ary economy increases are more common than decreases. The disutility of too little credibility is much greater in the case of statewide rate level than in the case of territory relativity calculations, since the former can lead to a dollar loss on all units while the latter results only in the lost opportunity for profit on some units. In other words, the mathematical and statistical tools cannot be separated from the decision procedure — the business situation — itself. If one accepts this premise, then the selection of P and k becomes an exercise in statistical decision theory once the disutilities of the various outcomes are postulated. This latter step is, of course, a subjective one and can be performed only if one accepts the subjective, or Bayesian, view of probability. Apparently, the classical Neyman-Pearson solution of credibility problems is only of use to Bayesian statisticians.

DISCUSSION BY JOHN S. McGUINNESS

This is a fundamental paper of great significance. Besides the concrete advance it explicitly reports, it also has several substantial implicit qualities. These are easy to miss but they are worthy of specific recognition for the valuable instruction they provide.

Primarily, the paper provides a sound and distribution-free theoretical basis that permits the development of a specific criterion (a precise number of claims) for ascribing full credibility to data that reflect both relative claim frequency and average claim cost for a single-parameter class of risks. It appears to provide an objective basis for overcoming two present and contrasting deficiencies in ratemaking practice. The first deficiency is the current neglect of the average claim cost element (and its variability) in determining a class credibility, and the second is the present common neglect of the relative frequency element (and its variability) in making time-series adjustments.

The author's approach, in using the total amount of claims (T) as the major variable of which the variation is measured, is a less easily visible quality. It is a clever and rewarding change in viewpoint. By focusing their attention on this aggregate or collective figure, rather than on the much smaller pure premium, they have for their analysis a statistically much more manageable datum and one for which it is much easier to determine a mean and an objective measure of variability. This difference in approach appears to parallel precisely the difference between approaching single-auto merit rating as (a) measuring and basing rates on the experience of an individual automobile or (b) measuring and basing individual auto rates on the experience of aggregates or classes of risks, where the claim experience of each class has defined characteristics that differ from those of other classes.

This difference in approach has made unit-risk merit rating demonstrably sound actuarially, and it has made development of a fully representative (of both frequency and mean severity) credibility criterion by the authors demonstrably feasible actuarially. It is the mental equivalent of a military outflanking movement, permitting attack from the side or the rea when frontal attack fails. The potential of this tool — shifting one's view point or perspective to find a solution to a problem — as an aid in making similar major strides in future actuarial progress deserves explicit recognition which it does not appear generally to have received.

The paper is important not only for what it accomplishes but also for what it seems to promise. The full credibility criterion it sets forth is not linked in the paper to any criteria for partial credibility. But one such link and set of criteria which suggest themselves involve a slight change, a change linked to the "asymptotic" approach of Prof. Mayerson's previous paper.[1] Once the maximum permitted percentage deviation of the observed pure premium from the expected pure premium (Perryman's k) is set, the probability P that the observed pure premium will differ no more than $100\,k\,\%$ from the expected pure premium can be used as the credibility. An infinite sample size N would be needed to give full credibility under this approach, but the credibility of large samples measured by this criterion would approach so close to 100 per cent that it would be feasible to treat any sample with a P exceeding a set threshold as having full credibility. The threshold could be set at 99%, 95%, or the 90% used by the authors.

It will be clear to the reader that this "threshold" can with equal reason and logic be considered

(1) the Perryman and the present authors' criterion or critical point for full credibility, it being also the logical point of departure for an (in the paper) undefined but potentially multi-valued set of criteria for degrees of partial credibility; also

[1] Allen L. Mayerson, "A Bayesian View of Credibility," *PCAS* Vol. LI, p. 85.

(2) the confidence level for a classical two-valued Neyman-Pearson decision concerning a hypothesis in which the only two permitted credibility values are zero (rejection) and 100 (acceptance); and further

(3) a full-credibility threshold for a precisely defined, multivalued, set of objective criteria for all degrees of credibility (in which the credibility or $z = P$).

The paper, with its collective approach, also has important underwriting and reinsurance, as well as ratemaking, implications. The quantitative objective of an underwriter is, in conjunction with maximization of total profit, to keep adverse fluctuations in the total amount of actual claims within a given maximum percentage of the expected total, to a given degree of probability, P. The present authors' statement of the Perryman full credibility criterion says exactly the same thing, but in ratemaking rather than in underwriting terms.

Because of this parallel, the criterion presented in the paper can immediately be extended beyond its single class, single time period limits. Rosenthal[2] has shown how, in types of insurance giving rise only to total losses, such a criterion can be determined for a combination of classes of risks with different mean sizes of loss (in his scheme, amounts of insurance) and different relative claim frequencies. His work has been extended[3] to embrace as well all the types of insurance that give rise to partial losses, to different rating territories and time periods, and to all other types of differences in the characteristics of various classes of risks; and also to handle the contagion or catastrophe (interdependence of risks) problem.

These two papers had also gone beyond another of Perryman's limitations. They removed the need to assume equality between mean and variance. Their common approach permits a skewed (for example, a positive or negative binomial as well as Poisson) or unskewed (normal curve as a limiting form of the binomial) type of distribution. The present paper, however, appears to provide for a much more general set of possible types of distribution. This is another very important contribution.

---

[2] Irving Rosenthal, "Limits of Retention for Ordinary Life Insurance," *Record* of the Institute of Actuaries, May 1947, pp. 8-22.

[3] John S. McGuinness, "Controlling the Effects of Catastrophes in Insurance Against Floods and Other Elemental Perils," *Transactions XV International Congress of Actuaries* Vol. IV, p. 190.

One may take but minor exceptions to the present paper. One exception is to eliminate the N from assumption "(b) that the random variables $N, X_1, X_2, \ldots$ are independent . . . ," since once that X_N is chosen, N is determined. Secondly, it would be clearer if it were stated whether the automobile property damage data were for total limits or were truncated to a standard limit. It would also be useful to mention the need, due to inflationary and other temperal changes, of frequently recalculating the size of N for each type of insurance and for at least some breakdowns or subclasses thereof. This might have to be done as often as yearly in order to produce an actual variance as small as that postulated from prior data.

The minor nature of these points simply tends to confirm an opinion that recognizes these authors' paper as a scholarly, clearly presented, and very important contribution.

DISCUSSION BY KENNETH L. McINTOSH

Assuming that necessary data could be made available, the only argument against recognition of claim cost variation in credibility calculations seems to be one advanced by Mr. Perryman himself. He noted that the resulting "great increases in credibility requirements could not very well be made in practice under present day conditions for they would greatly limit the employment of local data."[1] Mr. Perryman's "present day conditions" of 1932 are not, however, the "present day conditions" of 1969. Messrs. Mayerson, Jones, and Bowers have refocused attention upon the question, and perhaps the argument will bear re-examination. The data problem should not prove insoluble if it once is decided that the hidden cost of deficient credibility standards exceeds the out-of-pocket incident to data collection and processing.

When full credibility is defined by $P=90\%$, $k=0.05$, it is doubtful that retention of the 3rd and higher claim cost moments results in any significant increase in accuracy, except possibly in extreme cases. Assuming $\lambda_3=\lambda_2=\lambda$, then Eq. (E) of the paper becomes:

$$\lambda = \frac{Z_e^2}{k_1^2}\left[1+\frac{\mu_2}{\mu^2}\right] \tag{1}$$

[1] Perryman, F. S., "Some Notes on Credibility," *PCAS* Vol. XIX (1932), p. 73.