CHOOSING SHRINKAGE ESTIMATORS FOR REGRESSION PROBLEMS

PREPARED UNDER A GRANT FROM THE OFFICE OF ECONOMIC OPPORTUNITY

JOHN E. ROLPH

R-1640-OEO
FEBRUARY 1975

Rand
SANTA MONICA, CA. 90406

The research reported herein was performed pursuant to a grant from the Office of Economic Opportunity, Washington, D.C. The opinions and conclusions expressed herein are solely those of the author and should not be construed as representing the opinion or policy of any agency of the United States Government.

Publishd by The Rand Corporation

CHOOSING SHRINKAGE ESTIMATORS FOR REGRESSION PROBLEMS

PREPARED UNDER A GRANT FROM THE OFFICE OF ECONOMIC OPPORTUNITY

JOHN E. ROLPH

R-1640-OEO
FEBRUARY 1975

Rand
SANTA MONICA, CA. 90406

PREFACE

This report grew out of research begun and financed under Contract 90088 D-73-01 with the Office of Economic Opportunity. It was written while the author was on the staff of the Department of Statistics and Computer Science at University College, London, during the academic year 1973-1974.

It is expected that the shrinkage methods described in this report will be useful in analyzing the data produced by the Health Insurance Study, which is being conducted by Rand and sponsored by the Department of Health, Education, and Welfare.

SUMMARY

A Bayesian formulation of the canonical form of the standard regression model is used to compare various Stein-type estimators and the ridge estimator of regression coefficients. A particular ("constant prior") Stein-type estimator having the same pattern of shrinkage as the ridge estimator is recommended for use.

CONTENTS

PREFACE iii

SUMMARY v

Section

1. INTRODUCTION 1

2. CANONICAL FORM 2

3. STEIN-TYPE ESTIMATORS 3

4. APPLICATIONS 7
   - 4.1. Unknown Variance 7
   - 4.2. Partial Shrinkage 8
   - 4.3. Discussion 9

REFERENCES 13

CHOOSING SHRINKAGE ESTIMATORS FOR REGRESSION PROBLEMS

1. INTRODUCTION

For estimating the regression coefficients in linear regression, the practice of shrinking the least-squares estimates of the coefficients toward a point, or more generally toward a subspace, has received increasing attention. Two particularly popular approaches to estimating regression coefficients are the ridge estimator of Hoerl and Kennard [8] and the Stein-type estimators derived from the estimation methods given in the original papers by Stein [15] and by James and Stein [9]. The ridge estimator was designed as a method to improve on the unsatisfactory characteristics of the least-squares estimator when there is multicollinearity present--when $\underset{\sim}{X}^T\underset{\sim}{X}$ is badly conditioned. Stein-type estimators are frequently recommended because they reduce mean-square error and they can be regarded as empirical Bayes estimators [5]. Section 2 gives the standard regression model in canonical form, while Section 3 expresses both ridge and Stein-type estimators in canonical form. Here a Bayesian approach is used to exhibit the relationship between various Stein-type estimators and the ridge estimator. Translating some desirable properties of Stein-type estimators from canonical form to the standard regression model leads to the recommendation of a particular ("constant prior") Stein-type estimator that has the same pattern of shrinkage as the ridge estimator, but uses an empirical Bayes estimate of the degree of shrinkage. Section 4 covers the case of unknown error variance, partial shrinkage, and a purely Bayesian approach to the problem. It is hoped that this expository report will contribute to a better understanding of the relationship between ridge estimators and Stein-type estimators.

Several other shrinkage estimators for regression that are not described here but which deserve mention include methods of independent variable selection, such as stepwise regression, which make the shrinkage factor either zero or one. Another recent method of choosing shrinkage factors is the cross-validatory approach of Geisser [6] and Stone [17]. The formulation and treatment of shrinkage estimators in the canonical form of the standard regression problem by Dempster [3] and by Goldstein and Smith [7] are similar to this presentation, but do not focus on the identity between the constant prior Stein-type estimators and empirical Bayes ridge estimators.

2. CANONICAL FORM

Consider the standard regression model,

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \tag{2.1}$$

where $\underset{\sim}{Y}$ is an $n \times 1$ response vector, $\underset{\sim}{X}$ is a fixed $n \times p$ matrix of rank p, $\underset{\sim}{\beta}$ is a $p \times 1$ vector, and $\underset{\sim}{\varepsilon}$ is an $n \times 1$ vector of independent $N(0, \sigma^2)$ observations. For convenience, I assume that the independent variables are scaled so that $\underset{\sim}{X}^T\underset{\sim}{X}$ is in correlation form. Usually the expected value of $\underset{\sim}{Y}$ in (2.1) is written $\mu\underset{\sim}{1} + \underset{\sim}{X}\underset{\sim}{\beta}$. To keep notation simple, I omit the constant term μ. Since the interest is in estimating $\underset{\sim}{\beta}$, the constant term can be left out in an approximate analysis if $\underset{\sim}{Y}$ and $\underset{\sim}{X}$ are taken to be the deviations from sample means. Alternatively, the constant can be included by taking the first column of the $\underset{\sim}{X}$ matrix to be all ones, the remainder being deviations from the sample mean. In this situation, the notion of partial shrinkage given in Section 4.2 applies.

It will be convenient to reexpress the above model in canonical form so that (2.1) becomes

$$\underset{\sim}{Y}^* = \underset{\sim}{X}^*\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon}^*, \tag{2.2}$$

where $Y^* = \underset{\sim}{D}Y$, $\underset{\sim}{X}^* = \underset{\sim}{DXC}^T$, $\underset{\sim}{\alpha} = \underset{\sim}{C}\beta$, and $\varepsilon^* = \underset{\sim}{D}\varepsilon$, with $\underset{\sim}{C}$ and $\underset{\sim}{D}$ being $p \times p$ and $n \times n$ orthogonal matrices satisfying

$$\underset{\sim}{CX}^T\underset{\sim}{X}C^T = \mathrm{diag}\ (\lambda_1, \ldots, \lambda_p), \tag{2.3}$$

and

$$\underset{\sim}{X}^* = \underset{\sim}{DXC}^T = [\mathrm{diag}\ (\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p})].$$

Note that (2.2) simplifies to

$$Y_i^* \overset{\text{ind}}{\sim} N(\sqrt{\lambda_i}\alpha_i,\ \sigma^2) \qquad \text{if } i = 1, \ldots, p,$$

and

$$Y_i^* \overset{\text{ind}}{\sim} N(0,\ \sigma^2) \qquad \text{if } i = p+1, \ldots, n,$$

where "$\overset{\text{ind}}{\sim}$" means "independently distributed as."

The least-squares estimator of $\underset{\sim}{\beta}$ is

$$\underset{\sim}{b} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\ \underset{\sim}{X}^T\underset{\sim}{Y},$$

and since orthogonal transformations preserve distances, the least-squares estimator of $\underset{\sim}{\alpha}$, $\underset{\sim}{a} = \underset{\sim}{C}\underset{\sim}{b}$ has the same dispersion about $\underset{\sim}{\alpha}$ as $\underset{\sim}{b}$ does about $\underset{\sim}{\beta}$. Note that $\underset{\sim}{b} = \underset{\sim}{C}^T\underset{\sim}{a}$. The components $a_1 = Y_1^*/\sqrt{\lambda_1}, \ldots, a_p = Y_p^*/\sqrt{\lambda_p}$ are independent normally distributed random variables with means $\alpha_1, \ldots, \alpha_p$ and variances $\sigma^2/\lambda_1, \ldots, \sigma^2/\lambda_p$. Since estimators will be judged relative to a quadratic loss function, and since orthogonal transformations leave quadratic loss unchanged, we will estimate $\underset{\sim}{\alpha}$ and then multiply by $\underset{\sim}{C}^T$ to transform back to the original problem (2.1) to estimate $\underset{\sim}{\beta}$.

3. STEIN-TYPE ESTIMATORS

Assuming that σ^2 is known, the problem is to simultaneously estimate p means, $\alpha_1, \ldots, \alpha_p$, from p independent normally distributed random variables, $Y_1^*/\sqrt{\lambda_1}, \ldots, Y_p^*/\sqrt{\lambda_p}$, with known variances $\sigma^2/\lambda_1, \ldots, \sigma^2/\lambda_p$, respectively. A useful way of deriving Stein-type estimates of $\underset{\sim}{\alpha}$ is to formally take a Bayes approach. The prior distribution on α_i is assumed to be

$$\alpha_i \overset{\text{ind}}{\sim} N(0,\ \rho_i A), \qquad i = 1, \ldots, p, \tag{3.1}$$

where $\rho_1, \ldots, \rho_p$ are known constants scaled so that $\Sigma\, \rho_i = p$. The prior mean being zero in (3.1) corresponds to the ridge regression estimator given below. Section 4.2 describes how this analysis carries through for a prior distribution with an arbitrary mean. The posterior distribution of $\underset{\sim}{\alpha}$ given $\underset{\sim}{Y}^*$ is

$$\alpha_i | Y_i^* \overset{\text{ind}}{\sim} N\{(1 - B_i)Y_i^*/\sqrt{\lambda_i},\ (1 - B_i)\sigma^2/\lambda_i\}, \tag{3.2}$$

where

$$B_i = \frac{\sigma^2/\lambda_i}{\sigma^2/\lambda_i + \rho_i A} = \frac{\sigma^2}{\sigma^2 + \lambda_i \rho_i A} .$$

The usual Bayes estimator for a quadratic loss function is the mean of the posterior distribution

$$\alpha_i^* = (1 - B_i)Y_i^*/\sqrt{\lambda_i}. \tag{3.3}$$

Stein-type estimates can be regarded as empirical Bayes estimates by taking the parameter A of the prior distribution as unknown, estimating A or equivalently B_i from the sample, and using (3.3) with B_i replaced by an estimate $\hat{B}_i$ (see [5]). The "positive part" James-Stein estimate [5] occurs when $\rho_i \equiv 1$, $\lambda_i \equiv 1$, so that

$$B_i \equiv B = \frac{\sigma^2}{\sigma^2 + A} \quad \text{and} \quad \hat{B} = \min\left\{1, \frac{(p - 2)\sigma^2}{\Sigma\, Y_i^{*2}/\lambda_i}\right\} .$$

Then the positive part James-Stein estimator of α_i is

$$\hat{\alpha}_i = (1 - \hat{B})Y_i^*/\sqrt{\lambda_i}. \tag{3.4}$$

This situation occurs in orthogonal regression where $\underset{\sim}{X}^T\underset{\sim}{X} = I$, so that $\lambda_i \equiv 1$. Thus the corresponding empirical Bayes estimator of $\underset{\sim}{\beta}$ is

$$\hat{\underset{\sim}{\beta}} = \underset{\sim}{C}^T \hat{\underset{\sim}{\alpha}} = (1 - \hat{B})\underset{\sim}{b}. \tag{3.5}$$

Since $\Sigma\, Y_i^{*2}/\lambda_i = \Sigma\, b_i^2$, $\hat{B} = \min\,\{1,\ (p - 2)\sigma^2/\Sigma\, b_i^2\}$. Sclove [12] essentially uses (3.5) for orthogonal regression.

This "Stein-type" estimator used by most authors in regression problems ($\lambda_i \not\equiv \lambda$), ([1], [3], [7], [16]), corresponds to using a prior distribution on α_i, where $\rho_i = m\sigma^2/\lambda_i$. The posterior distribution (3.2) then has $B_1 \equiv B = 1/(1 + mA)$, so that the corresponding empirical Bayes estimator uses the common value of B_i in (3.3) as

$$\hat{B} = \min \left\{ 1,\ \frac{(p - 2)\sigma^2}{\Sigma\, Y_i^{*2}} \right\}. \tag{3.6}$$

This follows from applying the equal variance James-Stein estimator (3.4) to the variables Y_i^*. From (2.2) the estimator (3.6) of B yields the estimator

$$\hat{\beta} = (1 - \hat{B})\underset{\sim}{b}, \tag{3.7}$$

where $\hat{B} = \min\,\{1,\ (p - 2)\sigma^2/\underset{\sim}{b}^T \underset{\sim}{X}^T \underset{\sim}{Xb}\}$. The above estimate with $(p - 2)\sigma^2/\underset{\sim}{b}^T \underset{\sim}{X}^T \underset{\sim}{Xb}$ instead of $\hat{B}$ in (3.7) is referred to as a James-Stein estimator by Dempster [3], Goldstein and Smith [7], and Sclove [12]. An unappealing property of (3.7) is that all coefficient estimates b_i are shrunk by the same amount. As in [2], I will refer to (3.7) as a *proportional prior* Stein-type estimator because the variance of the prior distribution is proportional to the sampling variance σ^2/λ_i.

An appealing choice is to take $\rho_i \equiv 1$ so that all components (regression coefficients) have a *constant prior* variance. Under this assumption, the posterior distribution of α_i given $Y_i^*/\sqrt{\lambda_i}$ has mean $[1 - \{\sigma^2/(\sigma^2 + \lambda_i A)\}Y_i^*/\sqrt{\lambda_i}]$ and variance $\lambda_i A/(\sigma^2 + \lambda_i A)$. To transform the Bayes estimator $[1 - \{\sigma^2/(\sigma^2 + \lambda_i A)\}]Y_i^*/\sqrt{\lambda_i}$ of α_i into a Bayes estimator of β_i, observe from (2.2) and (2.3) that

$$\underset{\sim}{\alpha}^* = [\underset{\sim}{C}(\underset{\sim}{X}^T \underset{\sim}{X} + (\sigma^2/A)I_p)C^T]^{-1}\underset{\sim}{CX}^T \underset{\sim}{D}^T Y^*,$$

and hence, after some rearrangement,

$$\underset{\sim}{\beta}^* = \underset{\sim}{C}^T \underset{\sim}{\alpha}^* = (\underset{\sim}{X}^T \underset{\sim}{X} + (\sigma^2/A) I_p)^{-1} \underset{\sim}{X}^T \underset{\sim}{Y}. \tag{3.8}$$

If σ^2/A is replaced by k in (3.8), and k is chosen in the appropriate judgmental way, the resulting estimator is the *ridge estimator* of Hoerl and Kennard [8]. Alternatively, an empirical Bayes approach can be used so that $k = \sigma^2/A$ is estimated from the data and that estimate is substituted into (3.8), giving a Stein-type estimator of $\underset{\sim}{\beta}$.

To choose an empirical Bayes Stein-type estimator of $\underset{\sim}{\beta}$, consider the problem of finding a Stein-type estimator of $\underset{\sim}{\alpha}$. That is, find an estimator of $B_i = (\sigma^2/\lambda_i)/(\sigma^2/\lambda_i + A)$ to use the formula

$$\alpha_i^* = (1 - B_i) Y_i / \sqrt{\lambda_i}. \tag{3.9}$$

There are several methods that have been used to derive Stein-type estimators. Defining $S(A) = \Sigma \{a_i^2/(A + \sigma^2/\lambda_i)\}$, Carter and Rolph [2] choose $\hat{A}$ so that $S(\hat{A}) = p$, since p is the expected value of S(A). Then $\underset{\sim}{\alpha}$ is estimated by

$$\hat{\alpha}_i = (1 - \hat{B}_i) a_i, \tag{3.10}$$

where

$$\hat{B}_i = \min \left[1, \frac{(p - 2)\sigma^2}{\lambda_i S(\hat{A}) + (p - 2)\sigma^2 \{1 - (\lambda_i/p) \sum_j (1/\lambda_j)\}} \right].$$

In the equal variance case, where $\lambda_i \equiv \lambda$ for all i, (3.10) reduces to the usual positive part of the James-Stein estimator (3.4). Dempster [3] uses p rather than p - 2 in (3.10) to get what he designates a "ridge estimator." Efron and Morris [5] give an alternative to (3.10) that also reduces to the James-Stein estimator in the equal variance case. All of these estimators seem to differ very little and will be called *constant prior* Stein-type estimators. Our estimate of the regression coefficients $\underset{\sim}{\beta}$ is then

$$\hat{\underset{\sim}{\beta}} = C^T\hat{\alpha}, \tag{3.11}$$

where $\hat{\alpha}$ is defined in (3.10).

Constant prior estimates such as (3.11) are usually preferable to the proportional prior estimates given by (3.7). Empirical work by Hoerl and Kennard [8] shows that ridge estimates yield substantial savings over ordinary least squares. Their ridge estimators differ from (3.11) in that $\hat{A}$ in (3.10) is estimated judgmentally. More recently, Dempster and his colleagues [3] have found in a simulation study that that the constant prior Stein-type estimators dominate proportional prior Stein-type estimators for two different quadratic loss functions in a variety of realistic regression situations. On the analytical side, Goldstein and Smith [7] have proved that for *any* parameter value $\underset{\sim}{\alpha}$, there exists a value of $k = \sigma^2/A$ in (3.8), so that the resulting estimators have smaller expected squared error than the least-squares estimators b_i in *every* component. Thus, with a good estimator of k, one can *hope* to improve on least squares for *every* coefficient rather than just reduce the expected sum of the squared errors as Stein [15] and James and Stein [9] first guaranteed.

4. APPLICATIONS

4.1 Unknown Variance

In most realistic regression situations, the error variance σ^2 is unknown. The usual point estimator of σ^2 is the residual mean square

$$S^2 = \frac{1}{n - p} (\underset{\sim}{Y} - X\underset{\sim}{b})^T(\underset{\sim}{Y} - X\underset{\sim}{b}). \tag{4.1}$$

Rather than to substitute (4.1) in place of σ^2 in the Stein-type rules described thus far, we will use

$$V = \frac{1}{n - p + 2} (\underset{\sim}{Y} - \underset{\sim}{Xb})^T(\underset{\sim}{Y} - \underset{\sim}{Xb}), \tag{4.2}$$

which turns out to give an estimator with smaller mean-squared error in the equal variance case. In the equal variance case where $\rho_i \equiv 1$ and $\lambda_i \equiv 1$, the estimate of $\underset{\sim}{\beta}$ in (3.5) becomes

$$\left[1 - \min\left\{1, \frac{(p - 2)V}{\Sigma\, b_i^2}\right\}\right] \underset{\sim}{b}. \tag{4.3}$$

Similarly, the *proportional prior* estimator of $\underset{\sim}{\beta}$ given in (3.7) becomes

$$\left[1 - \min\left\{1, \frac{(p - 2)V}{\underset{\sim}{b}^T \underset{\sim}{X}^T \underset{\sim}{Xb}}\right\}\right] \underset{\sim}{b}. \tag{4.4}$$

The *constant prior* estimate of $\underset{\sim}{\beta}$ is given by (3.10) and (3.11), where V replaces σ^2. In most situations where σ^2 is unknown, this is a sensible choice of an estimator of $\underset{\sim}{\beta}$.

4.2. Partial Shrinkage

Regression coefficients need not all be pulled toward the origin. Efron and Morris [5], Sclove [12], Stein [13], and others present Stein-type estimators in which either the least-squares estimators are pulled toward a subspace rather than the single point 0 or, equivalently, only a prespecified subset of the components (say, regression coefficients) are pulled toward zero.

Sclove [12] gives an extensive discussion of Stein-type estimation for a subset containing r regression coefficients and compares this procedure with that of first testing the hypothesis that the r coefficients are zero and then using a regression model with either p - r or all p coefficients, depending on whether the test accepts or rejects the hypothesis, respectively. A common situation is one in which the least-squares estimator is pulled toward the one-dimensional subspace, where all the coefficients are equal (i.e., the mean).

In the canonical form of the problem with equal variances and $\rho_i \equiv 1$, D. V. Lindley (in his discussion of [14]) and others suggest pulling the components toward their average value rather than toward the origin. For the unequal variance case, in canonical form, Carter and Rolph [2] give the *constant prior* analogue of (3.10), using this suggestion of pulling toward the average. In the regression situation, this estimator is preferable if one a priori suspects that the regression coefficients are approximately equal but not necessarily zero. This corresponds to testing the hypothesis that all coefficients are equal and then estimating accordingly. In Bayesian terms, shrinking toward the average corresponds to the case in which the prior distribution on α_i is N(m, A) rather than N(0, A). Thus, in an empirical Bayes approach, one must estimate m as well as A from the data [2]. A detailed description of several applications of Stein-type estimators is given by Efron and Morris [4].

4.3. Discussion

Rather than to use the estimators of B_i in (3.3) as an empirical Bayes method to produce Stein-type estimators, a full-blown Bayesian approach can be taken, as in Lindley and Smith [11] and in Leonard [10]. For the equal variance case for shrinking toward the average, Leonard's Bayesian analogue of (3.4) is

$$\tilde{\alpha}_i = \tilde{B}\,\bar{a} + (1 - \tilde{B})a_i, \tag{4.5}$$

where

$$\tilde{B} = \frac{1}{y}\,\frac{\Gamma(q + 1,\ y)}{\Gamma(q,\ y)}, \qquad a_i = Y_i^*/\sqrt{\lambda_i},$$

$$q = \frac{1}{2}\,(p - 3), \qquad y = \frac{\Sigma\,(a_i - \bar{a})^2}{2\sigma^2},$$

and

$$\Gamma(q, y) = \int_0^y u^{q-1} e^{-u} \, du.$$

This contrasts to Lindley's modification of the James-Stein estimator, which is

$$\hat{\alpha}_i = \hat{B}\,\bar{a} + (1 - \hat{B})a_i,$$

where $\hat{B} = q/y$. From the properties of the incomplete gamma function $\Gamma(q, y)$, it is easily seen that the shrinkage of a_i is less using the Bayesian estimator $\tilde{\alpha}_i$ and that, furthermore, $\tilde{B}$ is strictly between 0 and 1, obviating the need to truncate $\hat{B}$ to [0, 1]. Leonard's estimator (4.5) can be calculated for the unequal variance case to give an analogue to the constant prior estimator (3.10), but the computations become considerably more involved than (4.5), so we omit them. Because of some of its attractive properties, like admissibility, one might wish to use a constant prior Bayes estimator in preference to Hoerl and Kennard's ridge estimator or the Stein-type constant prior estimator given by (3.10). In the equal variance case, Efron and Morris [5] show that there is very little difference in the risk functions using quadratic loss between the James-Stein estimator (3.4) and an estimator due to Strawderman [18], which is the same as Leonard's (4.5). Thus one can expect very little difference in the unequal variance constant prior case between our recommended estimator (3.10) and the analogue to Leonard's (4.5) estimator.

* * *

The purpose of this report has been to point out the relationship between two apparently different shrinkage estimators in regression problems: ridge estimators and Stein-type estimators. Unequal variance constant prior Stein-type estimators yield the same shrinkage *pattern* as ridge estimators, but the shrinking constant k is estimated from a statistical model rather than chosen in the more judgmental way that Hoerl and Kennard [8] recommend. Thus the constant prior Stein-type estimator seems to be an improvement over both the ridge estimator and the proportional prior Stein-type estimator, which shrinks all the least-squares coefficients the same amount even if their variances are unequal.

REFERENCES

1. Baranchik, A. J. (1973). "Inadmissibility of Maximum Likelihood Estimators in Some Multiple Regression Problems with Three or More Independent Variables," *Ann. Statist.*, Vol. 1, No. 2, pp. 312-321.

2. Carter, G. M., and J. E. Rolph (1974). "Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities," *J. Amer. Statist. Assoc.*, Vol. 69, No. 348, pp. 880-885.

3. Dempster, A. P. (1973). "Alternatives to Least Squares in Multiple Regression," in D. G. Kabe and R. P. Gupta (eds.), *Multivariate Statistical Inference* (Amsterdam: North Holland Publishing Co.).

4. Efron, B., and C. Morris (1974). *Data Anlaysis, Using Stein's Estimator and Its Generalizations*, The Rand Corporation, R-1394-OEO, March 1974.

5. ----- (1973). "Stein's Estimation Rule and Its Competitors--An Empirical Bayes Approach," *J. Amer. Statist. Assoc.*, Vol. 68, No. 341, pp. 117-130.

6. Geisser, S. (1974). "A Predictive Approach to the Random Effect Model," *Biometrika*, Vol. 61, No. 1, pp. 101-108.

7. Goldstein, M., and A.F.M. Smith (1974). "Ridge-type Estimators for Regression Analysis," *J. Roy. Statist. Soc.*, B, Vol. 36, No. 2, pp. 284-291.

8. Hoerl, A. E., and R. W. Kennard (1970). "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, Vol. 12, No. 1, pp. 55-68.

9. James, W., and C. Stein (1961). "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium*, Vol. 1 (Berkeley, California: University of California Press), pp. 361-379.

10. Leonard, T. (1974). "A Bayesian Method for the Simultaneous Estimation of Several Parameters" (submitted to *J. Amer. Statist. Assoc.*).

11. Lindley, D. V., and A.F.M. Smith (1972). "Bayes Estimates for the Linear Model (with Discussion)," *J. Roy. Statist. Soc.*, B, Vol. 34, No. 1, pp. 1-41.

12. Sclove, S. L. (1968). "Improved Estimators for Coefficients in Linear Regression," *J. Amer. Statist. Assoc.*, Vol. 63, No. 322, pp. 596-606.

13. Stein, C. (1966). "An Approach to the Recovery of Inter-block Information in Balanced Incomplete Block Designs," in F. N. David (ed.), *Festschrift for J. Neyman* (New York: John Wiley & Sons, Inc.), pp. 351-366.

14. ----- (1962). "Confidence Sets for the Mean of a Multivariate Normal Distribution (with Discussion)," *J. Roy. Statist. Soc.*, B, Vol. 24, No. 2, pp. 265-296.

15. ----- (1955). "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," *Proceedings of the Third Berkeley Symposium*, Vol. 1 (Berkeley, California: University of California Press), pp. 197-206.

16. ----- (1960). "Multiple Regression," in I. Olkin (ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (Stanford, California: Stanford University Press), pp. 424-443.

17. Stone, M. (1974). "Cross-validatory Choice and Assessment of Statistical Predictions (with Discussion)," *J. Roy. Statist. Soc.*, B, Vol. 36, No. 2, pp. 111-147.

18. Strawderman, W. E. (1971). "Proper Bayes Minimax Estimators of the Multivariate Normal Mean," *Ann. Math. Statist.*, Vol. 42, No. 1, pp. 385-388.