# Comment

**William E. Strawderman**

This article was a pleasure for me to read. I thank the authors for giving the time and effort required to present a thorough and efficient development of the key decision-theoretic aspects of variance estimation.

I would like to elaborate somewhat on the authors' comments concerning practical improvements by drawing some parallels with the problem of estimating a mean vector. Additionally, I would like to briefly discuss the problem of estimating variance components.

Consider the model of the second paragraph of Section 5. For simplicity I'll assume $\sigma^2 = 1$ when discussing estimation of $\mu$ with loss equal to $\| \delta - \mu \|^2$. In both problems (estimating $\mu$ and estimating $\sigma^2$), the potential gains in risk over the best equivariant estimator increase with the dimension $p$ of $\mu$. For the case of estimating $\mu$, the James–Stein estimator

$$\left(1 - \frac{p-2}{X_2' X_2}\right) X_2$$

(here $p \geq 3$) has minimum risk equal to 2 when $\mu = 0$ rising as a function of $\| \mu \|^2$ to $p$ as $\| \mu \|^2 \to \infty$. Hence the maximum relative savings is $(p - 2)/p$ which increases from ⅓ to 1 as $p$ increases. The larger the dimension $p$, the more one is able to "borrow strength" across coordinates. In this case, as well as for the estimator (5.1) of the variance, meaningful gains (for large $p$) will result only if our prior guess (of $\mu = 0$) for the true mean vector is accurate. In either case, if our guess is poor and $\| \mu \|$ is large, $X_2' X_2$ will be large with high probability. Hence both the James–Stein estimator and (5.1) will, with high probability, be equal (or close) to the respective best equivariant estimators and little improvement will result.

It is still possible to achieve meaningful gains in the two problems (for large $p$) even if we are unable to accurately guess the true value of $\mu$. For example, if we (accurately) guess that all of the components are equal but can't guess what the common value is, the Lindley–Smith estimator of $\mu$ (see Lindley and Smith, 1972) will work well. This estimator shrinks all components of $X_2$ toward the average of these components and results in an estimator of $\mu$ with risk equal to 3 when all components of $\mu$ are exactly equal and increases to $p$ as the variance of $\{\mu_1, \cdots, \mu_p\}$ tends to $\infty$. Similarly in the problem of estimating $\sigma^2$ an estimator similar to (5.1) but with

$$Y^2 = \| X_2 - \bar{X}_2 1 \|^2$$

and

$$\phi_S(Z) = \min\left(\frac{1}{\nu + 2}, \frac{1 + Z^2}{\nu + p + 1}\right)$$

will achieve meaningful gains for $p$ large relative to $\nu$ provided the coordinates of $\mu$ are all equal. Again gains will not be substantial unless the prior guess is accurate.

Similar parallel estimators are possible in the two problems based for example on guesses that the mean vector $\mu$ lies in a given subspace and the relative gains will be large provided the co-dimension of the subspace is large and the prior guess is good.

Of course, the parallels in the two problems are imperfect. In the problem of estimating the mean vector, the dimensions of the estimator and estimand are also growing as is the minimax risk. In the variance estimation problem, the corresponding values are fixed. In the variance estimation problem if $\nu$ is fixed, $p \to \infty$, and $\mu = (0, \cdots, 0)$ the estimator (5.1) $\to$ $\min(S^2/(\nu + 2), \sigma^2)$, and hence the relative savings in risk do not approach 1 even when the prior guess for $\mu$ is correct.

I believe it is also worthwhile to mention some work in variance component estimation that is directly related to Stein's method. Klotz, Milton and Zacks (1969) (KMZ) studied estimation of $\sigma_e^2$ and $\sigma_a^2$ for squared error loss in the balanced random effects one way layout.

Sufficiency reduces the problem to consideration of

$$Y \sim N(\mu, \sigma_e^2 + J\sigma_a^2/IJ),$$

$$S_a^2 \sim (\sigma_e^2 + J\sigma_a^2)\chi_{I-1}^2$$

and

$$S_e^2 \sim \sigma_e^2 \chi_{I(J-1)}^2,$$

where $Y$, $S_a^2$ and $S_e^2$ are independent. One aspect of this problem that is of interest as it contrasts with the fixed effects model of Section 5 is that location and scale invariance are not enough to produce a best equivariant estimator.

The fully invariant estimators of the $\sigma_e^2$ are of the form $S_e^2 g(S_a^2/S_e^2)$. "Reasonable" estimators in this

*William E. Strawderman is Professor and Chair, Department of Statistics, Hill Center, Busch Campus, Rutgers University, New Brunswick, New Jersey 08903.*

class include the UMVUE ($S_e^2/I(J-1)$) and the

$$\mathrm{MLE}\left(\min\left(\frac{S_e^2}{I(J-1)}, \frac{S_e^2+S_a^2}{IJ}\right)\right).$$

KMZ use a version of Stein's method to show that the MLE dominates the UMVUE. The authors go on to show, again using Stein's method, that any equivariant estimator of $\sigma_e^2$ which is greater than the sum of squares of all observations divided by $IJ + 2$ (i.e., $[IJY^2 + S_a^2 + S_e^2]/(IJ+2)$) with positive probability, is inadmissible. This last result implies that there is additional information about $\sigma_e^2$ in the overall mean $Y$ even though the variance of $Y$ is a multiple of $\sigma_e^2 + J\sigma_a^2$ and not of $\sigma_e^2$.

Estimation of $\sigma_a^2$, of course, involves the additional wrinkle that the

$$\text{UMVUE } E\left(\frac{1}{J}\left(\frac{S_a^2}{I-1} - \frac{S_a^2}{I(J-1)}\right)\right)$$

is negative with positive probability. KMZ use Stein's method to investigate dominance relations among several estimators and show that the overall mean can sometimes be used to construct improvements. Portnoy (1971) and others have constructed Bayes equivariant estimators with good sampling properties. Loh (1986) has studied the problem of estimating $\sigma_a^2/\sigma_e^2$ using similar methods.

In this setup, since there is only one degree of freedom for the grand mean, the likely improvement is small (once one has selected a good equivariant estimator). It is possible that larger gains could occur in higher way mixed models where several degrees of freedom are available for the mean vector.

Presumably extensions of Brown's and Zidek's methods can be applied in these models and improved confidence intervals can be constructed as well.

## ADDITIONAL REFERENCES

KLOTZ, J. H., MILTON, R. C. and ZACKS, S. (1969). Mean square efficiency of estimators of variance components. *J. Amer. Statist. Assoc.* **64** 1383–1402.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 1–41.

LOH, W.-Y. (1986). Improved estimators for ratios of variance components. *J. Amer. Statist. Assoc.* **81** 699–702.

PORTNOY, S. (1971). Formal Bayes estimation with application to a random effects model. *Ann. Math. Statist.* **42** 1379–1402.

# Rejoinder

**Jon M. Maatta and George Casella**

To begin, we thank all discussants for their kind remarks and stimulating comments. This project was started to enhance our understanding of the topic, but also helped to improve our knowledge and perspective. As mentioned by several discussants, the scope of our work was limited. This work was an intentional decision, because our relatively narrow focus presented a reasonable size task, and allowed us a fuller understanding of one part of this complicated subject. Many of the discussants had similar concerns, and we will structure our rejoinder to respond to the major topics mentioned.

### PRACTICAL CONSIDERATIONS

Though somewhat surprising to us, much concern was expressed over the magnitude of possible improvement. A major point was that the possible improvements in variance estimation seem small when compared to those possible in the estimation of means. This is true, but we feel that the improvement here is still worthy of consideration.

Berger expresses concerns about this and, in his inimitable way, anticipates some of our rejoinder. While the magnitude of improvement is small (as demonstrated by other discussants), it does increase in the generalized linear model case, which we do not consider "less realistic," but useful in practice. Very interesting calculations are provided by both Hwang and Rukhin, showing the limiting amount of improvement possible, approximately 25% in practical cases. Rather than interpreting these findings in the pessimistic way of Professor Hwang, however, we find more hope for future improvements (although we certainly agree that greater improvement seems possible in the estimation of means).

Some of our optimism is supported, and Hwang's pessimism negated, by the comments of George and Strawderman. They suggest that we have not yet fully exploited the structure of the problem. The risk (or interval length) improvement in variance estimation obtains when the means are close to the point to which we are shrinking. George and Strawderman each point out ways to shrink toward subspaces, and, further, George suggests that we can shrink toward multiple subspaces. Such estimators may provide substantial practical gains, since the region of improvement will be expanded. Another interesting possibility