METHODS FOR FITTING DISTRIBUTIONS TO INSURANCE LOSS DATA

CHARLES C. HEWITT, JR. AND BENJAMIN LEFKOWITZ

VOLUME LXVI

DISCUSSION BY LEE R. STEENECK

"While this paper, so suggestive of an austere scholarship, may seem directed to those of the avant-garde who delight in frolicking among the outer reaches of actuarial theory, Mr. Hewitt presents both a challenge and a promise to those members whose interests, like this reviewer's, may gravitate more towards the application of actuarial principles to current underwriting and rating problems." And so Robert Hurley began his review of an earlier Hewitt paper (*PCAS* LIII) titled "Distribution by Size of Risk—A Model."

I, too, have always been *intrigued* by actuarial theory put into practice to solve rating problems. Certainly the body of the Hewitt/Lefkowitz paper deals primarily with the practical manipulation of a fitted loss distribution's cumulative function for the purposes of determining deductible discounts, increased limits factors, and relative frequency and severity.

Lest the sharp reader point an accusing finger at me already, perhaps I should explain my use of the verb "intrigue." Although Webster's preferred definition is "to cheat or trick," certainly I imply no diabolic intent to actuarial theory. Rather, my "interest is aroused."

I have some expertise with models used to price deductibles and increased limits factors. It was in 1976 and the then recently past "first" modern era of double digit inflation that actuaries working with the ISO felt that an improvement on the uniform excess and so-called "layer-of-loss" approaches to increased limits rate-making were justly deserved. I became involved and was a past chairman of the standing ISO committee dealing with commercial lines increased limits rate-making. I chose to review this paper, comparing it with what has been done in the last several years at ISO. Please keep in mind that our individual pursuits were totally independent, but it will be shown that our thoughts followed similar courses. Only in the practical manipulation of the curve do we, in fact, see differences.

The Increased Limits Subcommittee has been working with general liability data. Our initial curve-fitting efforts centered around the log normal distribution to be used on "live" medical malpractice data. As work progressed on other lines, we noted the log-normal failed to adequately describe various loss processes. Frequently, there were too many smaller losses for a good fit. A second distribution, the Pareto, was then developed. But again, we have not been able to totally explain the loss process by this single distribution. Perhaps we should be investigating compound distributions as well. Instead, we have chosen to truncate the Pareto from below at a value of, say, $5,000 and assign a single probability mass for all losses in the range $1 to $5,000. Since this falls within basic limits, the distribution in the range lacks importance by comparison to the "tail" probabilities by loss amount.

As a reinsurance actuary, I am constantly asked to evaluate large loss potential given that an insured has suffered a variety of smaller losses. This is a problem most of us face at one point or another. Even ISO, with its substantial data base, is missing detail on large losses by either (a) their non-occurrence (the fact that we have not seen a whole distribution of losses larger than $300,000 each) or (b) the tendency of primary policy limits to cap those losses which have occurred and are reported. Also, losses from excess and umbrella policies cannot be used with other raw data for fitting purposes. They are just not available in sufficient detail to use. From my point of view, this is primarily where the Hewitt/Lefkowitz paper interests me—predicting the tail of the distribution.

As we follow the curve to the right beyond the fitted area, say, up to $300,000 limits, and move into the unknown larger loss area, the choice of curve is of primary importance. Whether it be gamma, log-gamma, log-normal, Pareto, etc., we are speculating on some increased limits losses. Substantial actuarial judgment is required.

The authors have analyzed the "tail" problem in a manner similar to ours at ISO. Those losses that are at policy limits are said to be censored. A particular curve is fit in such a way that the number of policy limits losses are retained and are said to come from somewhere within the smooth extrapolation of the curve beyond policy limits. The reasonability test is: do all the tail frequencies of the fitted curve, when added together, compare favorably to the number of losses at policy limits?

This sort of fitting process is performed many times on data split by policy limits. At ISO it is called a multi-censored model. Naturally, the lesser the policy limit below $300,000, the greater the number of losses being censored.

As I mentioned before, I am concentrating on increased limits rate-making. The derived deductible credit columns on Tables 2 and 3 in the paper lend themselves extremely well to this. The tables, of course, are unitized, as should be expected for cumulative distribution functions. Layers of loss as a percentage of the total are calculated by subtraction. At ISO, we use a variation on this theme. We choose to use average policy limits losses. They reflect losses uncapped by policy limits as well as those capped at the policy limits. Then increased limits factors (based on expected value pricing) are determined by policy limits average loss plus average allocated loss expense plus unallocated loss expense divided by basic limits average loss plus the same average allocated loss expense plus unallocated loss expense. Neither type of loss expense was included for discussion by the authors.

For all of us who enjoy working with calculators, we can derive some pleasure in manipulating the columns for deductible credit on Tables 2 and 3. But before I demonstrate one principle, let us briefly investigate whether 100% inflation is realistic (Table 2 to 3). At first glance it might appear high, but try raising 1.15 to the fifth power! Yes, for losses emanating from the 1975 policy year, if trend is level at 15% per annum losses will double when on-level calculations are performed for policy year 1980 rates. And policy year 1975 experience is a most integral part of rates and increased limits factors being made for 1980. The other point to consider is whether all losses trend by the same percentage regardless of amount. ISO has assumed so, based on a few limited tests. But certain lines do exhibit an apparent increasing trend by size of loss which requires further study.

Given the reasonability of Tables 2 and 3, let us price the time differential. From Table 2 it can be demonstrated that if basic limits are \$10,000 then a \$250,000 policy (exclusive of LAE) should be rated at 2.81 times basic limits. Roughly 36% of the total cost is in the basic limits area, that is, \$100 for each \$281. With inflation, under Table 3, the increased limit factor (exclusive of LAE) is 3.62. Only 28% of the total limits premium is in basic limits now, that is, \$200 for each \$724. Note that even though inflation is 100%, the excess of basic premium has risen nearly 200%, that is, from \$181 to \$524. If you are not acutely aware of the part excess limits experience plays on your underwriting results, you should be. The leveraged effect of inflation is a point well worth remembering.

The cumulative frequency of cases columns are also interesting. Again, if basic limits are \$10,000, a full 82.15% of all cases had losses falling in basic limits (later to become 71.09% after inflation). The dollars of loss in basic limits during this period dropped from 25.88% to 17.54%. An analysis like this could be used to decide what limits should be called basic. Without raising the limit occasionally, the insurance principle of loss spread becomes much more dominant over equity for a class or territory.

The cumulative frequency of cases columns also have use for excess of loss reinsurers as they plan on claim staff size and other operational costs associated with servicing different retentions to different sized treaty reinsureds. One of the methods for selecting a reasonable retention has to do with reinsuring only a small proportion of the number of claims, i.e. the largest claims with the most effect on loss experience. Using relative frequencies on fitted curves, the excess of loss reinsurer could suggest reasonable increases in retention to simplify the dialog and paperwork associated with administering a less than reasonable reinsurance program.

Again, I thank Messrs. Hewitt and Lefkowitz for sharing with us the results of a no doubt time-consuming and expensive study. This paper should foster more ''intelligent competition'' in rates.