Rejoinder

G. K. Robinson

INTRODUCTION

I would like to thank the discussants for their remarks. I hope that readers will find that the discussion helps to clarify the ideas that I tried to present in my paper. Mostly, I have chosen not to use this opportunity to restate my opinion on minor points where I disagree with the discussants or where I would give different emphasis.

In this introduction I will pass quickly over a number of issues which can each be presented briefly. Issues requiring longer discussion will be laid out as separate sections.

C. R. Henderson died in March 1989. Searle (1989) is an obituary.

Following comments by Harville and Speed, I think that my presentation would have been easier to understand if I had given greater emphasis to the way the linear model (1.1) would be handled if the random effects were not to be estimated. The linear model could be rewritten as

$$y = X\beta + \varepsilon,$$

where $\varepsilon = Zu + e$. Now $\operatorname{Var}(\varepsilon) = (ZGZ^T + R)\sigma^2$ and it is convenient to denote $ZGZ^T + R$ by V. The generalized least-squares estimate

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

is the same as the BLUP estimate as explained in Section 5.1.

As Harville and Thompson indicated, BLUP is often explained using a predictive formulation. Henderson frequently used such a formulation. (e.g., Henderson, 1973). Goldberger (1962) also used a predictive formulation. I find my presentation simpler, but I recommend that readers consider the alternative to see which they find easier to comprehend.

As pointed out by Spall, I did not clarify the distinction between smoothers and filters in my paper. His statement "it is a Kalman smoother ... that produces the BLUP estimate of u based on data y" might leave readers thinking that the Kalman filter is not BLUP. In fact, the Kalman filter is the BLUP estimate of u based on the data up to time t, y_t.

NOMENCLATURE

One of the major barriers to discussion in this area is the variety of nomenclature.

- I have used the term BLUP where many other people would use the term *parametric empirical Bayes*.
- I refer to *random effects* within *mixed models* whereas Steffey and Kass refer to *unit-specific parameters* within *conditionally independent hierarchical models*.
- Much terminology is application specific.

I do not wholeheartedly support the term BLUP because it includes the idea of predicting, and I do not believe that estimates of random effects are predictors for any greater fraction of their usage than estimates of fixed effects are predictors.

In ore reserve estimation I find it silly to speak of predicting something that happened millions of years ago. In time series, it is common to differentiate between smoothing, filtering and prediction. BLUP can be used for all three—which suggests that it is not merely *prediction*.

In the absence of general agreement about terminology, I would appeal for greater tolerance of other people's terminology.

COMPUTATIONAL ISSUES

As Speed hinted at in his discussion, when I first started working on the paper I was involved in the task of designing a computating strategy for estimating the genetic merits of dairy cattle using BLUP. (My first draft of the paper was dated February 22, 1982.)

Up to that time, BLUP for large numbers of sires had been done using several different models, but BLUP for models requiring the solution of sets of simultaneous equations with equations corresponding to both male and female animals (often referred to as *animal models*) had only been used for small number of animals. Henderson (1975b) had proposed the model for use within single herds. The Australian Dairy Herd Improvement Scheme accepted my opinion that an animal model was computationally practical for large numbers of animals and has been using it for several years. Details of the computing strategy are given in Robinson (1986). See also Jones and Goddard (1990). A nonessential development was a method for solving the sets of up to one million simultaneous linear equations which is described in Robinson (1988). Many other genetic evaluation schemes with large data sets now also use animal models rather than sire-only models.

ILL-POSED INVERSE PROBLEMS

I agree with most of Campbell's remarks about BLUP and ill-posed inverse problems. In thinking about the distinction between *uncertainty* (which she refers to as the epistemological or Bayesian interpretation of probability) and *variation* (which she refers to as the ontological or classical interpretation of probability), I would include as *variation* the variation between the true images that have been or are likely to be looked at and the variation between patterns of grade in mineral deposits. The characteristic that I look for in deciding whether probability is being used to describe *uncertainty* or *variation* is that when probability is used to describe *variation* it should be possible to estimate the probability distribution or to test hypotheses about the probability distribution using available data.

ILL-POSED THEORY

While thinking about things that are ill posed, I would like to take this opportunity to indicate my disrespect for a type of theoretical work that is seldom helpful for solving real problems. The Cramér-Lévy theorem, referred to by Spall, states that $X+Y$ can only be precisely normally distributed if both X and Y are precisely normally distributed. I regard the Cramér-Lévy theorem as an example of *ill-posed theory* because, although its conclusion does follow from its premises, a small departure from its premises allows a large departure from its conclusions. More specifically, it is not true that $X+Y$ can only be *approximately* normally distributed if both X and Y are *approximately* normally distributed, as is easy to see for the case of X and Y having smooth, unimodal distributions of opposite skewness.

Spall obviously realizes that it is desirable to supplement the Cramér–Lévy theorem's conclusions. Regrettably, he only quotes the statistical significance of the departure from normality of the Kalman filter errors for his simulation experiments, not the extent of the departure.

MOST LIKELY UNOBSERVABLES

Thompson asked whether the method of most likely unobservables can be used to construct confidence intervals for estimates of random effects. Assuming a multivariate normal distribution as in Section 4.1, the ratio of the density of the unobservables u and e for arbitrary β and u to the density given $\beta=\hat{\beta}$ and $u=\hat{u}$ is $\exp\{-Q/(2\sigma^2)\}$, where Q is the quadratic form

$$\begin{pmatrix} u \\ y-X\beta-Zu \end{pmatrix}^T \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}^{-1} \begin{pmatrix} u \\ y-X\beta-Zu \end{pmatrix} - \begin{pmatrix} \hat{u} \\ y-X\hat{\beta}-Z\hat{u} \end{pmatrix}^T \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}^{-1} \begin{pmatrix} \hat{u} \\ y-X\hat{\beta}-Z\hat{u} \end{pmatrix}.$$

Using equation (1.2), this can be shown to be equal to

$$\begin{pmatrix} \beta-\hat{\beta} \\ u-\hat{u} \end{pmatrix}^T \begin{pmatrix} X^TR^{-1}X & X^TR^{-1}Z \\ Z^TR^{-1}X & Z^TR^{-1}Z+G^{-1} \end{pmatrix} \begin{pmatrix} \beta-\hat{\beta} \\ u-\hat{u} \end{pmatrix}.$$

Hence using the method of most likely unobservables to construct confidence intervals is equivalent to assuming that the estimation errors have a multivariate normal distribution with the usual variance-covariance matrix which was given just below equation (1.2).

The meaning of Thompson's last question is not completely clear to me. There appear to be two likelihoods for β.

1. For the linear model $y=X\beta+\varepsilon$ the variance-covariance matrix of estimation errors is

$$E\left[(\hat{\beta}-\beta)^T(\hat{\beta}-\beta)\right] = (X^TV^{-1}X)^{-1}\sigma^2$$

and the likelihood is a multivariate normal distribution with mean $\beta=\hat{\beta}$ and this variance-covariance matrix.

2. The method of most likely unobservables gives a likelihood for β and u that is a multivariate normal distribution with mean of $\beta=\hat{\beta}$, $u=\hat{u}$ and variance

$$\begin{pmatrix} X^TR^{-1}X & X^TR^{-1}Z \\ Z^TR^{-1}X & Z^TR^{-1}Z+G^{-1} \end{pmatrix}^{-1}\sigma^2.$$

These two likelihoods give the same conclusions about fixed effects, so choosing between them is not an issue. To see that they give the same conclusions, remember that the likelihood arising from the method of most likely unobservables is $\exp\{-Q/(2\sigma^2)\}$ times the maximum at $\beta=\hat{\beta}$, $u=\hat{u}$. The derivative of the quadratic form with respect to u is

$$2Z^TR^{-1}X(\beta-\hat{\beta}) + 2(Z^TR^{-1}Z+G^{-1})(u-\hat{u}).$$

The minimum of Q over u is at

$$(u-\hat{u}) = -(Z^TR^{-1}Z+G^{-1})^{-1}Z^TR^{-1}X(\beta-\hat{\beta})$$

and is

$$(\beta-\hat{\beta})^T X^T\{ZGZ^T+R\}^{-1}X(\beta-\hat{\beta}).$$

Thus the maximum over u of the likelihood of the unobservables differs only by a constant factor from the likelihood for the linear model $y = X\beta + \varepsilon$.

In answer to Speed's question in his Section 8, I believe that the coverage of the confidence intervals is the usual sort, provided that they are interpreted highlighting the probability distribution of the random effects that might have occurred.

PRIOR INFORMATION AND ANIMAL BREEDING

Section 3 of the discussion by Steffey and Kass seems to me to require some specific comments.

I agree that considerable knowledge about herd and sire effects is available. However, as I stated in Section 7.3 of my paper, I would prefer to treat herd-year-season effects as fixed rather than random (or, from a Bayesian perspective, to put uniform priors on them) because I am worried about the potential biases in the information contained in between herd-year-season comparisons.

The *prior information*, which Gianola and Fernando (1986) say should be used to preclude *ridiculous estimates of heritability* and other anomalies, can be put into the form of restrictions on the parameter spaces. Readers might be misled if they thought that this was a type of information that non-Bayesian statisticians would be unwilling to use.

The third paragraph is inaccurate. In Section 4.2 of my paper I find that BLUP estimates are exactly (not approximately) Bayes estimates with uniform improper prior on β and a point distribution (not a uniform distribution) on θ. Example 2 from Kass and Steffey (1989) illustrates that using flat or informative priors on the random effect (u in my notation) can make a substantial difference. I agree with this, but I do not agree with the implication that users of BLUP should be concerned about having implicitly used a uniform prior for β.

CLOSURE

Before receiving copies of the discussion, I had wondered whether some discussion might be as intemperate as the comments by O. Kempthorne on a paper by D. V. Lindley at the Waterloo Symposium. The comments reported in Godambe and Sprott (1971, page 452) included the following:

> ... a former colleague, D. L. Harris, prepared a manuscript entitled "Estimation of Random Variables" in 1963. The title "bugged" some people, and the manuscript was rejected by *Biometrics* in all its wisdom. The argumentation was Bayesian, *of course*. Perhaps a manuscript with such a title will receive a slightly more open reception nowadays.

Kempthorne could surely not object to the openness of my paper's reception. Having met Dewey Harris and seen the revised version of his paper ("Estimation of normally distributed random elements of certain statistical models," Journal Paper No. J-4576, Iowa Agricultural and Home Economics Experiment Station, Ames, Project No. 1505, supported by National Science Foundation Grant G-18093), I see no evidence of lack of openness in 1963 either.

In fact, the complimentary and constructive tone of the discussion concerns me because readers might get the impression that there was general agreement with the ideas expressed in my paper. I doubt that this is the case. Two things which I believe readers should regard as controversial but which have not been raised in discussion are the following:

- In Section 5.8 I imply that much work on ranking and selection has been misguided.
- In Section 5.4 I ignore most of the methods used for estimating variance parameters.

ADDITIONAL REFERENCES

ANDERSON, B. D. O. and MOORE, J. B. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J.

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.

ANSLEY, C. F. and KOHN, R. (1986). Prediction mean squared error for state-space models with estimated parameters. *Biometrika* **73** 467–473.

BLIGHT, B. J. N. and OTT, L. (1975). A Bayesian approach to model inadequacy for polynomial regression. *Biometrika* **62** 79–88.

BROEMELING, L. D. (1985). *Bayesian Analysis of Linear Models*. Dekker, New York.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17** 453–555.

CARLIN, B. P. and GELFAND, A. E. (1990). Approaches for empirical Bayes confidence intervals. *J. Amer. Statist. Assoc.* **85** 105–114.

CASSEL, C.-M., SARNDAL, C.-E. and WRETMAN, J. H. (1977). *Foundations of Inference in Survey Sampling*. Wiley, New York.

CRESSIE, N. (1990). The origins of kriging. *Math. Geol.* **22** 239–252.

DE HOOG, F. R., SPEED, T. P. and WILLIAMS, E. R. (1990). On a matrix identity associated with generalized least squares. *Linear Algebra Appl.* To appear.

DEMPSTER, A. P., RUBIN, D. B. and TSUTAKAWA, R. K. (1981). Estimation in covariance component models. *J. Amer. Statist. Assoc.* **76** 343–353.

DEYST, J. J. (1973). Corrections to "Conditions for asymptotic stability of the discrete minimum variance linear estimator" by J. J. Deyst and C. F. Price. *IEEE Trans. Automat. Control* **AC-18** 562–563.

DEYST, J. J. and PRICE, C. F. (1968). Conditions for asymptotic stability of the discrete minimum variance linear estimator. *IEEE Trans. Automat. Control* **AC-13** 702-705.

DUBRULE, O. (1983). Two methods with different objectives: Splines and kriging. *Math. Geol.* **15** 249-261.

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*. Wiley, New York.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

GIANOLA, D. and FERNANDO, R. L. (1986). Bayesian methods in animal breeding theory. *J. Animal Sci.* **63** 217-244.

GODAMBE, V. P. and SPROTT, D. A. (1971). *Foundations of Statistical Inference*. Holt, Rinehart & Winston, Toronto.

GOODWIN, G. C. and PAYNE, R. L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Academic, New York.

GREEN, P. J. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika* **72** 527-537.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems (with discussion). *J. Amer. Statist. Assoc.* **72** 320-340.

HENDERSON, C. R. (1975b). Use of all relatives in intraherd prediction of breeding values and producing abilities. *J. Dairy Sci.* **58** 1910-1916.

HILL, J. R. (1990). A general framework for model-based statistics. *Biometrika* **77** 115-126.

ISAAKS, E. H. and SRIVASTAVA, R. M. (1988). Spatial continuity measures for probabilistic and deterministic geostatistics. *Math. Geol.* **20** 313-341.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361-379. Univ. California Press, Berkeley.

JAZWINSKI, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic, New York.

JONES, L. P. and GODDARD, M. E. (1990). Five years experience with the animal model for dairy evaluation in Australia. *Proceedings of the Fourth World Congress on Genetics Applied to Livestock Production*, vol. XII, 382-385.

KADANE, J. B. (1990). Comment on "The 1988 Wald Memorial Lectures: The present position in Bayesian statistics" by D. V. Lindley. *Statist. Sci.* **5** 80-82.

KADANE, J. B., DICKEY, J. M., WINKLER, R. L., SMITH, W. S. and PETERS, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *J. Amer. Statist. Assoc.* **75** 845-854.

KASS, R. E. and STEFFEY, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Amer. Statist. Assoc.* **84** 717-726.

LAIRD, N. M. and LOUIS, T. A. (1987). Empirical Bayes intervals based on bootstrap samples. *J. Amer. Statist. Assoc.* **82** 739-750.

LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963-974.

MARROQUIN, J., MITTER, S. and POGGIO, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *J. Amer. Statist. Soc.* **82** 76-89.

NYCHKA, D. (1988). Bayesian confidence intervals for smoothing splines. *J. Amer. Statist. Assoc.* **83** 1134-1143.

O'HAGAN, A. (1978). Curve fitting and optimal design for prediction (with discussion). *J. Roy. Statist. Soc. Ser. B* **40** 1-42.

O'SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statist. Sci.* **1** 502-527.

ROBINSON, G. K. (1988). A method of accelerating stationary iterative methods for solving linear systems. *J. Austral. Math. Soc. Ser. B* **30** 1-23.

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments (with discussion). *Statist. Sci.* **4** 409-435.

SALLAS, W. M. and HARVILLE, D. A. (1988). Noninformative priors and restricted maximum likelihood estimation in the Kalman filter. In *Bayesian Analysis of Time Series and Dynamic Models* (J. C. Spall, ed.) 477-508. Dekker, New York.

SEARLE, S. R. (1989). Obituary: Charles Roy Henderson 1911-1989. *Biometrics* **45** 1333-1335.

SPALL, J. C. (1988). An overview of key developments in dynamic modeling and estimation. In *Bayesian Analysis of Time Series and Dynamic Models* (J. C. Spall, ed.) xv-xxvii. Dekker, New York.

SPALL, J. C. and GARNER, J. P. (1990). Parameter identification for state-space models with nuisance parameters. *IEEE Trans. Aerospace Electron. Syst.* **26**. In press.

SPALL, J. C. and WALL, K. D. (1984). Asymptotic distribution theory for the Kalman filter state estimator. *Comm. Statist. Theory Methods* **13** 1981-2003.

STEINBERG, D. M. (1985). Model robust response surface designs: Scaling two-level factorials. *Biometrika* **72** 513-526.

TARANTOLA, A. (1987). *Inverse Problem Theory. Methods for Data Fitting and Model Parameter Estimation*. North-Holland, Amsterdam.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions*. Academic, New York.

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

WHITTAKER, J. M. (1923). On a new method of graduation. *Proc. Edinburgh Math. Soc.* (2) **41** 63-75.

WOODBURY, A. D. (1989). Bayesian updating revisited. *Math. Geol.* **21** 285-308.