# Minimaxity of Empirical Bayes Estimators Derived From Subjective Hyperpriors

by

James O. Berger*
Purdue University

and

Shun-Yu Chen
Cheng-Kung University

Technical Report #86-14

Department of Statistics
Purdue University

May 1986

*Research Supported by National Science Foundation Grant DMS 84-01996.

# Minimaxity of Empirical Bayes Estimators Derived From Subjective Hyperpriors

by

James O. Berger* and Shun-Yu Chen
Purdue University — Cheng-Kung University

## Abstract

A p-vector $X = (X_1,...,X_p)^t$ has a normal distribution with unknown mean vector $\theta = (\theta_1,...,\theta_p)^t$ and covariance matrix $\sigma^2 I$, $\sigma^2$ known. It is desired to estimate $\theta$ under sum of squares error loss, in the empirical or hierarchical Bayes scenario where $\theta$ is modelled as having a $\eta_p(\mu,\lambda^{-1}I)$ distribution, $\lambda$ and possibly $\mu$ unknown. When p is small or moderate, empirical Bayes estimates of $\lambda$ can be quite inaccurate, and it can be very beneficial to also utilize subjective knowledge concerning $\lambda$. Empirical Bayes estimators which allow incorporation of such knowledge are developed, are shown to be minimax and are compared with more common empirical Bayes estimators.

*Research Supported by National Science Foundation Grant DMS 84-01996.

## 1. Introduction

Let $X = (X_1, \ldots, X_p)^t$ have a p-variate normal distribution with unknown mean vector $\theta = (\theta_1, \ldots, \theta_p)^t$ and nonsingular known covariance matrix $\Sigma$. In estimating $\theta$, a variety of shrinkage estimators have been proposed from decision theoretic and Bayesian perspectives. It has been argued (cf. Berger (1980, 1982, 1985)) that minimax estimators developed in the decision theoretic approach must usefully incorporate available prior information to offer significant advantages. Often the most attractive manner of doing this is to develop a Bayesian estimator which clearly incorporates such information, and then to establish the minimaxity of the estimator. In this paper we follow such a program for several hierarchical Bayesian situations with informative second stage prior distributions.

The usual hierarchical Bayes formulation for this problem (see Lindley and Smith (1972) or Berger (1985)) assumes that, given the p-vector $\mu$ and the $p \times p$ positive definite matrix A, the unknown $\theta$ has a $\eta(\mu, A)$ (first stage) prior distribution. In addition, however, $\mu$ and A are considered unknown with a (second stage) prior distribution $\pi(\mu, A)$. One can then calculate that the corresponding (hierarchical) Bayes estimator of $\theta$ is

$$(1.1) \qquad \delta^{\pi}(x) = x - \Sigma E^{\pi(\mu, A|x)}((\Sigma + A)^{-1}(x - \mu)),$$

where $\pi(\mu, A|x)$ is the posterior distribution of $\mu$ and A given x (see Berger (1985) for formulas). Verification of the minimaxity of such estimators has only been done for a few essentially degenerate choices of $\pi(\mu,A)$ (cf. Strawderman (1971) and Berger (1980)), partly because of the difficulty of mathematically working with $\pi(\mu,A|x)$. Somewhat more success has been achieved with the empirical Bayes approximation to the above estimator; one determines $\hat{\mu}$ and $\hat{A}$, the maximum likelihood estimates of $\mu$ and $A$ with respect to the posterior distribution $\pi(\mu,A|x)$, and then considers the estimator in (1.1) with $\mu$ and $A$ replaced by $\hat{\mu}$ and $\hat{A}$. (Extensive discussion of such approximations can be found in Lindley and Smith (1972) and Berger (1985).) In proving minimaxity for such empirical Bayes estimators, it is common to consider positive multiples of the shrinkage term, leading to a final form of the estimator of

$$(1.2) \qquad \delta^{B*}(x) = x - t\, \ddagger\, (\ddagger+\hat{A})^{-1}(x-\hat{\mu}),$$

$t$ being a positive scalar.

The most extensively studied special case of this formulation is that in which $\mu = \mu_0(1,\ldots,1)^t$ and $A = \lambda^{-1}I_p$, $\mu_0$ and $\lambda$ unknown, being a model of the exchangeable scenario in which the $\theta_i$ are i.i.d. from an unknown distribution. If $\ddagger = \sigma^2 I_p$, the estimates of $\hat{\mu}$ and $\hat{A}$ are

$$(1.3) \qquad \hat{\mu} = \bar{x}, \quad \hat{A} = \max\{0, \tfrac{1}{p}\Sigma(x_i-\bar{x})^2-\sigma^2\}I_p;$$

the estimator in (1.2) then becomes

$$\delta^{B*}(x) = \begin{cases} x - t(x-\bar{x}\underset{\sim}{1}) & \text{if } \sum_{i=1}^{p} (x_i-\bar{x})^2 \leq p\sigma^2 \\ x - \dfrac{tp\,\sigma^2}{\Sigma(x_i-\bar{x})^2}(x-\bar{x}\underset{\sim}{1}) & \text{else.} \end{cases}$$

with $t = (p-3)/p$, this can be recognized as a truncated version of the usual James-Stein estimator which shrinks to a common mean, as analyzed in, say, Efron and Morris (1973).

When p is small it was argued in Berger (1982) that it might be preferable to use subjective estimates of $\mu_0$ and $\lambda$ (or $\mu$ and A in general) leading to the minimax robust generalized Bayes estimator in Berger (1980). The point is that the estimates in (1.3) will be very inaccurate for small p, and the overall risk performance can be substantially improved through use of subjective estimates of $\mu$ and A. (Note that all estimators here are minimax, so that the criterion of interest would be some measure of overall average risk improvement.)

The natural Bayesian solution to the above dilemma is to give $\mu_0$ and $\lambda$ (or $\mu$ and A in general) an informative (second stage) prior distribution which incorporates the available subjective information. This allows optimum estimation of $\mu_0$ and $\lambda$, and hence optimum overall performance of the resulting estimator of $\theta$. The difficulty is that verification of minimaxity of the resulting estimator of $\theta$ can be very difficult especially since the resulting empirical Bayes estimators of $\mu_0$ and $\lambda$ are only defined implicitely as solutions of likelihood equations.

In this paper we make substantial progress on a special case of the above problem. The specific scenario considered is the symmetric one where $\ddagger = \sigma^2 I_p$, $\sigma^2$ known, and $A = \lambda^{-1} I_p$. Thus we are assuming $X \sim \eta_p(\theta, \sigma^2 I_p)$ and $\theta \sim \eta_p(\mu, \lambda^{-1} I_p)$. The scond stage prior distribution for $(\mu, \lambda)$ is as follows:

(i) Either $\mu$ is assumed to be known, or it is assumed to be of the form $\mu = B\gamma$, where B is a given matrix of rank q, and $\gamma$ is unknown with noninformative prior $\pi(\gamma) \equiv 1$;

(ii) The distribution of $\lambda$ is chosen to be Gamma $(\alpha, \beta/2)$ (independently of $\mu$), where $\alpha$ and $\beta$ are subjectively specified constants with $\beta \geq (\alpha-1)\sigma^2$.

The gamma family of priors for $\lambda$ is sufficiently general to allow reflection of most beliefs about $\lambda$. Note that $\lambda^{-1}$ can be thought of as the common variance of the $\theta_i$, and that

$$E[\lambda^{-1}] = \beta/[2(\alpha-1)].$$

Recall that we assume $\beta \geq (\alpha-1)\sigma^2$; thus only those priors for which $E[\lambda^{-1}] \geq \sigma^2/2$ are allowed. (This is a rather mild constraint, since the variance of the $\theta_i$ will typically be larger than $\sigma^2$, the sample varicance.) Also,

$$\mathrm{Var}(\lambda^{-1}) = \beta^2/[4(\alpha-1)^2(\alpha-2)].$$

Thus one could subjectively specify $E[\lambda^{-1}]$ (a "best guess" for the variability of the $\theta_i$), and $\mathrm{Var}(\lambda^{-1})$ (say, the square of the estimated accuracy of this "best quess"), and solve for the corresponding $\alpha$ and $\beta$.

One could similarly allow for a more general subjective prior on $\mu$, but we do not do so for two reasons. First, it seems to be somewhat less important than utilization of information about $\lambda^{-1}$. Mainly, however, we were unable to handle the ensuing complexity; proof of minimaxity of the resulting estimators is formidable.

## 2. Results When $\mu$ Is Known

When $\mu$ is known, the joint density of X, $\theta$, $\lambda$ is

$$m(x,\theta,\lambda) \propto \lambda^{\frac{p}{2}+\alpha-1} \exp\{-\frac{1}{2}[\frac{1}{\sigma^2}(x-\theta)'(x-\theta)+\lambda(\theta-\mu)'(\theta-\mu)+\beta\lambda]\}$$

and the marginal posterior density of $\lambda$ given X is

$$\pi(\lambda|x) \propto \frac{\lambda^{\alpha-1}}{(\sigma^2+\lambda^{-1})^{p/2}} \exp\{-\frac{1}{2}[(\lambda^{-1}+\sigma^2)^{-1}v+\beta\lambda]\}, \tag{2.1}$$

where $v = \|x-\mu\|^2 = (x-\mu)'(x-\mu)$.

From (2.1), it is easy to show that the MLE of $\lambda$, $\hat{\lambda}$, satisfies the equation $\ell(\hat{\lambda}) = 0$, where

$$\begin{aligned} \ell(\lambda) = {} & (\beta\sigma^4)\lambda^3 + (2\beta\sigma^2-2(\alpha-1)\sigma^4)\lambda^2 \\ & + (v-p\sigma^2-4(\alpha-1)\sigma^2+\beta)\lambda - (p+2(\alpha-1)). \end{aligned} \tag{2.2}$$

Note that the coefficient of $\lambda^2$ in $\ell(\lambda)$ is nonnegative (since $\beta \geq (\alpha-1)\sigma^2$), so that the equation $\ell(\lambda) = 0$ has a unique positive solution. The hierarchical estimator (1.2) can then be written

$$\delta^1(x) = x - \frac{t\hat{\lambda}\sigma^2}{(1+\sigma^2\hat{\lambda})}(x-\mu) \tag{2.3}$$

(recall we are assuming here that $\mu$ is known).

Although $\hat{\lambda}$ is quite complicated (being the solution to a cubic equation), it is possible to verify minimaxity of $\delta^1$; that this can be done for such complicated estimators is one of the main messages of this paper.

Theorem 1. Under sum of squares error loss, the estimator $\delta^1$ is minimax for

(2.4) $$0 \leq t \leq \frac{2p}{t^*} - \frac{4}{p+2(\alpha-1)},$$

where

(2.5) $$t^* = \max\{p+2(\alpha-1),\ p+4(\alpha-1)-\beta/\sigma^2\}.$$

Proof. The familiar Stein identity (see Stein (1981)) shows that, for any estimator

$$\delta(x) = x - \phi(x)$$

satisfying certain mild conditions (all of which are trivially satisfied by the estimators in this paper),

(2.6) $$R(\theta,\delta) = E_\theta|\theta-\delta(x)|^2 = p\sigma^2+E_\theta[\mathcal{D}\phi(x)],$$

where

(2.7) $$\mathcal{D}\phi(x) = |\phi(x)|^2-2\sigma^2\sum_{i=1}^{p}\frac{\partial}{\partial x_i}\phi_i(x).$$

For $\delta^1$,

$$\mathcal{D}\phi(x) = -\sigma^4\frac{t\hat{\lambda}}{(1+\sigma^2\hat{\lambda})}\{2p+\frac{4v(d\hat{\lambda}/dv)}{(\hat{\lambda}+\sigma^2\hat{\lambda}^2)}-\frac{tv\hat{\lambda}}{(1+\sigma^2\hat{\lambda})}\}.$$

Differentiating with respect to v in the equation $\ell(\hat{\lambda}) = 0$ and rearranging terms yields

$$(2.8) \qquad \frac{d\hat{\lambda}}{dv} = -\hat{\lambda}[3\beta\sigma^2\hat{\lambda}^2+4(\beta\sigma^2-(\alpha-1)\sigma^4)\hat{\lambda}+(v-p\sigma^2-4(\alpha-1)\sigma^2+\beta)]^{-1}$$

$$= -\hat{\lambda}^2[2\beta\sigma^4\hat{\lambda}^3+2(\beta\sigma^2-(\alpha-1)\sigma^4)\hat{\lambda}^2+(p+2(\alpha-1))]^{-1}$$

$$\geq -\hat{\lambda}^2(p+2(\alpha-1))^{-1}.$$

Also, from the equation $\ell(\hat{\lambda}) = 0$ it follows that

$$v\hat{\lambda} \leq (p\sigma^2+4(\alpha-1)\sigma^2-\beta)\hat{\lambda}+(p+2(\alpha-1)),$$

which implies that

$$(2.9) \qquad \frac{v\hat{\lambda}}{1+\sigma^2\hat{\lambda}} \leq \frac{(p\sigma^2+4(\alpha-1)\sigma^2-\beta)\hat{\lambda}+(p+2(\alpha-1))}{1+\sigma^2\hat{\lambda}} \leq t^*.$$

Hence

$$(2.10) \qquad \mathcal{D}\phi(x) \leq -\sigma^4\frac{t\hat{\lambda}}{(1+\sigma^2\hat{\lambda})}\{2p+\frac{4v(-\hat{\lambda}^2)}{(\hat{\lambda}+\sigma^2\hat{\lambda}^2)[p+2(\alpha-1)]}-\frac{tv\hat{\lambda}}{(1+\sigma^2\hat{\lambda})}\}$$

$$= -\frac{\sigma^4 t\hat{\lambda}}{(1+\sigma^2\hat{\lambda})}\{2p-\frac{v\hat{\lambda}}{(1+\sigma^2\hat{\lambda})}[\frac{4}{p+2(\alpha-1)}+t]\}$$

$$\leq -\frac{\sigma^4 t\hat{\lambda}}{(1+\sigma^2\hat{\lambda})}\{2p-t^*[\frac{4}{p+2(\alpha-1)}+t]\}.$$

From (2.6) it is clear that $\delta^1$ is minimax (and has risk less than the minimax risk $p\sigma^2$), if $\mathcal{D}\phi(x) \leq 0$. Equation (2.10) assures that $\mathcal{D}\phi(x) \leq 0$ when (2.4) is satisfied (note that $\hat{\lambda} \geq 0$). This completes the proof. ///

There is no clearly optimal choice of t in $\delta^1$ for this problem. Certain asymptotic arguments suggest that the choice

$$t = \frac{p}{p+2(\alpha-1)}$$

is attractive for larger p or larger $\beta/2(\alpha-1)$, suggesting use of

$$(2.11)\qquad \tilde{t} = \min\{\frac{2p}{t^*} - \frac{4}{p+2(\alpha-1)}, \frac{p}{p+2(\alpha-1)}\}.$$

Note that $t^* = p+2(\alpha-1)$ if $\beta \geq 2(\alpha-1)\sigma^2$ (which will occur when the "guess" for the variance of the $\theta_i$ exceeds $\sigma^2$), so that

$$(2.12)\qquad \tilde{t} = \frac{\min\{2(p-2),p\}}{p+2(\alpha-1)} \qquad \text{if } \beta \geq 2(\alpha-1)\sigma^2,$$

## 3. Results When $\mu$ Is Partially Unknown.

When $\mu$ is known to be of the form $\mu = B\gamma$, where B is a given matrix of rank q and $\gamma$ is unknown with noninformative prior $\pi(\gamma) = 1$, the joint (improper) density of X, $\lambda$, $\gamma$ is

$$f(x,\lambda,\gamma) \propto (\sigma^2+\lambda^{-1})^{-\frac{p}{2}}\exp(-\tfrac{1}{2}(\sigma^2+\lambda^{-1})^{-1}(x-B\gamma)'(x-B\gamma))\lambda^{\alpha-1}\exp(-\tfrac{\lambda\beta}{2})$$

$$\propto \lambda^{\alpha-1}(\sigma^2+\lambda^{-1})^{-\frac{p}{2}}\exp[-\tfrac{1}{2}(\sigma^2+\lambda^{-1})^{-1}(x-B\gamma_x)'(x-B\gamma_x)-\tfrac{\lambda\beta}{2}]$$

$$\cdot\exp[-\tfrac{1}{2}(\sigma^2+\lambda^{-1})^{-1}(\gamma-\gamma_x)'B'B(\gamma-\gamma_x)],$$

where

$$(3.1)\qquad \gamma_x = (B'B)^{-}B'x.$$

(Here $D^{-}$ denotes the generalized inverse of D). It is clear that the MLE of $\gamma$ is $\gamma_x$ and the marginal posterior density of $\lambda$ given x (with $\gamma$ replaced by $\gamma_x$) is

$$f(\lambda|x) \propto \lambda^{\alpha-1}(\sigma^2+\lambda^{-1})^{-\frac{p-q}{2}}\exp[-\tfrac{1}{2}(\sigma^2+\lambda^{-1})^{-1}(x-B\gamma_x)'(x-B\gamma_x)-\tfrac{\lambda\beta}{2}].$$

It follows as before that the MLE, $\hat{\lambda}$, is the solution of the equation $\ell_1(\lambda) = 0$, where

$$(3.2)\qquad \ell_1(\lambda) = (\beta\sigma^4)\lambda^3+(2\beta\sigma^2-2(\alpha-1)\sigma^4)\lambda^2$$
$$+(x'Mx-(p-q)\sigma^2-4(\alpha-1)\sigma^2+\beta)\lambda-((p-q)+2(\alpha-1))$$

and

(3.3) $$M = I - B(B'B)^{-}B'.$$

The hierarchical estimator (1.2) can be written in this case as

$$\delta^2(x) = x - \frac{t\hat{\lambda}\sigma^2}{(1+\sigma^2\hat{\lambda})}(x - B\gamma_x).$$

<u>Theorem 2.</u> Under sum of squares error loss, $\delta^2$ is minimax for

(3.4) $$0 \leq t \leq \frac{2(p-q)}{t^{**}} - \frac{4}{(p-q)+2(\alpha-1)},$$

where

(3.5) $$t^{**} = \max\{p-q+2(\alpha-1),\ p-q+4(\alpha-1)-\beta/\sigma^2\}.$$

<u>Proof.</u> Analogous to that of Theorem 1. ///

The suggested choice of t in $\delta^2$ is

$$\tilde{\tilde{t}} = \min\{\frac{2(p-q)}{t^{**}} - \frac{4}{[p-q+2(\alpha-1)]}, \frac{p-q}{[p-q+2(\alpha-1)]}\},$$

which for $\beta \geq 2(\alpha-1)$ becomes

$$\tilde{\tilde{t}} = \frac{\min\{2(p-q-2),\ p-q\}}{[p-q+2(\alpha-1)]}.$$

## 4. Comparisons and Conclusions.

Since all estimators considered in this paper are minimax, the main question of interest is to investigate overall average performance. Since we are mainly considering application in empirical Bayes or hierarchical Bayes scenarios, suppose that $\theta$ actually has a $\eta_p(\mu,\tau^2 I)$ prior distribution. We consider $\sigma^2 = 1$

and the known $\mu$ case for simplicity; the case of unknown $\mu$ yields similar conclusions.

A convenient way to measure performance with respect to a prior $\pi$ is through the relative savings loss discussed by Efron and Morris (1973); this is given by

$$RSL(\delta) = \frac{r(\pi,\delta)-r(\pi,\delta^{\pi})}{r(\pi,\delta^{0})-r(\pi,\delta^{\pi})},$$

where $\delta^{0}(x) = x$, $\delta^{\pi}$ is the Bayes rule with respect to $\pi$, and

$$r(\pi,\delta) = E^{\pi}R(\theta,\delta).$$

RSL measures the additional overall risk incurred by using $\delta$ instead of the optimal $\delta^{\pi}$, scaled by the total possible improvement over the standard estimator $\delta^{0}$. Thus RSL near zero indicates optimal Bayesian performance with respect to $\pi$, while RSL near one indicates negligible overall improvement over $\delta^{0}$.

The usual empirical Bayes estimator (for the known $\mu$ case) is

$$\delta^{J-S}(x) = x-\min\{1, \frac{(p-2)}{|x-\mu|^{2}}\}(x-\mu),$$

the James-Stein positive part estimator. This assumes no knowledge of $\tau^{2}$, and performs reasonably well for any $\tau^{2}$.

The new estimator $\delta^{1}(x)$ is similar to $\delta^{J-S}$, except that it is designed to do particularly well for $\tau^{2}$ near $E[\lambda^{-1}] = \beta/[2(\alpha-1)]$. Tables 1 and 2 indicate that this is indeed so, for larger p or $E[\lambda^{-1}]$. (When p and $E[\lambda^{-1}]$ are small, the behavior is somewhat different.) For table 1, $E[\lambda^{-1}] = 1$ while for table 2, $E[\lambda^{-1}] = 3$. The choice of $\alpha = 3$ merely implies that the standard deviation of $\lambda^{-1}$ equals the mean, corresponding to a situation of moderate undertainty in the prior mean for $\lambda^{-1}$. In all cases, t was chosen using (2.12).

Table 1. RSL for various P, $\tau^2$, and $\delta^{J-S}$ and $\delta^1$, when $\alpha$ = 3.0 and $\beta$ = 2($\alpha$-1).

| | P | | | | | |
|---|---|---|---|---|---|---|
| | 3 | | 6 | | 10 | |
| $\tau^2$ | RSL($\delta^{J-S}$) | RSL($\delta^1$) | RSL($\delta^{J-S}$) | RSL($\delta^1$) | RSL($\delta^{J-S}$) | RSL($\delta^1$) |
| .5 | .5393 | .6331 | .2427 | .3080 | .1480 | .2040 |
| 1.0 | .5472 | .5579 | .2642 | .2156 | .1699 | .1274 |
| 1.5 | .5547 | .4994 | .2804 | .1618 | .1824 | .0940 |
| 2.0 | .5612 | .4541 | .2920 | .1335 | .1892 | .0845 |
| 2.5 | .5669 | .4188 | .3003 | .1217 | .1931 | .0875 |
| 3.0 | .5718 | .3914 | .3064 | .1202 | .1954 | .0965 |
| 3.5 | .5760 | .3700 | .3110 | .1253 | .1968 | .1080 |
| 4.0 | .5798 | .3535 | .3145 | .1344 | .1978 | .1201 |
| 4.5 | .5832 | .3408 | .3173 | .1459 | .1984 | .1318 |
| 5.0 | .5862 | .3311 | .3195 | .1586 | .1988 | .1428 |
| 5.5 | .5889 | .3238 | .3213 | .1719 | .1991 | .1529 |
| 6.0 | .5913 | .3185 | .3227 | .1854 | .1993 | .1619 |
| 6.5 | .5935 | .3148 | .3239 | .1986 | .1994 | .1701 |
| 7.0 | .5956 | .3124 | .3250 | .2115 | .1996 | .1773 |
| 7.5 | .5975 | .3111 | .3258 | .2238 | .1996 | .1838 |
| 8.0 | .5992 | .3106 | .3265 | .2357 | .1997 | .1896 |
| 8.5 | .6009 | .3109 | .3272 | .2470 | .1998 | .1947 |
| 9.0 | .6024 | .3118 | .3277 | .2577 | .1998 | .1993 |
| 9.5 | .6038 | .3132 | .3282 | .2678 | .1998 | .2034 |
| 10.0 | .6051 | .3150 | .3286 | .2775 | .1999 | .2070 |

Table 2. RSL for various P, $\tau^2$, and $\delta^{J-S}$ and $\delta^1$, when $\alpha$ = 3.0 and $\beta$ = 6($\alpha$-1).

| | P | | | | | |
|---|---|---|---|---|---|---|
| | 3 | | 6 | | 10 | |
| $\tau^2$ | RSL($\delta^{J-S}$) | RSL($\delta^1$) | RSL($\delta^{J-S}$) | RSL($\delta^1$) | RSL($\delta^{J-S}$) | RSL($\delta^1$) |
| .5 | .5393 | .7739 | .2427 | .5124 | .1480 | .3922 |
| 1.0 | .5472 | .7199 | .2642 | .4200 | .1699 | .2964 |
| 1.5 | .5547 | .6736 | .2804 | .3499 | .1824 | .2306 |
| 2.0 | .5612 | .6338 | .2920 | .2965 | .1892 | .1857 |
| 2.5 | .5669 | .5994 | .3003 | .2561 | .1931 | .1554 |
| 3.0 | .5718 | .5695 | .3064 | .2254 | .1954 | .1353 |
| 3.5 | .5760 | .5433 | .3110 | .2022 | .1968 | .1222 |
| 4.0 | .5798 | .5204 | .3145 | .1849 | .1978 | .1142 |
| 4.5 | .5832 | .5002 | .3173 | .1722 | .1984 | .1097 |
| 5.0 | .5862 | .4824 | .3195 | .1630 | .1988 | .1077 |
| 5.5 | .5889 | .4666 | .3213 | .1566 | .1991 | .1074 |
| 6.0 | .5913 | .4525 | .3227 | .1525 | .1993 | .1084 |
| 6.5 | .5935 | .4400 | .3239 | .1501 | .1994 | .1103 |
| 7.0 | .5956 | .4288 | .3250 | .1492 | .1996 | .1127 |
| 7.5 | .5975 | .4188 | .3258 | .1493 | .1996 | .1155 |
| 8.0 | .5992 | .4099 | .3265 | .1504 | .1997 | .1185 |
| 8.5 | .6009 | .4019 | .3272 | .1522 | .1998 | .1218 |
| 9.0 | .6024 | .3947 | .3277 | .1545 | .1998 | .1250 |
| 9.5 | .6038 | .3883 | .3282 | .1573 | .1998 | .1283 |
| 10.0 | .6051 | .3825 | .3286 | .1604 | .1999 | .1316 |

Although the tables only deal with small and moderate $\tau^2$, it is interesting to note that, as $\tau^2 \to \infty$, $RSL(\delta^1) \to RSL(\delta^*)$, where (assuming $\beta \geq 2(\alpha-1)\sigma^2$ for convenience)

$$\delta^*(x) = x - \frac{\min\{2(p-2),p\}\sigma^2}{|x-\mu|^2}(x-\mu).$$

The RSL of $\delta^*$ is very similar to that of $\delta^{J-S}$, especially when p is moderate or large. Thus, even if the prior information concerning the variance of the $\theta_i$ is completely wrong and $\tau^2$ is huge, $\delta^1$ will be comparable to $\delta^{J-S}$.

We cannot give unqualified endorsement of $\delta^1$ or $\delta^2$ over the more familiar James-Stein type estimators, or over, say, the robust generalized Bayes estimators in Berger (1980), because there are too many variables to study all possibilities. Furthermore, from a practical perspective it may be questionable to demand complete minimaxity.

In any case, the results here are of theoretical interest because they

(i) Deal for the first time in the "Stein estimation" literature with estimators which combine empirical Bayes type exchangeability structure with subjective inputs;

(ii) Indicate that verification of minimaxity is possible even for highly complicated estimators which cannot even be easily written in closed form.

## REFERENCES


[1] Berger, J.(1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist., 8, 716-761.

[2] Berger, J.(1982). Selecting a minimax estimator of a multivariate normal mean. Ann. Statist., 10, 81-92.

[3] Berger, J.(1985). Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York.

[4] Chen, S. Y.(1983). Restricted Risk Bayes Estimator. Ph.D. Thesis. Department of Statistics, Purdue University.

[5] Efron, B and Morris, C.(1973). Stein's estimation rule and its competitors-an empirical Bayes approach. J. Amer. Statist. Assoc., 68, 117-130.

[6] James, W. and Stein, C.(1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Stat. Prob., 1, 361-379. University of California Press.

[7] Lindley, D. V. and Smith, A. F. M.(1972). Bayes estimators for linear model (with discussion). J. Roy. Statist. Soc. B, 34, 1-41.

[8] Stein, C.(1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist., 9, 1135-1151.

[9] Strawderman, W. E.(1971). Proper Bayes minimax estimators of the multivariate normal mean. Ann. Math. Statist., 42, 385-388.