ESTIMATING CASUALTY INSURANCE LOSS AMOUNT DISTRIBUTIONS

GARY PATRIK

VOLUME LXVII

DISCUSSION BY STEPHEN PHILBRICK AND JEROME JURSCHAK

Gary Patrik has written a paper which is significant from several points of view. It provides:

- a well-conceived methodology for selecting a model for an empirical loss amount distribution;
- thoughtful remarks suggesting more than usual intuitive familiarity with the subject matter;
- a synthesis of a large body of existing literature interpreted to speak directly to the concerns of the actuary.

Mr. Patrik has discussed a number of reasons for seeking models for loss amount distributions. Successful model building requires a level of abstraction and understanding which goes beyond the mere analysis of data. Useful models have typically isolated those factors of marginal importance—the less cluttered the model, the more easily it can be communicated and the more likely cross-fertilization with other disciplines can be accomplished.

Since the K-S statistic is distribution-free and takes into account the natural ordering of the sample, it is a particularly useful goodness-of-fit test. However, the author states that it may be *too* powerful for certain actuarial considerations, since it has rejected (at the 5% level) all probability models yet tried. This observation is certainly not unexpected, partly for the reasons suggested by the author, but also due to clustering.

As practitioners, more than theoreticians, we know that real data rarely conform to the ideal. A number of arguments can be given as to why loss amounts cluster about particular levels. This observation is found frequently enough to be considered something of a norm for certain classes of business. This fact is a powerful argument in support of a statistical test which is less powerful than the K-S test. The χ^2 test, for example, is simple to use and easily communicated. While the choice of intervals is subject to manipulation, this liability can be an asset when dealing with the question of clustering. However, one must be careful, in any situation allowing manipulation, that adjustments which improve the fit can be realistically defended.

Another alternative is to modify the rejection percentile. Based on the expected discrepancies from the ideal, perhaps the K-S statistic should be used at other significance levels. In any case, the p-value (the smallest value of α for which the hypothesis would have been rejected) should be stated, thereby permitting different conclusions to investigators with differing qualitative assessments of the data itself.

It is also important to note a difference between the application of the K-S test to the simulated data in Appendix C and the application to the OL&T data. The simulated data consist of individual points, whereas the OL&T data is grouped (classified). Hoel (*Introduction to Mathematical Statistics*, p. 326) points out: " . . . the test then is no longer an exact one because the maximum difference for classified and unclassified data may not be the same; however, the discrepancy is usually slight if the classification is not too coarse." In the case of the OL&T data in Appendix E, Part 3, the first interval contains 41% of the data points. This is probably too coarse. However, if the point of the test is to compare the K-S statistic for competing distributions this may not be a problem.

One of the author's main conclusions is that the method of maximum likelihood should be used to estimate the parameters of the particular model. Although we agree with this conclusion, two points need to be stressed.

1. It must be recognized that comparison of method-of-moments estimates and the MLE estimates in Table 5.1 and Table 5.2 are not on the same basis. The MLE estimates are derived under the assumption that losses are censored. The method-of-moments calculations ignore this assumption. Hence, it is not surprising that the method-of-moments estimates are so poor. The author recognizes this fact, since he later states: ". . . we could compute correct method-of-moments estimates accounting for the policy limit censorship. But the equations that must be solved are much more complicated than the general equation (5.5)."
2. The maximum likelihood estimates for the parameters of the normal distribution are the sample mean and sample variance (Fraser, *Statistics—An Introduction*, p. 226). Hence, the MLE estimates are equivalent to the method-of-moments estimates. It then also follows that the method-of-moments applied to the logs of claim sizes (Method-of-Moments II

in Table 5.2 *if* it had been applied to the unlimited data) should be equivalent to the MLE estimates for the lognormal distribution. Note that this applies only to the unlimited distribution, not to a censored distribution.

The EVC test suggested by Mr. Patrik can be a very useful one. After all, as we are reminded, the expected value of loss is the most important component of most insurance premiums. One suggestion which would have the effect of making the computation of the vector statistic

$$\left\{\frac{G(x_i|\theta) - G_n(x_i)}{G(x_i|\theta)}\right\}$$

less cumbersome, and which would recognize the importance of policy limits, would be evaluation of the alternative statistic

$$\left\{\frac{G(P_i|\theta) - G_S(P_i)}{G(x_i|\theta)}\right\}$$

where the P_i, $i = 1, \ldots, L$ are some of the commonly used policy limits or retentions (such as \$100,000, \$250,000) and $G_S(P_i)$ is the sample average with censor P_i. It is the expected value of loss at policy limits which is a premier consideration.

Before ending with some comments on the use of the Pareto distribution, a few additional points will be discussed.

1. Our experience indicates that failure to modify data for trend and development before solving for the maximum likelihood estimates can produce future loss estimates differing significantly from those obtained with adjusted data. To the extent that IBNR losses tend to be larger than average, this would *partially* account for the observation that the data has too many small losses. However, note that even adjusting the individual claims for case development will not solve this problem. We suggest that unadjusted data be used for illustrative purposes only.
2. The author notes that the method-of-moments technique forces the value of δ to be greater than 2. This is a problem since typical values of δ are often less than 2. It should be noted that the single parameter Pareto with distribution function

   $$F(x) = 1 - x^{-\delta} \quad \delta > 1, x \geq 1$$

   has a less severe restriction, namely that $\delta > 1$. However, our experience generally indicates that the single parameter Pareto should be restricted to fitting excess losses, where the truncation point is approximately $10,000 or more.

3. In Table 6.1, the author fits the Pareto to the overall distribution and to the excess portion. The estimate of p for MLE I is .95 as shown in the table. It may be of interest to note that the value of p implied by MLE II, namely F(8000/347, .877) is equal to .9385. Although MLE II produces poor estimates of tail probabilities, it does a reasonably good job of estimating the proportion of losses less than the truncation point.
4. In Section VI, the author states that it is "convenient" to specify t (the truncation point at which the distribution splits into two distinct pieces) so that it is not an unknown. It should be pointed out that the choice of t is not an innocent one—different values of t can produce model estimates of tail probabilities which are quite different.

The final part of this review will deal directly with the Pareto distribution as a model for loss amount. While Mr. Patrik does not specifically advocate its use in any particular situation, he does state that both the ISO Increased Limits Subcommittee and he personally have found the two parameter Pareto very useful. The authors of this review have used the Pareto distribution to model large property and casualty losses in a wide range of circumstances including estimating property damage losses at large petrochemical complexes, forecasting corporate casualty losses excess of various self-insured retentions, pricing working cover excess of loss reinsurance, and establishing contributions to hospital trusts which serve to fund hospital professional liability losses. The particular model we have used is the single parameter distribution mentioned above.

In choosing to use almost exclusively the single parameter distribution, we have been guided by two considerations. First, its analytical form is simple enough to make the MLE parameter estimation routine ($\delta = n/\Sigma \ln x_i$) and to make accessible answers to such questions as sensitivity of forecast results to parameter value, the relationship between sample size and confidence in the parameter estimate, and the comparative impact on forecast losses of using unlimited, truncated, and censored distributions. Second, a single parameter gives a good fit to a variety of empirical data. For example, when fitting a one parameter Pareto distribution to the censored data in Appendix E, Part 1, the EVC statistic has components which range in magnitude from -6.04% to 2.72% (versus -5.60% to 1.71% for the two parameter model). This type of variation is small when compared to that inherent in the sampling distribution of δ itself. However, as mentioned earlier, the single parameter Pareto is generally appropriate only for excess losses.

Finally, we would like to discuss several areas in which additional research would be helpful.

1. Although the methodology for the calculation of MLE estimates of parameters should be well within the grasp of all actuaries, it might be the case that relatively few would spend the time necessary to pursue this concept. Is it possible that there are alternative methods which may sacrifice a little accuracy for a large savings in time and computation? For example, equating the 5th and 95th percentiles of the simulated data in Appendix C to the corresponding theoretical percentiles of the theoretical two parameter distribution and solving the resulting two equations yields parameter estimates $\beta = 28{,}339$ and $\delta = 1.623$ which in this case compare favorably to the actual values, as the probabilities that X is greater than 100,000 or 1,000,000 are .086 and .003 respectively. (See Quandt (1966) for additional discussion of this method.)
2. Suppose the estimates of parameters for a large set of data are calculated and also those for a small subset. For example, let the large set be all hospitals and the subset be a single hospital. Is it reasonable to derive parameter estimates for the single hospital by credibility weighting the two sets of parameters? If so, how does one determine the credibilities?
3. If parameters are estimated for various accident years, the values of the parameters will differ. To what extent can real changes in the shape of the distribution be measured by the changes in the parameter values? Equivalently, how sensitive are the parameter values to various sets of losses?
4. Can the concept of order statistics be used to draw inferences about the shape of the tail? For example, the expected largest loss from a finite sample generated by a Pareto distribution generally, in our experience, exceeds the greatest sample value. This may imply that the tail is too "thick," or possibly that a truncated (from above) Pareto is a better descriptor of reality.
5. In our experience, we have found that we can get reasonably good fits to loss data in excess of $25,000 with a one parameter Pareto (occasionally we split the distribution into two or more parts and estimate a sequence of parameters for a sequence of censored Pareto distributions). Although it is clear that two (or more) parameters are necessary to fit the distribution from ground zero, is it necessary for the distribution to have such a wide range? In many cases, an estimate of aggregate losses below some value will suffice; in other cases a different distribution may be a better choice for small losses. It may sound more complex to have two

distributions, one for losses up to a truncation point t, and another for losses in excess of t, but in fact the estimation of parameters may be easier.

Finally, we would like to make it clear that we do not advocate abandonment of a two parameter Pareto model. Anyone with the computer procedures for this distribution will certainly get good use out of them. We are merely suggesting to those without such techniques already developed, that there may be several suitable alternatives.

The following is a short extension to the bibliography in Mr. Patrik's paper.

Benktander, G. (1961). "On the Correlation in Results from Different Layers in Excess Reinsurance," *Astin Colloquium* 1961, pp. 203–209.

Haung, J. S. (1975). "A Note on Order Statistics from Pareto Distribution," *Scandinavian Actuarial Journal*, pp. 187–190.

Lwin, Thaung (1974). "Empirical Bayes Approach to Statistical Estimation in the Paretian Law," *Scandinavian Actuarial Journal*, pp. 221–236.

Malik, H. J. (1970). "Estimation of the Parameters of Pareto Distribution," *Metrika*, Vol. 15, pp. 126–132.

Quandt, R. E. (1966). "Old and New Methods of Estimation and the Pareto Distribution," *Metrika*, Vol. 10, pp. 55–82.