DISCUSSION BY CHRISTIAN SVENDSGAARD AND PAUL BRAITHWAITE

INTRODUCTION

Having worked on Empirical Bayes credibility for a combined total of over ten years, we share Mr. Buck's frustration at the slow acceptance in practice of Empirical Bayes techniques. Part of the reason, we believe, is the inherent conservatism of the insurance business. Considering the sums at stake, practicing actuaries are reluctant to adopt new methods until they have been thoroughly researched and tested. It is gladdening, then, to see further discussion of credibility in the *Proceedings*. Only after undergoing thorough scrutiny can new methods hope to be adopted in practice.

Mr. Buck has written a paper that considers one aspect of credibility—Stein estimation—from a theoretical point of view. Lay actuaries hoping to see a comparison based on real data of an Empirical Bayes credibility procedure and (say) the square root rule must look elsewhere. But Mr. Buck's paper could still have relevance to lay actuaries. If a method can be shown to be theoretically incorrect, there is no reason to test it on real data.

While we applaud further exchange of ideas on credibility, we find parts of Mr. Buck's fundamental approach, and several of his conclusions, problematic. Our thoughts regarding his approach can be summarized as follows:

1. There are three schools of statistical thought: classical, Bayesian, and Empirical Bayesian. A case may be made for one school or another on philosophical grounds, or possibly on practical grounds. But from a mathematical viewpoint, arguing against one school based on the assumptions underlying another begs the question. This, Mr. Buck has done, treating Stein (classical) estimation from a Bayesian viewpoint.

2. The circle distribution example is interesting but not really relevant to actuarial problems.

More specific criticisms are:

1. In the normal case, the mean is the Bayes estimator *only* for a flat prior. In that case, the Stein estimator approaches the mean with probability one. (The concept of a "flat prior" is an attempt to extend the concept of a uniform prior distribution to an infinitely large parameter space. This is done by examining a sequence of uniform prior distributions, each covering a larger area of the parameter space. Loosely speaking, a flat prior gives the Bayes estimate if there is no prior belief.)

2. The circle distribution example, besides suffering from irrelevance, has the same problems as the normal case.

3. The adapted Morris-Van Slyke procedure, while akin to the Stein estimator, is based on Empirical Bayes rather than classical ideas. While Mr. Buck is correct in pointing out the bias in the procedure, the bias is due to the logical constraint that estimates of variances should not be negative. Because the bias is non-linear, it cannot be corrected by a linear transformation of the estimate. By making distributional assumptions, it might be possible to construct an unbiased estimator. However, the procedure would be valid only in situations where the posited distribution held and would lose its generality. In practice, Empirical Bayes credibility procedures have been applied to loss ratios. The distributional properties of loss ratios are complicated and it seems unlikely that an unbiased Empirical Bayes estimator could be constructed based on a realistic loss ratio distribution.

4. Tests on simulated data show that the $3/k$ adjustment factor that Mr. Buck criticizes should be used whether or not individual classes are trending at different rates from one another.

We explain and elaborate on these comments below.

COMMENTS ON THE APPROACH

Currently, there exist at least three schools of statistics: classical, Bayesian, and Empirical Bayesian. Each school makes different assumptions. Bayesians assume a prior distribution; classicists do not. Empirical Bayesians assume the parameters of the prior distribution are unknown; Bayesians do not.

It is easy to "prove" that one school is wrong by examining it from the viewpoint of another school. However, this makes no more sense than "proving" non-Euclidean geometry is wrong by making Euclidean assumptions. ("Assuming the parallel postulate holds, then any geometry where it does not hold has a contradiction. Therefore the parallel postulate holds.")

One set of assumptions may be more useful than another, because it fits reality better. Prior to general relativity, non-Euclidean geometry was an interesting curiosity. Afterwards, non-Euclidean geometry became *the* geometry. It is conceivable that *in practice* (say) Bayesian estimators will always perform best. The data could tell us which school of statistics is right. But it cannot be decided *a priori*.

The approach we advocate for selecting estimators in practice is:

(1) Selection of reasonable models;
(2) Testing of the model assumptions using the data;
(3) Derivation of estimators based on the models; and
(4) Testing of the estimators using the data.

The different schools of statistical thought might select different models, and testing of the model assumptions might not eliminate any of the models. This seems especially true of (pure) Bayesian models, which incorporate prior belief. But the various estimators derived will yield different results when tested on the data. Given enough data, one estimator will prove most attractive.

One of the major themes of Mr. Buck's paper is an argument that, from a certain Bayesian viewpoint, Stein estimators (i.e., classical estimators) do not make sense. He presents no empirical data. In our view, this argument is no more convincing than the argument against non-Euclidean geometry from a Euclidean point of view.

While you cannot "prove" one school is wrong from the point of view of another, it may be that the assumptions underlying one school are self-contradictory. Mr. Buck hints that the disturbing property of the Stein estimator, that it is not translation-invariant (i.e., that for a given data point, the Stein estimator could be anywhere, depending on the location of the origin you are shrinking toward) is such a contradiction.

While the non translation-invariance of the Stein estimator is disturbing, Mr. Buck has not shown that it is paradoxical. As a footnote, the Morris-Van Slyke and Bühlmann-Straub Empirical Bayes credibility procedures *are* translation-invariant. This is accomplished by shrinking towards the group mean, rather than the origin.

Mr. Buck attempts to illustrate the failings of Stein estimation by means of a similar estimator derived for the circle distribution. Reasoning by analogy is, of course, inappropriate in a mathematical context. The success or failure of the illustration must therefore be judged on its effectiveness as a pedagogical device. In our case, at least, we were not convinced by the illustration.

SPECIFIC COMMENTS

Inadmissible Estimator is Bayes Only For Flat Prior

We argue above that it is incorrect to criticize a classical estimator by making Bayesian assumptions. You will not find us making the converse mistake here. However, we do wish to show that in the normal case the mean is not a Bayes estimator except in the case of a flat prior.

We believe this is an important fact because, strictly speaking, a flat prior is not a prior distribution at all. The concept of a flat prior is based on a sequence of ever-flatter distributions. For any distribution in the sequence, the mean is not the Bayes estimate; this follows from Stein's result, as we will show below. Thus, while it is true that in the limit the mean is the Bayes estimate, it is not true for any intermediate point.

Moreover, in the limit, the Stein estimator approaches the mean with probability one. This means that under a flat prior the mean is not better than the Stein estimator: it is essentially equal to it.

Proof

We are attempting to estimate an (at least) 3-dimensional vector of means, θ, given a vector of observations, X, distributed normally around θ with covariance matrix, I, the identity:

$$X \sim N(\theta, I).$$

For an estimator $\hat{\theta}$ of θ, the squared error loss is

$$\text{Loss} = L(\hat{\theta},\theta) = |\hat{\theta} - \theta|^2$$

The risk is the expectation given θ:

$$\text{Risk} = R_\theta(\hat{\theta}) = E(L(\hat{\theta},\theta)|\theta)$$

The Bayes risk is

$$B(\theta) = E_\theta(R_\theta(\hat{\theta})) = \int_\Omega R_\theta(\hat{\theta})\, dF(\theta)$$

where E_θ denotes "expectation with respect to θ," the integral is over all possible values of θ (i.e., the sample space is Ω), and $\int \ldots dF(\theta)$ denotes Riemann-Stieltjes integral where $F(\theta)$ is the CDF of the prior of θ.

When Stein proved that the mean is not admissible, he proved it by showing that

$$R_\theta(\hat{\theta}_{\text{Stein}}) < R_\theta(\hat{\theta}_{\text{Mean}})$$

for all θ's, where $\hat{\theta}_{\text{Stein}}$ denotes the Stein estimator, and $\hat{\theta}_{\text{Mean}}$ denotes the mean (i.e., X). (See "Estimation with Quadratic Loss," p. 363 [2].)

The size of the difference

$$R_\theta(\hat{\theta}_{\text{Mean}}) - R_\theta(\hat{\theta}_{\text{Stein}})$$

depends on the value of θ. It is greatest at the origin (which makes sense, since the Stein estimator shrinks the estimate towards the origin). It decreases as θ moves away from the origin, but it is always positive.

How does the Bayes risk of $\hat{\theta}_{\text{Mean}}$ compare to the Bayes risk of $\hat{\theta}_{\text{Stein}}$?

$$B(\hat{\theta}_{\text{Mean}}) - B(\hat{\theta}_{\text{Stein}}) =$$

$$\int_\Omega R_\theta(\hat{\theta}_{\text{Mean}})\, dF(\theta) - \int_\Omega R_\theta(\hat{\theta}_{\text{Stein}})\, dF(\theta) =$$

$$\int_\Omega [R_\theta(\hat{\theta}_{\text{Mean}}) - R_\theta(\hat{\theta}_{\text{Stein}})]\, dF(\theta).$$

The expression inside the integral

$$R_\theta(\hat{\theta}_{\text{Mean}}) - R_\theta(\hat{\theta}_{\text{Stein}})$$

is greatest at the origin and decreases as θ moves away from the origin. But it is always positive. This is what Stein proved.

The value of the integral will depend on $F(\theta)$, the prior distribution. The more weight given to θ's away from the origin, the smaller the integral will be. *But it will always be positive.*

Since the Bayes estimator minimizes Bayes risk, the mean cannot be the Bayes estimator for any prior. Only by taking the limit of distributions throwing more and more weight away from the origin can the mean be made to approach the Bayes estimate.

Note, however, that in the limit the difference in Bayes risk is *zero*. The mean and the Stein estimator are equal in the limit. In the expression

$$\left(1 - \frac{n-2}{|\vec{x} - \vec{p}|^2}\right)\vec{x} + \left(\frac{n-2}{|\vec{x} - \vec{p}|^2}\right)\vec{p},$$

$|x - p|^2$ is greater than 100 (1,000,000, 10^{26}, ...) with probability .99 (.999, $1 - 10^{-500}$, ...) in the limit: the Stein estimator reduces to the mean $\bar{x}$.

A careful re-reading of the above proof should convince the reader that it can be made entirely general. An inadmissible estimator cannot be a Bayes estimate. In other words, Bayes estimators are admissible. (See [3].)

Generality of the Circle Distribution Example

To repeat our main concern, this is not relevant. If normal-distribution Stein estimation has faults, they cannot be discerned by examining circle distribution Stein-like estimators.

We showed earlier that Bayes estimators are admissible. Mr. Buck claims to "prove" that his Stein-like estimator dominates the mean (in Appendix I). He then derives the mean as a Bayes estimator. This is a contradiction due to the use of a flat prior. Note that in Appendix I, he claims to show that the Stein-like estimator has risk less than one. Then he shows that the risk of the mean is one. In the limit, the Stein-like estimator is the mean.

There is also a mistake in the derivation of the Stein-like estimator in Appendix I. The quantity c is treated as a constant in all the integrals—but at the end "... we chose $c \geq (|\vec{p} - \bar{x}| + 1)^2$ ...," i.e., c depends on $\bar{x}$.

ISO's Empirical Bayes Credibility Procedure

Mr. Buck says that the Morris-Van Slyke Empirical Bayes credibility procedure is "based on 'Stein estimation'." This is not entirely accurate. While the Morris-Van Slyke procedure is similar to, and to an extent suggested by, the Stein procedure, it is developed in an Empirical Bayes framework. In fact, Efron and Morris, in [4], show that the Stein estimator itself can be developed as an Empirical Bayes estimator.

Mr. Buck states that the adapted Morris-Van Slyke procedure with the $3/k$ factor is biased upwards and that, because the tests included groups where the expected class loss ratios trended up or down over time, the results were slanted in favor of the $3/k$ factor. While the testing included the "residual trend" case, the original testing was done on the no residual trend case. For instance, in [5], page 79 ff., among other things, the adapted Morris-Van Slyke procedure is tested on simulated data against the same procedure *without* the $3/k$ factor where no residual trend is in effect. The with-$3/k$ procedure does better than the without-$3/k$ procedure in 86 out of 110 cases.

For instance, Table 1 reproduces the results given in [5] of simulated consecutive reviews for six different groups of simulation parameters. The error ("premium weighted test statistic," which is defined as the premium-weighted sum over all classes of the squared difference between the class loss ratio after the rate change and the expected loss ratio, see [6] p. II-15) is shown for the first through fifth reviews after the implementation of the new credibility procedure. Each entry is the average of 21 independent simulations.

As a footnote, group 4 was constructed with a very low original between-variance. This is why the procedure without the $3/k$ correction did better—lower credibilities were called for.

Mr. Buck says that Empirical Bayes credibility procedures using the $3/k$ correction factor are biased. The adapted Morris-Van Slyke procedure is biased, but not due to the $3/k$ factor. The bias is caused by logical constraints imposed on the variance estimators and there are good Empirical Bayesian reasons for these constraints.

The credibility formula depends on using an estimate of between-variance (parameter variance) in the denominator. Even though the between-variance estimator is unbiased, the credibility is not unbiased, because the credibility is not a *linear* function of the between-variance. To correct this, the indicated credibility is adjusted as follows:

$$Z^{\text{Adjusted}} = \frac{k-3}{k} Z^{\text{Indicated}} + \frac{3}{k}$$

where k is the number of classes. The derivation of this bias correction is given in [6].

At this point, we have an unbiased estimate of the credibility. Technically, the procedure is only unbiased for a highly restrictive set of assumptions. But, even without these assumptions, the bias correction is in the right direction.

Unfortunately, this unbiasedness depends on allowing the estimate of the between-variance to be negative. While the explanation of this [6] is complicated, the reader can see this intuitively by looking at the above equation and trying to imagine what value of $Z^{\text{Indicated}}$ is necessary for a class whose "true" credibility is less than $3/k$.

A priori, negative between-variances are impossible. Restricting the estimate of the between-variance to be non-negative leads to the minimum credibility of $3/k$ that Mr. Buck mentions. Unfortunately, restricting the estimate of the

TABLE 1
COMPARISON OF MORRIS-VAN SLYKE PROCEDURE WITH AND WITHOUT THE $3/k$ FACTOR
(ERRORS $\times\ 10^{-3}$)

Group	Review	Error With $3/k$	Error Without $3/k$	% Reduction in Error With $3/k$
1	1	2.30	2.43	5.3
	2	2.11	2.39	11.7
	3	1.84	2.25	18.2
	4	1.66	1.97	15.7
	5	1.47	1.69	13.0
2	1	1.49	1.58	5.7
	2	1.43	1.53	6.5
	3	1.38	1.52	9.2
	4	1.25	1.49	16.1
	5	1.14	1.23	7.3
3	1	2.61	2.67	2.2
	2	2.40	2.67	10.1
	3	2.18	2.62	16.8
	4	1.96	2.53	22.5
	5	1.66	2.11	21.3
4	1	0.72	0.66	−9.1
	2	0.83	0.65	−27.7
	3	1.03	0.74	−39.2
	4	1.16	0.73	−58.9
	5	1.19	0.72	−65.3
5	1	2.59	2.55	−1.6
	2	2.26	2.51	10.0
	3	2.09	2.37	11.8
	4	1.89	2.23	15.2
	5	1.77	2.21	19.9
6	1	3.10	3.40	8.8
	2	3.04	3.43	11.4
	3	3.03	3.32	8.7
	4	3.10	3.21	3.4
	5	2.62	2.99	12.4

between-variance also biases the procedure. Because the bias is a non-linear function of the between-variance, it cannot be corrected for all values of the between-variance. The bias is less in cases where there are many classes (k is large) and the underlying between-variance is high (so the probability of a negative between-variance estimate is low). These two conditions tend to hold in ISO's Products review.

As a practical matter, we do not recommend applying Empirical Bayes procedures if k is less than six. At least for the current generation of Empirical Bayes procedures, there simply aren't enough degrees of freedom to get a good estimate of the between-variance when there are five or fewer classes.

We do not regard current Empirical Bayes credibility procedures as the final word. Empirical Bayes procedures are, in general, the best credibility procedures we've seen thoroughly tested. However, other families of techniques offer hope for greater accuracy, particularly when data are classified by more than one rating variable (e.g., class and territory). As estimates of loss costs improve, insurers, insured, and society as a whole receive benefits. By focusing interest on improving credibility procedures, Jim Buck has performed a valuable service.

REFERENCES

[1] James E. Buck, "On Stein Estimators: 'Inadmissibility' of Admissibility as a Criterion for Selecting Estimators," *PCAS* LXXII, 1985.

[2] W. James and C. Stein, "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, University of California Press, Berkeley, 1961, pp. 361–379.

[3] Thomas S. Ferguson, *Mathematical Statistics: A Decision-Theoretic Approach,* New York, Academic Press, 1967.

[4] Bradley Efron and Carl Morris, "Stein's Estimation Rule and Its Competitors—An Empirical Bayes Approach," *JASA*, March 1973, Volume 68, Number 341.

[5] Minutes, ISO Credibility Subcommittee Meeting of December 13, 1979.

[6] Insurance Services Office, Report of the Credibility Subcommittee: Development and Testing of Empirical Bayes Credibility Procedures for Classification Ratemaking, September, 1980.