	1.97
22	13.07	3.81	1.11	0.46
23	6.29	2.65	1.11	0.81
24	3.82	2.00	1.05	1.16
25	4.15	1.85	0.82	0.66
26	12.02	4.77	1.90	1.25
27	5.41	2.71	1.36	1.36
28	10.58	3.64	1.25	0.66
	252.49	106.41	46.64	41.51

EXHIBIT 5

Regional Frequencies by Time Period

	Interval of Ratio R							
	$8\% < R \leq 16\%$		$16\% < R \leq 32\%$		$32\% < R \leq 64\%$		$R > 64\%$	
Region	1949-69	1970-89	1949-69	1970-89	1949-69	1970-89	1949-69	1970-89
1	1	2	1	0	1	1	0	0
2	2	8	3	1	0	1	0	1
3	10	12	0	1	1	3	1	2
4	4	10	0	3	4	1	3	2
5	2	2	1	4	1	1	5	0
6	4	4	3	3	2	2	1	1
7	9	14	4	4	0	1	0	0
8	0	4	0	1	0	0	1	0
9	2	2	3	2	1	0	1	0
10	3	10	2	4	3	2	1	0
11	9	13	3	6	3	1	1	0
12	7	4	4	2	0	0	0	0
13	4	2	2	0	1	0	0	1
14	2	4	0	1	2	2	0	0
15	3	3	2	0	0	1	2	0
16	1	1	1	1	1	0	0	0
17	1	6	4	1	0	0	2	0
18	0	3	2	1	0	1	0	0
19	3	5	2	1	0	0	1	0
20	3	5	3	4	1	1	3	3
21	7	11	3	1	3	0	0	1
22	7	7	4	0	0	0	0	0
23	4	3	3	0	0	1	0	0
24	1	0	0	2	1	0	2	0
25	1	2	1	0	1	1	0	0
26	7	4	1	3	1	2	1	0
27	2	3	1	1	2	1	1	0
28	1	8	3	1	0	2	0	0

APPENDIX A

DETAILS OF REGRESSIONS

The "center" of a region is defined as the point such that half the area is to the north, half to the east, half to the west, and half to the south. For each of the 28 regions, the latitude and longitude of the centers of the regions were estimated and considered to be the latitude and longitude of the region. The "distance to the coast" of a region is defined as the length of the shortest line from the center to any ocean.

The independent variables used in the regression were x_1, x_2, x_3, and x_4, such that, for each region,

$$x_1 = \text{latitude of region;}$$

$$x_2 = \begin{cases} 0, & \text{if } 92 \le \text{longitude of region} \le 99, \\ |\text{longitude} - 99|, & \text{if } 99 < \text{longitude} < 105, \\ 6, & \text{if longitude} \ge 105, \\ |\text{longitude} - 92|, & \text{if } 86 < \text{longitude} < 92, \\ 6, & \text{if longitude} \le 86; \end{cases}$$

$$x_3 = \ln(\ln(\text{area of region, in thousands of miles}));$$

$$x_4 = \ln(\ln(\text{distance from coast of region, in miles})).$$

The values of x_1, x_2, x_3, and x_4, for the 28 regions are given in Exhibit 6.

For each of the seven intervals for R, the dependent variable used in the regression for the interval was ln(frequency of catastrophes). (In cases where the frequency was zero, ln (⅓) was judgmentally used instead of the undefined ln (0).) This dependent variable was chosen so that, for each independent variable, a given amount of change would produce a fixed multiplicative effect on the fitted frequencies defined below.

This approach produced a better fit than any other dependent variable and avoided the problem of negative or unreasonably small fitted values. An attempt was made to use (frequency of catastrophes)$^{0.5}$ as the dependent variable, since its variance is relatively close to being independent of the expected frequency of catastrophes, and this is desirable when using regression. However, it did not produce the most acceptable fitted values.

The use of ln (ln (x)) for x_3 and x_4 resulted from the observation that it produced values of x_3 and x_4 that came reasonably close to having the desired linear relationship with the values of the dependent variable.

For each interval I_i of R values, there is a corresponding set of frequencies by region $\{f_{i,j}\}$, where j is an integer from 1 to 28.

Fitted values $\hat{y}_{i,j}$ were produced by regression. Then the function

$$g_i(\hat{y}_{i,j}) = \exp(\hat{y}_{i,j}) \left(\left(\sum_{j=1}^{28} f_{i,j} \right) / \sum_{j=1}^{28} \exp\left(\hat{y}_{i,j}\right) \right) \tag{A.1}$$

was used to produce values $g_i(\hat{y}_{i,j})$ such that

$$\sum_{j=1}^{28} g_i(\hat{y}_{i,j}) = \sum_{j=1}^{28} f_{i,j}\,.$$

The values $g_i(\hat{y}_{i,j})$, rather than $\hat{y}_{i,j}$, were used as final fitted values for the frequencies $f_{i,j}$.

Tornadoes are more prevalent in the region between longitudes 92 and 99, which helps explain the motivation for the definition of the variable x_2.

The interval $R > 64\%$ was the only one for which x_4 was used. It appears that distance from the coast is a useful variable for large hurricanes, but not for smaller catastrophes such as tornadoes. The variable x_4 didn't work well for intervals for which $R \leq 64\%$, possibly due to collinearity with the longitude variable. The coefficient came out only negligibly negative or even positive.

Positive coefficients for any of the variables x_1, x_2, x_3, and x_4 were considered counter to the overall indications of the data and not appropriate for use in the study. For all intervals, all the variables x_1, x_2, and x_3 were used unless one of them had a positive coefficient. In these cases, a regression was done without using that variable.

To find confidence intervals for the regression coefficients or for the expected values of the dependent variables, it would have to be true that:

1. A linear relationship exists between the independent variables used and the dependent variable used.
2. The conditional distributions of the dependent variables, given values of the independent variables, are uncorrellated and have a common variance.

Neither condition is satisfied. Nothing can be done to satisfy the first condition unless a way is known to transform the variables so that they satisfy a linear relationship. Therefore, it was considered better to avoid the complications involved in transforming variables to come closer to satisfying the second condition. The results of the regression are considered to be simply a useful method of smoothing the data.

The functions resulting from the regressions are shown in Table 4.

TABLE 4
REGRESSION FUNCTIONS

i	Interval	Function
1	$8\% < R \le 16\%$	$-0.024x_1 - 0.167x_2 - 0.083x_3 + 3.694$
2	$16\% < R \le 32\%$	$-0.00005x_1 - 0.108x_2 - 0.461x_3 + 2.312$
3	$32\% < R \le 64\%$	$-0.095x_1 - 0.035x_2 + 4.169$
4	$R > 64\%$	$-0.030x_1 - 0.069x_2 - 0.241x_3 - 2.719x_4 + 6.457$
5	$R > 32\%$	$-0.102x_1 - 0.002x_2 - 0.808x_3 + 6.150$
6	$R > 16\%$	$-0.047x_1 - 0.087x_2 - 0.720x_3 + 5.172$
7	$R > 8\%$	$-0.035x_1 - 0.119x_2 - 0.596x_3 + 5.393$

EXHIBIT 6

VALUES OF INDEPENDENT VARIABLES

Region	x_1	x_2	x_3	x_4
1	37	6	1.6.12	1.535
2	37	6	1.838	1.824
3	31.5	0	1.715	1.708
4	31.5	0	1.596	1.513
5	28	6	1.381	1.303
6	34	6	1.596	1.582
7	35.5	0	1.626	1.790
8	44.5	6	1.703	1.758
9	45.5	6	1.787	1.924
10	45.5	0	1.581	1.936
11	40	0	1.626	1.903
12	40	0	1.654	1.909
13	41.5	6	1.581	1.818
14	38.5	6	1.548	1.767
15	38.5	6	1.405	1.582
16	43.5	6	1.405	1.652
17	44	6	1.381	1.303
18	37	6	1.876	1.780
19	45	6	1.862	1.868
20	31.5	0	1.796	1.684
21	33	6	1.767	1.504
22	41.5	0	1.813	1.902
23	40	6	1.703	1.817
24	42	6	1.640	1.629
25	40.5	6	1.970	1.868
26	35.5	0	1.935	1.798
27	37.5	6	1.854	1.740
28	38.5	0	2.078	1.870

APPENDIX B

Derivation of Credibility Formula

To approximate an experience rating formula, we assume:

1. Given that $g_i\,(\hat{y}_{i,j})$ is the fitted value for interval i and region j in the smoothing method of this paper, the probability distribution of the random variable $E_{i,j}$, which represents the expected value of the frequency of catastrophes in interval i and region j, has mean $g_i\,(\hat{y}_{i,j})$.
2. For each i, the probability distribution of $E_{i,j}$ has the same coefficient of variation C_i for each j.

It follows that, for each interval i and each region j, the Z such that

$$Z\,(\text{actual frequency in interval } i \text{ and region } j) + (1-Z)\,g_i\,(\hat{y}_{i,j}) \tag{B.1}$$

is the best least squares estimate of the expected value of the frequency in interval i and region j is

$$Z = g_i\,(\hat{y}_{i,j})/(g_i\,(\hat{y}_{i,j}) + 1/C_i^2\,). \tag{B.2}$$

The proof is as follows. By Bühlmann's theorem (Bühlmann [5], Herzog [10]), $Z = H_{i,j}/(H_{i,j} + V_{i,j})$ where $H_{i,j}$ equals the variance of the probability distribution of the expected value of the frequency for interval i and region j, and $V_{i,j}$ equals the expected value of the variance of the frequency, given the above probability distribution for the expected value of the frequency.

For each possible value $e_{i,j}$ for the expected value of the frequency, the probability distribution of actual values is assumed to be Poisson and thus has variance $e_{i,j}$. Therefore, by Assumption 1 above, $V_{i,j} = g_i\,(\hat{y}_{i,j})$. By Assumption 2 above, $H_{i,j} = (C_i\,g_i\,(\hat{y}_{i,j}))^2$. Therefore,

$$\begin{aligned} Z &= C_i^2\, g_i\,(\hat{y}_{i,j})^2/(C_i^2\, g_i\,(\hat{y}_{i,j})^2 + g_i\,(\hat{y}_{i,j})) \\ &= g_i\,(\hat{y}_{i,j})/(g_i\,(\hat{y}_{i,j}) + 1/C_i^2). \end{aligned} \tag{B.3}$$

This completes the proof.

The estimates of the numbers C_i^2 will now be discussed.

Consider the frequency in interval i and region j during the 41-year period used for the data to be the outcome of an experiment. Let the random variable $X_{i,j}$ represent the outcome. The expected value of $(g_i\,(\hat{y}_{i,j}) - X_{i,j})^2$, given that $e_{i,j}$ is the expected value of the frequency, equals $(g_i\,(\hat{y}_{i,j}) - e_{i,j})^2$ plus the expected value, given that $e_{i,j}$ is the expected value of the frequency, of $(e_{i,j} - X_{i,j})^2$. (This is left for the reader to verify.) Therefore, the mean of $(g_i\,(\hat{y}_{i,j}) - X_{i,j})^2$ equals the mean of $(g_i\,(\hat{y}_{i,j}) - E_{i,j})^2$ plus the mean of $(E_{i,j} - X_{i,j})^2$.

By Assumption 2 above, the mean of $(g_i\,(\hat{y}_{i,j}) - E_{i,j})^2$ equals $C_i^2\,(g_i\,(\hat{y}_{i,j}))^2$.

Given that $e_{i,j}$ is the expected value of the frequency, the mean of $(e_{i,j} - X_{i,j})^2$ is $e_{i,j}$. Therefore, the mean of $(E_{i,j} - X_{i,j})^2$ equals the mean of $E_{i,j}$, which is $g_i\,(\hat{y}_{i,j})$.

Therefore, the mean of $(g_i\,(\hat{y}_{i,j}) - X_{i,j})^2$ equals $C_i^2\,(g_i\,(\hat{y}_{i,j}))^2 + g_i\,(\hat{y}_{i,j})$. So C_i^2 equals the expected value of

$$\left(\sum_{j=1}^{28} (g_i\,(\hat{y}_{i,j}) - X_{i,j})^2 - \sum_{j=1}^{28} g_i\,(\hat{y}_{i,j})\right) \Big/ \sum_{j=1}^{28} g_i\,(\hat{y}_{i,j})^2. \tag{B.4}$$

The estimate of the expected value of

$$\sum_{j=1}^{28} (g_i\,(\hat{y}_{i,j}) - X_{i,j})^2$$

will depend partly on judgment and intuition, due to problems in estimating it purely mathematically.

Assume for the sake of approximation that the following two conditions are satisfied.

1. The values $g_i(\hat{y}_{i,j})$ are the function values produced directly by a regression, and a linear relationship with coefficients $a_{i,j}$ actually exists between the independent variables used and the expected values of the dependent variables.
2. The differences between the dependent variables and their expected values have independent probability distributions with a common variance σ^2.

Under these conditions,

$$\left(\sum_{j=1}^{28} (g_i(\hat{y}_{i,j}) - f_{i,j})^2\right) / (\text{degrees of freedom}) \,, \tag{B.5}$$

where $f_{i,j}$ is the actual frequency in interval i and region j, is an unbiased estimate of σ^2 (Draper and Smith [8]). If the values $g_i(\hat{y}_{i,j})$ are not the true expected values of the frequencies in interval i and region j, then the expected value of

$$\left(\sum_{j=1}^{28} (g_i(\hat{y}_{i,j}) - X_{i,j})^2\right) / 28$$

is greater than σ^2.

Assuming Equation B.5 is equal to or less than the expected value of

$$\left(\sum_{j=1}^{28} (g_i(\hat{y}_{i,j}) - X_{i,j})^2\right) / 28,$$

Equation B.4 gives the following lower bound for C_i^2 :

$$(\text{Equation B.5} - \sum_{j=1}^{28} g_i(\hat{y}_{i,j}))/\sum_{j=1}^{28} (g_i(\hat{y}_{i,j}))^2. \tag{B.6}$$

We now discuss an upper bound for C_i^2.

It clearly appears that the expected value of

$$\sum_{j=1}^{28} (g_i(\hat{y}_{i,j}) - X_{i,j})^2$$

is less than

$$\sum_{j=1}^{28} \left(\left(\sum_{j=1}^{28} g_i(\hat{y}_{i,j}) \right) /28 - f_{i,j} \right)^2,$$

where $f_{i,j}$ is the actual frequency in interval i and region j. The value

$$\left(\sum_{j=1}^{28} g_i(\hat{y}_{i,j}) \right) /28$$

is a mere average of the values $g_i(\hat{y}_{i,j})$, so the individual estimates $g_i(\hat{y}_{i,j})$ intuitively appear to be better estimators for the expected values of the variables $X_{i,j}$ than is

$$\left(\sum_{j=1}^{28} g_i(\hat{y}_{i,j}) \right) /28.$$

Therefore, it follows, based on the above arguments and Equation B.4, that the following is an upper bound for C_i^2:

$$\sum_{j=1}^{28} \left(\left(\sum_{j=1}^{28} (g_i(\hat{y}_{i,j})/28 - f_{i,j})^2 - \sum_{j=1}^{28} g_i(\hat{y}_{i,j}) \right) / \sum_{j=1}^{28} (g_i(\hat{y}_{i,j}))^2 \right). \tag{B.7}$$

Thus we have (Equation B.6) < C_i^2 < (Equation B.7). Using the actual values of the expressions in Equations B.6 and B.7 for $i = 1$ through 7, and averaging inequalities, gives

$$0.049 < ((C_1^2 + C_2^2 + C_5^2 + C_6^2 + C_7^2)/5) < 0.146 \tag{B.8}$$

and

$$0.065 < ((C_3^2 + C_4^2)/2) < 0.215. \tag{B.9}$$

The reason for considering C_3 and C_4 separately from C_1, C_2, C_5, C_6, and C_7 is that the numbers $g_i(\hat{y}_{i,j})$ for $i = 3$ and $i = 4$ were based on less data than for $i = 1, 2, 5, 6$, and 7. Thus, the expectation is that they are less accurate. Therefore, it can be seen from Equation B.1 that C_i^2 would be expected to be greater for those intervals.

By Equation B.2, the choices of $k_i = 9$ for $i = 1, 2, 5, 6$, or 7 and $k_i = 6$ for $i = 3$ or 4 in Subsection 2.C imply choices of 1/9 for each of C_1^2, C_2^2, C_5^2, C_6^2 and C_7^2 and 1/6 for C_3^2 and C_4^2.

Thus, the selected values for k_i are toward the low end of the range of inequalities B.8 and B.9. Still, the numbers $g_i(\hat{y}_{i,j})$ have a much greater effect than the numbers $f_{i,j}$ on the tails of the loss distributions selected by region in Subsection 2E.

APPENDIX C

METHOD OF FITTING PARETO

Iteration was used to find the single parameter Pareto distribution that minimizes

$$\sum_{i=1}^{4} ((f_i - P_i)^2 / P_i^{1.5}),$$

where f_i is as defined in Subsection 2E, and P_i is the corresponding fraction for the Pareto distribution.

The above method of fitting a Pareto to the numbers f_i is different, for theoretical reasons, from methods that would be used to fit a Pareto to actual frequencies. An explanation of the method is as follows.

Let the random variable X_i equal the f_i produced by performing the experiment of using the method of this paper on the data for the 41-year period. Assume that there is some Pareto distribution A such that each A_i, defined similarly to P_i, is the mean of X_i.

The Pareto that minimizes

$$\sum_{i=1}^{4} ((f_i - P_i) / \sigma_i)^2 ,$$

where σ_i is the standard deviation of X_i, is an estimate of A. If the probability distribution of X_i is Normal, then it is the maximum likelihood estimate of A.

If $P_i = A_i$, then based on the process used in computing the number f_i, it is judgmentally estimated that, for some constant c, each σ_i^2 equals approximately $cP_i^{1.5}$. Each f_i results from a weighting of actual data and a smoothed estimate. If only actual data were used, each σ_i^2

would be approximately in the same proportion to P_i. On the other hand, if only smoothed estimates were used to produce each f_i, and if the coefficient of variation were the same for each X_i, each σ_i^2 would be in the same proportion to P_i^2. The value $P_i^{1.5}$ was selected above because it is approximately midway between P_i and P_i^2. Thus the Pareto that minimizes

$$\sum_{i=1}^{4} ((f_i - P_i)^2 / P_i^{1.5})$$

is an estimate of the Pareto that minimizes

$$\sum_{i=1}^{4} ((f_i - P_i) / \sigma_i)^2.$$

A viable alternative method, which avoids the somewhat arbitrary choice of exponent on P, would be to use iteration to find the Pareto that maximizes the likelihood function $\Pi \, P_i^{f_i}$. This is numerically no more difficult than the approach used.