OMAN, S. D. (1984). A different empirical Bayes interpretation of ridge and Stein estimators. *J. Roy. Statist. Soc. Ser. B* **46** 544–557.
PERNG, S. K. (1970). Inadmissibility of various "good" statistical procedures which are translation invariant. *Ann. Math. Statist.* **41** 1311–1321.
RUBINSTEIN, R. and MARKUS, R. (1982). *Improved estimation using control variables*. Report, Technion University, Haifa, Israel.
SAVAGE, L. J. (1976). On rereading R. A. Fisher (J. W. Pratt, ed.) *Ann. Statist.* **4** 441–500.
SCLOVE, S. L., MORRIS, C. and RADHAKRISHNAN, R. (1972). Nonoptimality of preliminary test estimators for the multinormal mean. *Ann. Math. Statist.* **43** 1481–1490.
STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. California Press.
STEIN, C. (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann. Math. Statist.* **30** 970–979.
STEIN, C. (1960). Multiple regression. In *Contributions to Probability and Statistics Essays in Honor of Harold Hotelling* (I. Olkin, et al., eds.) 424–443. Stanford Univ. Press.
STEIN, C. (1973). Estimation of the mean of a multivariate distribution. In *Proc. Prague Symp. on Asymptotic Statist.* 345–381. Charles Univ., Prague.
STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.
TAKADA, Y. (1979). A family of minimax estimators in some multiple regression problems. *Ann. Statist.* **7** 1144–1147.

DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-7901

DISCUSSION

JAMES BERGER

Purdue University

The paper presents an exciting and rich mix of foundational issues concerning conditional reasoning and methodological developments involving improved estimation in multiple linear regression. My discussion will focus on the foundational issues, though certain features of the improved estimators will be used to illustrate some of the issues.

My first attempt at understanding the fundamental issue raised by the paper was along the following lines (sticking with the criterion of "admissibility" for preciseness):

> Ancillarity Paradox—A procedure which is conditionally admissible for each value of an ancillary statistic can be unconditionally inadmissible.

As I thought about it, however, this did not seem to capture the true novelty of the paper, because this ancillarity paradox has long been known, going back at least as far as the Cox example concerning testing with two randomly differing sample sizes. Brown notes that there is a difference between tests and estimation in regard to the role of reported accuracy, and hence views the estimation examples as more telling, but the difference strikes me as being at most one of degree. In the situation of Section 2.1, for instance, the conditional risk, given ω, of $\delta^* = \omega' d^*$ is certainly not irrelevant; if this were enormous for a particular ω_0 and ω_0 were actually observed, the good unconditional performance of δ^* would hardly seem satisfying. (Brown does not argue otherwise, but simply suggests that unconditional performance is clearly relevant in estimation, perhaps less so in testing.) To focus on the true novelty of Brown's examples, consider the following estimation example which trivially demonstrates the ancillarity paradox.

EXAMPLE. Observed are $X \sim N(\theta, 1)$ and (independently) $Y \sim \text{Bernoulli}(\frac{1}{2})$. It is desired to estimate θ under squared error loss. The estimator $\delta(X, Y) = YX$ is conditionally admissible given $Y = 0$ or $Y = 1$, but is globally inadmissible and is dominated by the Rao–Blackwellized $\delta^*(X, Y) = \frac{1}{2}X$.

While this is an example of the ancillarity paradox in estimation, I would lump it with the Cox example and differentiate it from Brown's examples by adding a phrase to the paradox:

> Alternative Ancillarity Paradox—A procedure which is conditionally admissible for each value of an ancillary statistic can be unconditionally inadmissible, and it can be impossible to determine its inadequacy using conditional reasoning.

To understand the motivation for the last phrase above, observe that it is possible to realize the inadequacy of $\delta(X, Y) = YX$ in the earlier example using only conditional reasoning. One observes that, given $Y = 0$, $\delta(X, 0) = 0$ is admissible because it is Bayes with respect to a point mass at zero, while, given $Y = 1$, $\delta(X, 1) = X$ is admissible because it is generalized Bayes with respect to a constant prior density (known to yield admissible estimators in one-dimensional location problems). Since the implied prior distribution changes with Y, one is immediately suspicious of $\delta(X, Y)$, either from Bayesian reasoning or from the folklore metatheorem that any admissible rule must be more or less Bayes with respect to a fixed prior. The Cox example exhibits the same feature: Fixed α-level tests for differing sample sizes are Bayes with respect to differing priors. In these examples, one can thus realize the inadequacy of the procedures without resorting to any frequentist calculations.

Brown's examples are quite different and much more surprising. The (generalized) prior yielding the standard procedure in his examples is the *same* for all values of the ancillary statistic. Indeed, I know of no way to conditionally criticize these procedures. Only by resorting to unconditional frequentist calculations can one realize their possible inadequacy. This strikes much deeper than a mere criticism of the actually rather ad hoc practice of conditioning on an ancillary and performing a conditional frequentist analysis; it suggests that solely conditional reasoning may sometimes be inadequate.

I somewhat reluctantly agree with this conclusion, reluctantly because of my support for the likelihood principle. One can remain an uncompromising conditionalist only by strict adherence to the subjective (proper prior) Bayesian paradigm or the more recent robust Bayesian version thereof [see Berger (1985) for references]. Since I find it pragmatically necessary to sometimes leave the subjective Bayesian paradigm—at least to allow improper priors and even perhaps likelihood methods—I accept the possible relevance of unconditional frequentist calculations.

I should add a practical qualification to this endorsement of a role for unconditional frequentist analysis. One of the primary attractions of conditional analysis is that it is often much easier than unconditional analysis, allowing much more sophisticated modelling or utilization of prior information. The study of dominating minimax estimators in Brown's paper is a case in point. Brown admits that it is far from clear how to choose among the various possibilities, or even if any of the suggested estimators will ultimately be judged suitable. My own view is that approaching the problem from the unconditional side is far too difficult. Much easier is to approach it in a conditional Bayesian fashion, utilizing priors that incorporate available information about (α, β). (In this regard, I found Brown's explanations of certain estimators in terms of Bayes or empirical Bayes motivations very interesting.) The functional form chosen for the prior may be influenced by general unconditional frequentist considerations (e.g., flat tails are desirable, though not too flat), but one can often get by without actually doing a frequentist computation.

I do not feel that one can operate in the other direction as easily. It is rare for estimators developed solely as unconditional frequentist dominating estimators to be of much use in practice. Basically, they are unlikely to "shrink" right. Also, in this regard, if one has *no* prior information about (α, β) (even very vague knowledge that, say, β is near zero), then there is no real practical advantage to using alternatives to δ_0. If one has no idea where to shrink, it would be a waste of effort to do so.

There is also, of course, the generic motivation for conditional development of procedures based on a desire for good conditional behavior. An unconditionally satisfactory procedure can be terrible conditionally on certain sets; an example is (2.1.4) for x near zero. [Of course, the positive part version (2.1.7) would partially correct this problem.] Developing the procedure conditionally to assure conditional soundness and then checking its unconditional behavior is usually much easier than the reverse.

From the conditional frequentist perspective there is also concern in using a procedure developed from the unconditional side. For instance, in Section 2.1 where $\theta = \omega'\mu$ is of interest, it is certainly relevant upon observing ω to look at the conditional risk given ω. This conditional risk can be very unappealing; for instance, it can be as large as $|\omega|^2 p/4$ for the estimator (2.1.4) when $\rho = p - 2$ and $\Sigma = \Omega = I$ [cf. Efron and Morris (1972)], and can be much larger in nonsymmetric problems. (Note that $\omega' x$ has conditional risk, given ω, of only $|\omega|^2$.) This should be disturbing to frequentists and would probably lead to use of limited translation versions of (2.1.4) [Efron and Morris (1972), Stein (1981), Berger and Dey (1985)], which have much smaller maximum conditional risks.

The above comments are not meant to imply that estimators developed solely as unconditional dominating estimators are of no interest. Quite to the contrary, they are often very helpful in guiding the type of development I recommend. My view is simply that such unconditional estimators tend to be mainly of theoretical interest, as opposed to being statistical methodology that can be recommended in practice. I am sure, for instance, that Brown's estimators and his interesting discussion of the role of information about V will be intently studied by those wanting to create conditionally (and unconditionally) satisfactory methodology for the multiple linear regression problem.

The following points summarize my comments:

1. Brown has convincingly demonstrated that, except for subjective proper prior Bayesian analysis (single-prior or robust versions, chosen—not formally—but to reflect actual subjective beliefs), conditional reasoning is not necessarily self-sufficient; unconditional inadmissibility can occur in ways that are not conditionally recognizable.
2. While unconditional inadmissibility is useful as a check on conditional reasoning, actual procedure development should be done in the conditional arena; procedures developed solely as unconditional dominating procedures can be very interesting and helpful theoretically, but are rarely practical solutions.
3. The ancillarity paradox poses severe problems for frequentist theory. Conditional evaluations are clearly important, but one cannot condition and then simply apply usual frequentist criteria. A mix of good conditional and good unconditional frequentist behavior is needed, but this opens up a Pandora's box of possibilities. I do not feel that a good mix can generally be found within the frequentist domain. The conditional development of procedures and conditional evaluation thereof is most naturally a Bayesian enterprise. I see unconditional frequentist performance in a supporting role, possibly guiding certain aspects of the conditional development.

REFERENCES

BERGER, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BERGER, J. and DEY, D. K. (1985). Truncation of shrinkage estimators of normal means in the nonsymmetric case. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 43–56. North-Holland, Amsterdam.

EFRON, B. and MORRIS, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators II: The empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907