Frequentist Behavior of Robust Bayes Estimates of Normal Means

by

Anirban DasGupta* and William J. Studden**
Purdue University

Technical Report #88-36C

Department of Statistics
Purdue University

August 1988
March 1989

* Research supported by NSF Grant DMS 8702620.
** Research supported by NSF Grant DMS 8802535.

◇

Abstract

In the canonical normal problem where $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, \Sigma_0)$ (Σ_0 known) and $\underset{\sim}{\theta}$ has a prior π belonging to a suitable family of priors Γ, the important practical issue of choosing one specific Bayes action from the collection of all Bayes actions corresponding to the priors in Γ is addressed. Several different methods are considered and the consequences are evaluated. It turns out that the method of posterior Γ-minimaxity gives a sound answer if the priors that are not "compatible" with the observed data are eliminated before evaluating the posterior minimax action. The results derived involve application of the rich minimax theory due to Lehmann, Wald and others to a novel geometric game of independent mathematical interest. The frequentist risks of the derived procedures are also considered and it is found that sometimes the procedures derived are frequentist minimax, and even when they are not, they mostly have very satisfactory risk behavior.

1. Introduction. Considerable attention has recently been given to the issue of sensitivity of Bayesian analysis to the choice of the prior. Thus, given a likelihood function ℓ, a loss function L, and a prior π belonging to a suitable family of priors Γ, substantial research is now being done by many workers on the amount of variation in various Bayesian measures of interest (like the Bayes rule itself or the posterior risk etc.) due to variation in the prior. Robustness is present if these measures of variations are small. For many specific results and general exposition, see Berger and Sivaganesan (1986), Berger (1987), DasGupta and Studden (1988a,b), DeRobertis and Hartigan (1981), Good and Crook (1987), Kadane and Chuang (1978), Leamer (1982), Polasek (1985) etc. Virtually nothing, however, has been done on a question of practical importance described in the following paragraph.

Considerable evidence has now accumulated that whenever the observed data are "compatible" with the priors in the family Γ, robustness with respect to the choice of the prior will usually obtain and otherwise robustness will usually not obtain. For example, if $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, I)$ and the priors on $\underset{\sim}{\theta}$ have a fixed mean $\underset{\sim}{\mu}$, robustness will *typically* obtain if the observed $\underset{\sim}{y}$ is "near" $\underset{\sim}{\mu}$. The question of practical importance is what do we do when $\underset{\sim}{y}$ is not near $\underset{\sim}{\mu}$ and robustness is absent. A number of workers in the area believe that in such a case the observed data "rule out" some of the priors that were originally considered plausible. For example, if a normal prior with mean 0 and a Cauchy prior with median 0 were considered a priori plausible in a one dimensional normal problem, then it is argued that an observed $y = 10$ rules out the normal prior, but not the Cauchy prior. In general, the collection of plausible priors will shrink on observing the data and therefore one might hope that a repetition of the sensitivity analysis with the new smaller class of priors will demonstrate robustness.

While sympathizing with the stand that the observed data *may* rule out some of the priors originally considered plausible, we propose in this article a specific method of selecting *one Bayes action* from the collection of the totality of possible Bayes actions as the prior changes in the class Γ. We do this within the conditional Bayes framework, i.e., while selecting one specific action for actual use and referral to the user, we do not integrate on the unobserved data (although, once an action is selected for each specific y, we do look at the frequentist behavior of the strategy resulting from the actions combined for different y). An important reason for providing a concrete method to select one action is that there is no guarantee that if the "unlikely" priors are eliminated, then a repetition of the sensitivity analysis will automatically give robustness.

In the following we briefly describe the set up considered and results obtained in this article and also describe the organization of the paper.

Consider the canonical normal problem where $\underset{\sim}{Y}_{p\times 1} \sim N(\underset{\sim}{\theta}, \Sigma_0)$, where $\underset{\sim}{\theta}$ is unknown and Σ_0 is a known p.d. matrix. We let $\underset{\sim}{\theta}$ have a prior belonging to the class

$$\Gamma = \{\pi\colon\ \pi \text{ is } N(\underset{\sim}{\mu}, \Sigma), \quad \underset{\sim}{\mu} \text{ fixed}, \Sigma_1 \leq \Sigma \leq \Sigma_2\}. \tag{1.1}$$

The reason for considering conjugate priors is that in many problems they provide a rich enough class of priors for a conscientious sensitivity analysis and yet they are mathematically attractive. Families of nonconjugate priors are considered in DasGupta and Studden (1988c). Note that $\underset{\sim}{\mu}$ is kept fixed but not the prior variance-covariance matrix Σ because it will typically be easier to elicit the location of the prior than its higher moments. Priors of the type (1.1) were first suggested by Leamer (1978, 1982), and Polasek (1984). In addition, also see DasGupta and Studden (1988a) for an extensive discussion of the family (1.1). Note that the canonical normal linear model where the error variance is known is subsumed in our set up. Non normal priors as well as non normal models will be considered elsewhere.

With the regression model in mind, we consider in this article estimation of $\underset{\sim}{\theta}$ as well as a linear combination $\underset{\sim}{c}'\underset{\sim}{\theta}$ under squared-error loss. This will cover problems of estimating the vector of regression coefficients and estimating the mean response (or predicting a future value) for a fixed level of the regressor variables. In each of these problems, we derive the conditional Γ-minimax procedure, i.e., the procedure δ that minimizes (for fixed data $\underset{\sim}{y}$) $\sup_{\pi\in\Gamma} r_y(\pi,\delta)$ where $r_y(\pi,\delta)$ denotes the posterior expected loss of δ corresponding to the prior π. It is seen that there exists an estimate $\hat{\underset{\sim}{\theta}}$ for $\underset{\sim}{\theta}$ such that $\underset{\sim}{c}'\hat{\underset{\sim}{\theta}}$ is conditional Γ-minimax for $\underset{\sim}{c}'\underset{\sim}{\theta}$ for every $\underset{\sim}{c}$ and for every $\underset{\sim}{y}$, but $\hat{\underset{\sim}{\theta}}$ *may not be* conditional Γ-minimax for $\underset{\sim}{\theta}$ for every $\underset{\sim}{y}$. It is shown that if Σ_0, Σ_1, and Σ_2 are each proportional to the identity matrix then in fact $\hat{\underset{\sim}{\theta}}$ is conditional Γ-minimax for $\underset{\sim}{\theta}$ for every $\underset{\sim}{y}$. These results are derived in section 2.

In this context, we have also considered the Bayes action for the Type-II maximum likelihood prior (the Type-II maximum likelihood 'prior' is defined as the π that maximizes, for given $\underset{\sim}{y}$, the marginal likelihood of $\underset{\sim}{y}$; see Good (1965) and Berger and Berliner (1986)). In the process, a new Bayesian interpretation of the Stein-estimate emerges. These results are contained in section 3. The Type-II ML prior is connected with the idea of "ruling out" some priors which are considered unlikely after observing the data. To put it in another way, the Type-II ML prior is the most likely prior after observing the data. Consequently, priors that are "far" from the ML prior are the priors that should be considered unlikely after observing the data. Keeping this in mind as well as the need to be robust with respect to other priors in Γ, we consider in section 4 the problems of deriving the action that is conditional Γ-minimax subject to being ε-Bayes with respect to the Type-II ML prior or conversely the action that is Bayes with respect to the ML prior subject to being ε-minimax. The second of these problems is related to the restricted-risk Bayes approach, first suggested by Hodges and Lehmann (1952). For more recent works see Berger (1982), Kempthorne (1986), DasGupta and Berger (1986), DasGupta and Rubin (1987a), DasGupta and Bose (1987b) etc. The first problem relates to restricted minimaxity, considered among others by Bickel (1980).

Even though from a conditional perspective it may be desirable to derive one's decision rule by conditioning on the given data $\underset{\sim}{y}$, it is important to consider its average performance over repeated use. This is done by considering the usual risk functions of the different procedures derived in this article. Using a result of Strawderman (1971), it is actually proved that the Type-II ML Bayes procedure is frequentist minimax for $p \geq 3$. Some of the other procedures, although not minimax, have remarkably good risk properties. This is considered in section 5. Section 6 contains some concluding remarks and a brief discussion.

2. Estimation of an arbitrary linear combination $\underset{\sim}{c}'\underset{\sim}{\theta}$. Although the vector estimation problem usually arises more naturally than estimation of a linear combination, we consider the linear combination problem first because our results are more general in this case and more importantly, because the conditional Γ-minimax procedure for this case naturally suggests what the solution may be for the vector estimation case. We need the following result from DasGupta and Studden (1988a).

Theorem 2.1. For the family of priors (1.1) and any arbitrary p-dimensional vector $\underset{\sim}{c}$, let

$S(\underset{\sim}{c})$ denote the set of two-dimensional vectors of posterior mean and posterior variance of $\underset{\sim}{c}'\underset{\sim}{\theta}$, i.e.,

$$S(\underset{\sim}{c}) = \left\{\left(E(\underset{\sim}{c}'\underset{\sim}{\theta}|\underset{\sim}{y}),\ \ \mathrm{Var}(c'\underset{\sim}{\theta}|\underset{\sim}{y})\right)|\pi \in \Gamma\right\}. \tag{2.1}$$

The set $S(\underset{\sim}{c})$ is an ellipse

$$\{\underset{\sim}{u}\colon\ (\underset{\sim}{u} - \underset{\sim}{u}_0)'D^{-1}(\underset{\sim}{u} - \underset{\sim}{u}_0) \le 1\}, \tag{2.2}$$

where $\underset{\sim}{u}_0 = (\underset{\sim}{c}'\underset{\sim}{\mu} + \underset{\sim}{c}'\overline{\Lambda}\underset{\sim}{v},\ \underset{\sim}{c}'\overline{\Lambda}\underset{\sim}{c})'$, and

$$D = A^2\begin{pmatrix} \underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} & \underset{\sim}{c}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} \\ & \underset{\sim}{c}'(\Lambda_2 - \Lambda_1)\underset{\sim}{c} \end{pmatrix},$$

where $\Lambda_i = (\Sigma_0^{-1} + \Sigma_i^{-1})^{-1}$, $i = 1, 2, \overline{\Lambda} = \frac{\Lambda_2+\Lambda_1}{2}$, $\underset{\sim}{v} = \Sigma_0^{-1}(\underset{\sim}{y} - \underset{\sim}{\mu})$, and $A^2 = \dfrac{\underset{\sim}{c}'(\Lambda_2 - \Lambda_1)\underset{\sim}{c}}{4}$.

Henceforth, the matrix D defined above will be denoted by

$$D = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ & \sigma_2^2 \end{pmatrix},$$

where σ_1^2, σ_2^2, ρ are defined in the obvious way. Before deriving the conditional Γ-minimax rule for estimation of $\underset{\sim}{c}'\underset{\sim}{\theta}$, we note that under any prior π and for any action δ, the posterior expected loss of δ is given by

$$\begin{aligned} r_{\underset{\sim}{y}}(\pi, \delta) &= E\left((\delta - \underset{\sim}{c}'\underset{\sim}{\theta})^2|\underset{\sim}{y}\right) \\ &= (\delta - u)^2 + v, \end{aligned} \tag{2.3}$$

where $u = E(\underset{\sim}{c}'\underset{\sim}{\theta}|\underset{\sim}{y})$ and $v =$ $\mathrm{Var}(c'\underset{\sim}{\theta}|\underset{\sim}{y})$. Since Theorem 2.1 gives that for Γ as in (1.1), the points (u, v) form an ellipse $S = S(\underset{\sim}{c})$, the problem of deriving the conditional Γ-minimax action reduces to the geometric problem of finding a δ that minimizes $\sup\limits_{(u,v)\in S}\{(\delta - u)^2 + v\}$. This geometrical representation of the minimax problem later turns out to be very useful in visualizing various aspects of the proofs. Note that easy arguments show that the minimax action will have to belong to the projection of S onto the u-axis even if δ is permitted to belong to the entire real line. There will, therefore, be no loss of generality in assuming that $(\delta - u_{01})^2 \le \sigma_1^2$, where $u_{01} = \underset{\sim}{c}'\underset{\sim}{\mu} + \underset{\sim}{c}'\overline{\Lambda}\underset{\sim}{v}$.

Theorem 2.2. Let S be the ellipse (2.2). Then the δ minimizing $\sup\limits_{(u,v)\in S}\{(\delta - u)^2 + v\}$ is given by

$$\begin{aligned} \delta &= u_{01} + \rho\sigma_1 && \text{if } \frac{2\sigma_1^2}{\sigma_2} \le 1 \\ &= u_{01} + \frac{\rho\sigma_2}{2\sigma_1} && \text{if } \frac{2\sigma_1^2}{\sigma_2} > 1. \end{aligned}$$

Proof: It's easily seen that u_{01} can be assumed to be zero. We will first consider the case $\frac{2\sigma_1^2}{\sigma_2} > 1$.

First note that the supremum clearly occurs for (u, v) on the upper boundary of the ellipse. We then have the representation

$$v = u_{02} + \rho\frac{\sigma_2}{\sigma_1}u + \sigma_2\sqrt{(1-\rho^2)(1-\frac{u^2}{\sigma_1^2})}, \tag{2.4}$$

for any point on the upper boundary of S. Consequently for the action $\delta = \rho\frac{\sigma_2}{2\sigma_1}$, the (posterior) risk at a point (u, v) on the upper boundary of S equals

$$r_1(u) = u_{02} + \left(\frac{\rho\sigma_2}{2\sigma_1} - u\right)^2 + \frac{\rho\sigma_2}{\sigma_1}u + \sigma_2\sqrt{(1-\rho^2)(1-\frac{u^2}{\sigma_1^2})}, \tag{2.5}$$

where $-\sigma_1 \le u \le \sigma_1$.

The derivative of $r_1(u)$ equals

$$r_1'(u) = u\left[2 - \frac{\sigma_2\sqrt{1-\rho^2}}{\sigma_1^2\sqrt{1-\frac{u^2}{\sigma_1^2}}}\right], \tag{2.6}$$

from which it follows that $r_1(u)$ is monotone increasing for u near $-\sigma_1$, monotone decreasing for u near σ_1, and $r_1'(u) = 0$ at $u = 0, \pm\sigma_1\sqrt{1-\frac{\sigma_2^2(1-\rho^2)}{4\sigma_1^4}}$. Note $\frac{\sigma_2^2(1-\rho^2)}{4\sigma_1^4} < 1$ since $\frac{2\sigma_1^2}{\sigma_2} > 1$ and $\rho^2 \le 1$. It is, therefore, clear that the maximum of $r_1(u)$ is attained at $u = \pm u^*$ where $u^* = \pm\sigma_1\sqrt{1-\frac{\sigma_2^2(1-\rho^2)}{4\sigma_1^4}}$. It is easy to check that under the condition $\frac{2\sigma_1^2}{\sigma_2} > 1$, the action $\delta = \rho\frac{\sigma_2}{2\sigma_1}$ lies between $\pm u^*$. In this situation $\rho\frac{\sigma_2}{2\sigma_1}$ must minimize $\sup_{(u,v)\in S}\{(\delta - u)^2 + v\}$ since by moving away from the action $\rho\frac{\sigma_2}{2\sigma_1}$ the risk at one of the two points $\pm u^*$ increases. This completes the proof in the case $\frac{2\sigma_1^2}{\sigma_2} > 1$.

For the case $\frac{2\sigma_1^2}{\sigma_2} \le 1$, we will prove that the maximum (posterior) risk of the action $\delta = \rho\sigma_1$ is attained at the point (u, v) on the upper boundary of S corresponding to $u = \rho\sigma_1$. This will then directly imply that $\rho\sigma_1$ must be the minimax action in this case.

Thus we are required to show that the quantity $(\rho\sigma_1 - u)^2 + v$, where v is given by (2.4) has a maxmum at $u = \rho\sigma_1$. The value at $u = \rho\sigma_1$ is readily seen to be $u_{02} + \sigma_2$, so it suffices to show

$$(u-\rho\sigma_1)^2+\rho\frac{\sigma_2}{\sigma_1}u+\sigma_2\sqrt{(1-\rho^2)(1-\frac{u^2}{\sigma_1^2})}\leq\sigma_2.$$

However, this can be seen to be true, with a small amount of algebra, by substituting $u=\sigma_1\cos\theta$ for $0\leq\theta\leq\pi$ and using $2\sigma_1^2<\sigma_2$.

The proof of the theorem is now complete.

Corollary 2.3. For the problem of estimating $\underset{\sim}{c}'\underset{\sim}{\theta}$ under squared error loss and for the family of priors as in (1.1), the conditional Γ-minimax action is given by

$$\begin{aligned}\delta_{\underset{\sim}{c}}(\underset{\sim}{y})&=\underset{\sim}{c}'(\Lambda_2\underset{\sim}{v}+\underset{\sim}{\mu})&&\text{if } \underset{\sim}{v}'(\Lambda_2-\Lambda_1)\underset{\sim}{v}\leq 1\\&=\underset{\sim}{c}'(\Lambda^*\underset{\sim}{v}+\underset{\sim}{\mu})&&\text{if } \underset{\sim}{v}'(\Lambda_2-\Lambda_1)\underset{\sim}{v}>1,\end{aligned}\tag{2.7}$$

where $\Lambda^*=\frac{\Lambda_2+\Lambda_1}{2}+\frac{\Lambda_2-\Lambda_1}{2\underset{\sim}{v}'(\Lambda_2-\Lambda_1)\underset{\sim}{v}}$, and $\underset{\sim}{v}$, Λ_1, Λ_2 are as in Theorem 2.1.

Proof: Direct algebra using Theorem 2.2 and the definitions of σ_1^2, σ_2^2, and ρ in the matrix D.

Notice that corollary 2.3 implies that there exists an estimate $\hat{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$, independent of $\underset{\sim}{c}$, such that $\underset{\sim}{c}'\hat{\underset{\sim}{\theta}}$ is posterior Γ-minimax for $\underset{\sim}{c}'\underset{\sim}{\theta}$ for every $\underset{\sim}{c}$. The estimate

$$\begin{aligned}\hat{\underset{\sim}{\theta}}&=\Lambda_2\underset{\sim}{v}+\underset{\sim}{\mu}&&\text{if }\ \underset{\sim}{v}'(\Lambda_2-\Lambda_1)\underset{\sim}{v}\leq 1\\&=\Lambda^*\underset{\sim}{v}+\underset{\sim}{\mu}&&\text{if }\ \underset{\sim}{v}'(\Lambda_2-\Lambda_1)\underset{\sim}{v}>1\end{aligned}$$

serves this purpose. A natural guess for a posterior Γ-minimax rule for estimating $\underset{\sim}{\theta}$ is this estimate $\hat{\underset{\sim}{\theta}}$. We will shortly give an example showing that this is not the case in general. But before that we need a representation analogous to (2.3) for the (posterior) risk of an action $\underset{\sim}{\delta}$ in the vector estimation problem. Note that for an action $\underset{\sim}{\delta}$ and a generic prior π in (1.1), the posterior risk equals

$$\begin{aligned}r_{\underset{\sim}{y}}(\pi,\delta)&=E||\underset{\sim}{\theta}-\underset{\sim}{\delta}||^2|\underset{\sim}{y}\\&=||\hat{\underset{\sim}{\theta}}_\pi-\underset{\sim}{\delta}||^2+\ \text{tr}\ D_y(\underset{\sim}{\theta}),\end{aligned}\tag{2.8}$$

where $\hat{\underset{\sim}{\theta}}_\pi$ is the posterior mean of $\underset{\sim}{\theta}$ under the prior π and $D_y(\underset{\sim}{\theta})$ denotes the posterior variance-covariance matrix of $\underset{\sim}{\theta}$ under π. If π is the $N(\underset{\sim}{\mu},\Sigma)$ prior, then $\hat{\theta}_\pi=\Lambda\underset{\sim}{v}+\underset{\sim}{\mu}$ where

$\Lambda = (\Sigma_0^{-1} + \Sigma^{-1})^{-1}$, and $D_y(\underset{\sim}{\theta}) = \Lambda$. Consequently, the risk (2.8), parametrized by Σ (or equivalently, by Λ), equals

$$r(\Lambda, \underset{\sim}{\delta}) = ||\underset{\sim}{\delta} - (\Lambda \underset{\sim}{v} + \underset{\sim}{\mu})||^2 + \text{ tr } \Lambda.$$

Clearly, $\underset{\sim}{\mu}$ could be assumed zero without any loss of generality. We thus need to solve the game where the payoff is

$$r(\Lambda, \underset{\sim}{\delta}) = ||\underset{\sim}{\delta} - \Lambda \underset{\sim}{v}||^2 + \text{ tr } \Lambda, \tag{2.9}$$

where $\Lambda_1 \leq \Lambda \leq \Lambda_2$ and $\underset{\sim}{\delta}$, without any loss, can be assumed to belong to the (convex) set of all possible vectors $\Lambda \underset{\sim}{v}$ for $\Lambda_1 \leq \Lambda \leq \Lambda_2$ (even if $\underset{\sim}{\delta}$ is permitted to belong to the entire $\mathbb{R}^p$, the minimax rule must belong to this smaller convex set. For the sake of completeness, we note here that this convex set is a p-dimensional ellipsoid. See Theorem 2.1 in DasGupta and Studden (1988a), and also Leamer (1982) and Polasek (1984)).

The following example shows that $\hat{\theta}$ suggested by Corollary 2.3 is not the posterior minimax rule in general.

Example 1. Let $p = 2$, $\Lambda_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\Lambda_2 = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$, $\underset{\sim}{\mu} = \underset{\sim}{0}$. It is easy to prove that for each unit vector $\underset{\sim}{a}$, the matrix $\Lambda = \Lambda_2 - (\Lambda_2 - \Lambda_1)^{1/2} \underset{\sim}{a}\underset{\sim}{a}'(\Lambda_2 - \Lambda_1)^{1/2}$ satisfies $\Lambda_1 \leq \Lambda \leq \Lambda_2$. We choose two specific unit vectors $\underset{\sim}{a}_1$, $\underset{\sim}{a}_2$ defined as

$$\underset{\sim}{a}_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \underset{\sim}{a}_2 = \begin{pmatrix} -\cos\theta \\ \sin\theta \end{pmatrix},$$

where $0 < \theta < \frac{\pi}{2}$ is defined as $\cos^2\theta = \frac{2-\sqrt{3}}{2}$ ($\theta \approx .381\pi$). Also let $\underset{\sim}{v} = (1,1)$, so that $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\ \underset{\sim}{v} = 2 > 1$. The rule $\hat{\underset{\sim}{\theta}} = \Lambda^* \underset{\sim}{v}$ is easily seen to be equal to (1.75, 1.75). However, the action $\hat{\underset{\sim}{\theta}}^* = (\sqrt{3}, \sqrt{3})$ attains its maximum risk at $\Lambda = \Lambda_0$, Λ_2, Λ_3, where Λ_0, Λ_3 are matrices obtained by using the formula $\Lambda_2 - (\Lambda_2 - \Lambda_1)^{1/2} \underset{\sim}{a}\underset{\sim}{a}'(\Lambda_2 - \Lambda_1)^{1/2}$ with $\underset{\sim}{a} = \underset{\sim}{a}_1, \underset{\sim}{a}_2$ respectively (the proof involves relatively simple matrix algebra and calculus; the details can be obtained from the authors). But the vector $(\sqrt{3}, \sqrt{3})$ is inside the convex hull of the points $\Lambda_i \underset{\sim}{v}$, $i = 0, 2, 3$; hence it must be the minimax action (by moving away from the point $(\sqrt{3}, \sqrt{3})$, the risk at at least one of the three Λ's defined above increases). Interestingly, although $\Lambda^* \underset{\sim}{v} = (1.75, 1.75)$ is not minimax, the minimax action is extremely close to it.

We comment without proof that whenever $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v}$ is less than or equal to 1, $\hat{\underset{\sim}{\theta}} = \Lambda_2 \underset{\sim}{v} + \underset{\sim}{\mu}$ is in fact the conditional Γ-minimax action. It's only in the case $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} > 1$ that $\hat{\underset{\sim}{\theta}}$ may fail to be minimax. Even in this case, $\hat{\underset{\sim}{\theta}}$ is minimax provided $\Lambda_2 - \Lambda_1$ is a multiple of the identity matrix. This is the assertion of the next theorem. We do not know if the condition that $\Lambda_2 - \Lambda_1$ be proportional to the identity matrix is necessary for the theorem to be valid, but some crucial steps in the proof of the theorem are false without this assumption.

Theorem 2.4. For the problem of estimating $\underset{\sim}{\theta}$ under squared error loss, the rule

$$\begin{aligned} \hat{\underset{\sim}{\theta}} &= \Lambda_2 \underset{\sim}{v} + \underset{\sim}{\mu} && \text{if } \underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} \le 1 \\ &= \Lambda^* \underset{\sim}{v} + \underset{\sim}{\mu} && \text{if } \underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} > 1 \end{aligned}$$

is conditional Γ-minimax if $\Lambda_2 - \Lambda_1$ is a multiple of the identity matrix.

Proof: Let $\overline{\Lambda} = \frac{\Lambda_1 + \Lambda_2}{2}$. Then any Λ such that $\Lambda_1 \le \Lambda \le \Lambda_2$ satisfies $-\frac{\Lambda_2 - \Lambda_1}{2} \le \Lambda - \overline{\Lambda} \le \frac{\Lambda_2 - \Lambda_1}{2}$. We will give the proof of the theorem in the case $\frac{\Lambda_2 - \Lambda_1}{2} = I$; the case of a general multiple is exactly similar.

Under the assumption $\frac{\Lambda_2 - \Lambda_1}{2} = I$, Λ satisfies $\Lambda = \overline{\Lambda} + C$, where $-I \le C \le I$. Also notice that $\underset{\sim}{\mu}$ can be assumed to be zero without any loss of generality. For the case $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} \le 1$ (i.e., $\underset{\sim}{v}'\underset{\sim}{v} \le \frac{1}{2}$), we will prove that the action $\Lambda_2 \underset{\sim}{v}$ attains its maximum risk at $\Lambda = \Lambda_2$, which will imply that $\Lambda_2 \underset{\sim}{v}$ must be minimax. To do this, we will need to show that for $-I \le C \le I$,

$$\begin{aligned} & ||\Lambda_2 \underset{\sim}{v} - (\overline{\Lambda} \underset{\sim}{v} + C \underset{\sim}{v})||^2 + \text{ tr } \overline{\Lambda} + \text{ tr } C \le \text{ tr } \Lambda_2 \\ \Leftrightarrow & ||(I - C)\underset{\sim}{v}||^2 + \text{ tr } C \le p \end{aligned} \tag{2.10}$$

But,

$$\begin{aligned} & ||(I - C)\underset{\sim}{v}||^2 + \text{ tr } C \\ & = \underset{\sim}{v}'(I - C)^2 \underset{\sim}{v} + \text{ tr } C \\ & \le 2\underset{\sim}{v}'(I - C)\underset{\sim}{v} + \text{ tr } C (\because\ 0 \le I - C \le 2I) \\ & \le \lambda_{\max}(I - C) + \text{ tr } C \\ & = 1 + \sum_{i=2}^{p} \lambda_{(i)} \quad \text{(where } \lambda_{(1)} \le \ldots \le \lambda_{(p)} \text{ are the eigenvalues of } C) \\ & \le p, \text{ as was required to be proved.} \end{aligned}$$

For the case $\underset{\sim}{v}'\underset{\sim}{v} > \frac{1}{2}$, we will prove that the action $\Lambda^*\underset{\sim}{v}$ attains its maximum risk at Λ_1^* and Λ_2 where $\Lambda_1^* = \Lambda_1 + 2(I - \frac{\underset{\sim}{v}\underset{\sim}{v}'}{\underset{\sim}{v}'\underset{\sim}{v}})$ and that $\Lambda^*\underset{\sim}{v}$ is a convex combination of $\Lambda_1^*\underset{\sim}{v}$ and $\Lambda_2\underset{\sim}{v}$, which will imply that $\Lambda^*\underset{\sim}{v}$ must be minimax (incidentally Λ_1^* defined above maximizes $\text{tr}\Lambda$ among Λ satisfying $\Lambda\underset{\sim}{v} = \Lambda_1\underset{\sim}{v}$).

First we will prove that the risk of $\Lambda^*\underset{\sim}{v}$ at any Λ never exceeds its risk at Λ_2. Thus, we need to show that for $-I \leq C \leq I$,

$$\begin{aligned} &||\Lambda^*\underset{\sim}{v} - (\overline{\Lambda}\underset{\sim}{v} + C\underset{\sim}{v})||^2 + \text{ tr } \overline{\Lambda} + \text{ tr } C \leq ||\Lambda^*\underset{\sim}{v} - \Lambda_2\underset{\sim}{v}||^2 + \text{ tr } \Lambda_2 \\ \Leftrightarrow &||\frac{\underset{\sim}{v}}{2\underset{\sim}{v}'\underset{\sim}{v}} - C\underset{\sim}{v}||^2 + \text{ tr } C \leq ||\frac{\underset{\sim}{v}}{2\underset{\sim}{v}'\underset{\sim}{v}} - \underset{\sim}{v}||^2 + p \\ \Leftrightarrow &\underset{\sim}{v}'C^2\underset{\sim}{v} - \frac{\underset{\sim}{v}'C\underset{\sim}{v}}{\underset{\sim}{v}'\underset{\sim}{v}} + \text{ tr } C \leq \underset{\sim}{v}'\underset{\sim}{v} + p - 1. \end{aligned} \tag{2.11}$$

Using the facts that $\underset{\sim}{v}'C^2\underset{\sim}{v} \leq \underset{\sim}{v}'\underset{\sim}{v}$ and $\frac{\underset{\sim}{v}'C\underset{\sim}{v}}{\underset{\sim}{v}'\underset{\sim}{v}} \geq \lambda_{\min}(C)$, (2.11) follows immediately. To show that $\Lambda^*\underset{\sim}{v}$ also attains its maximum risk at Λ_1^*, we need to show that equality holds in (2.11) for $C = I - \frac{2\underset{\sim}{v}\underset{\sim}{v}'}{\underset{\sim}{v}'\underset{\sim}{v}}$; this can be seen very easily. That $\Lambda^*\underset{\sim}{v}$ is a convex combination of $\Lambda_1^*\underset{\sim}{v}$ and $\Lambda_2\underset{\sim}{v}$ follows on using $\underset{\sim}{v}'\underset{\sim}{v} > \frac{1}{2}$.

It thus follows that for the game described in (2.9), nature's least favorable prior concentrates on only Λ_2 for $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} \leq 1$ and randomizes between Λ_1^* and Λ_2 for $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} > 1$. It is informative that Λ_1 is not in the support of the least favorable prior, but the matrix maximizing $\text{tr}\Lambda$ (i.e., the overall posterior risk) among Λ making $\Lambda\underset{\sim}{v} = \Lambda_1\underset{\sim}{v}$ (i.e., Λ for which the Bayes rule is the same as that for Λ_1) is in the support of the least favorable prior. Also, easy calculation gives that if $\Sigma_0 = I$, $\Sigma_1 = \alpha I$, and $\Sigma_2 = \infty I$ (i.e., Σ varies in the range $\Sigma \geq \alpha I$), then the conditional Γ-minimax rule $\hat{\underset{\sim}{\theta}}$ takes the form (for $\underset{\sim}{\mu} = \underset{\sim}{0}$)

$$\begin{aligned} \hat{\underset{\sim}{\theta}} &= \underset{\sim}{y} \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \leq \alpha + 1 \\ &= \underset{\sim}{y} - \frac{1}{2(\alpha+1)}[1 - \frac{\alpha+1}{\underset{\sim}{y}'\underset{\sim}{y}}]\underset{\sim}{y} \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} > \alpha + 1. \end{aligned} \tag{2.12}$$

While a connection to the James-Stein estimate is immediate from (2.12), the exact form seems to be different from various estimators known in the literature. The striking resemblance of $\hat{\underset{\sim}{\theta}}$ to the restricted-risk Bayesian estimates in Berger (1982) and DasGupta and Rubin (1987a) is interesting.

3. Type-II Maximum likelihood. In this section, we derive the Type-II ML prior and the corresponding Bayes actions. Since we have a parametric family of priors, the Type-II ML approach is simply the usual parametric empirical Bayes approach where unknown hyperparameters of a prior are estimated from the data using the maximum likelihood method (although other methods such as unbiasedness and the method of moments have also been used). There is an enormous literature on empirical Bayesian methodology. See Maritz (1970) and Morris (1983) for discussion and other references. Note that we are not suggesting that the prior ought to or can (in a philosophical sense) depend on the data; we are simply suggesting use of the type-II ML prior as a usually effective operational technique for weeding out the 'unlikely' priors. Indeed, our main goal is to carry out a conditional minimax analysis after such unlikely priors have been eliminated. The reason for such an elimination is that a straight minimax rule sometimes tries to protect against priors that are simply unlikely in the light of the obtained data. The type-II ML analysis in this section should be viewed as simply a step towards that goal. However, the results of this section are of some independent interest as well; for example, one by-product is a formally new Bayesian interpretation of the Stein-type estimates. For a lucid discussion of the type-II ML methods, see sections 1 and 6 in Berger and Berliner (1986).

The following theorem is the main result required for the derivation of the type-II ML Bayes rule.

Theorem 3.1. Let $\underset{\sim}{Y} \sim N(\underset{\sim}{0}, \Sigma)$, where for some constants $\lambda \geq 0$, $1 < k < \infty$, $(k\lambda)^{-1}I \leq \Sigma \leq \lambda^{-1}I$. Then the maximum likelihood estimate of Σ equals

$$
\begin{aligned}
\hat{\Sigma} &= \frac{1}{k\lambda}[I + \frac{k-1}{\underset{\sim}{y}'\underset{\sim}{y}}\underset{\sim}{y}\underset{\sim}{y}'] \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \geq \frac{1}{\lambda} \\
&= \frac{1}{k\lambda}[I + (k\lambda - \frac{1}{\underset{\sim}{y}'\underset{\sim}{y}})\underset{\sim}{y}\underset{\sim}{y}'] \qquad \text{if } \frac{1}{k\lambda} \leq \underset{\sim}{y}'\underset{\sim}{y} \leq \frac{1}{\lambda} \qquad (3.1) \\
&= \frac{I}{k\lambda} \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \leq \frac{1}{k\lambda}.
\end{aligned}
$$

Proof: Let P be an orthogonal matrix such that $P\underset{\sim}{y} = (0\,0\,\ldots 0\,||y||)'$. Note that denoting $R = \begin{pmatrix} R_{11} & \underset{\sim}{u} \\ \underset{\sim}{u}' & r_{pp} \end{pmatrix} = P\Sigma^{-1}P'$ and $c = \underset{\sim}{y}'\underset{\sim}{y}$, the problem of finding the mle of Σ reduces to minimizing $cr_{pp} - \log|R|$ where $\lambda I \leq R \leq k\lambda I$. Since $|R| = |R_{11}|(r_{pp} - \underset{\sim}{u}'R_{11}^{-1}\underset{\sim}{u})$, and since for every R such that $\lambda I \le R \le k\lambda I$, the matrix

$$R^* = \begin{pmatrix} R_{11} & \underset{\sim}{0} \\ \underset{\sim}{0} & r_{pp} \end{pmatrix}$$

also satisfies $\lambda I \le R^* \le k\lambda I$, it follows that for fixed R_{11} and r_{pp}, $cr_{pp} - \log|R|$ is minimized by $R = R^*$. Next, for fixed r_{pp}, $|R_{11}|$ is maximized by $R_{11} = k\lambda I_0$, where I_0 is the identity matrix of order $p-1$. Thus the minimum of $cr_{pp} - \log|R|$ is attained at

$$\hat{R} = \begin{pmatrix} k\lambda I_0 & \underset{\sim}{0} \\ \underset{\sim}{0} & r_0 \end{pmatrix}, \tag{3.2}$$

where r_0 minimizes $cx - \log x$ in the interval $\lambda \le x \le k\lambda$. Note that

$$\begin{aligned} r_0 &= \lambda && \text{if } \frac{1}{c} \le \lambda \\ &= \frac{1}{c} && \text{if } \lambda \le \frac{1}{c} \le k\lambda \\ &= k\lambda && \text{if } \frac{1}{c} \ge k\lambda. \end{aligned} \tag{3.3}$$

Now (3.2) gives

$$\begin{aligned} \hat{R} &= k\lambda I - \begin{pmatrix} 0 & \underset{\sim}{0} \\ \underset{\sim}{0} & k\lambda - r_0 \end{pmatrix}_{-1} \\ \Leftrightarrow \hat{\Sigma} &= \left(k\lambda I - P' \begin{pmatrix} 0 & \underset{\sim}{0} \\ \underset{\sim}{0} & k\lambda - r_0 \end{pmatrix} P \right)^{-1} \\ &= \frac{1}{k\lambda} \left(I - P' \begin{pmatrix} 0 & \underset{\sim}{0} \\ \underset{\sim}{0} & 1 - \frac{r_0}{k\lambda} \end{pmatrix} P \right)^{-1}. \end{aligned} \tag{3.4}$$

Using the definition of P and the well-known matrix identity $(I + \underset{\sim}{u}\underset{\sim}{v}')^{-1} = I - \dfrac{\underset{\sim}{u}\underset{\sim}{v}'}{1 + \underset{\sim}{u}'\underset{\sim}{v}}$, the result now follows using (3.3) and (3.4).

In our set up we have $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, \Sigma_0)$ and $\underset{\sim}{\theta} \sim N(\underset{\sim}{\mu}, \Sigma)$ where $\Sigma_1 \le \Sigma \le \Sigma_2$. Thus the marginal distribution of $\underset{\sim}{Y} - \underset{\sim}{\mu}$ is $N(\underset{\sim}{0}, \Sigma_0 + \Sigma)$. If now Σ_0, Σ_1, Σ_2 are each proportional to the identity matrix, then using Theorem 3.1, the mle of Σ may be derived. This in turn gives the mle of $\Lambda = (\Sigma_0^{-1} + \Sigma^{-1})^{-1}$, which produces an expression for the corresponding Bayes action $\Lambda\underset{\sim}{v} + \underset{\sim}{\mu}$. That, of course, is the type-II ML Bayes rule. Thus the following theorem is straightforward. We omit the proof.

Theorem 3.2. Let $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, I)$ and $\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, \Sigma)$, where $n_1 I \le \Sigma \le n_2 I$ for some $n_2 > n_1 \ge 0$. Then the mle of $\Lambda = (I + \Sigma^{-1})^{-1}$ based on the marginal distribution of $\underset{\sim}{Y}$

is given by

$$\begin{aligned}\Lambda_m &= \frac{n_1}{n_1+1}I \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \le n_1+1 \\ &= \frac{n_1}{n_1+1}\Big(I + \frac{(1-\frac{n_1+1}{\underset{\sim}{y}'\underset{\sim}{y}})\underset{\sim}{y}\underset{\sim}{y}'}{n_1\underset{\sim}{y}'\underset{\sim}{y}}\Big) \qquad \text{if } n_1+1 \le \underset{\sim}{y}'\underset{\sim}{y} \le n_2+1 \\ &= \frac{n_1}{n_1+1}\Big(I + \frac{(n_2-n_1)\underset{\sim}{y}\underset{\sim}{y}'}{n_1(n_2+1)\underset{\sim}{y}'\underset{\sim}{y}}\Big) \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \ge n_2+1;\end{aligned} \tag{3.5}$$

the corresponding Bayes action (i.e., the type-II ML rule) is

$$\begin{aligned}\hat{\underset{\sim}{\theta}}_m = \Lambda_m \underset{\sim}{v} = \Lambda_1 \underset{\sim}{y} &= \frac{n_1}{n_1+1}\underset{\sim}{y} \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \le n_1+1 \\ &= (1-\frac{1}{\underset{\sim}{y}'\underset{\sim}{y}})\underset{\sim}{y} \qquad \text{if } n_1+1 \le \underset{\sim}{y}'\underset{\sim}{y} \le n_2+1 \\ = \Lambda_2 \underset{\sim}{y} &= \frac{n_2}{n_2+1}\underset{\sim}{y} \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \ge n_2+1.\end{aligned} \tag{3.6}$$

If the prior mean $\underset{\sim}{\mu}$ for $\underset{\sim}{\theta}$ is not zero, (3.6) is valid with $\underset{\sim}{y}$ replaced by $\underset{\sim}{y} - \underset{\sim}{\mu}$.

Remark. First note that the conditional minimax rule of Theorem 2.4 and the type-II ML rule described above differ fundamentally in their forms in that the minimax rule uses the action $\Lambda_2 \underset{\sim}{y}$ for $\underset{\sim}{y}$ near $\underset{\sim}{0}$ while the ML rule uses the action $\Lambda_2 \underset{\sim}{y}$ for large $\underset{\sim}{y}$ and the action $\Lambda_1 \underset{\sim}{y}$ for $\underset{\sim}{y}$ near $\underset{\sim}{0}$. To put it in another way, the ML rule shrinks more towards the prior mean when the data is near the prior mean. This may be actually desirable if we believe that data near the prior mean indicates a sharper tail of the prior. Otherwise, the extra shrinkage may be regarded as unnecessarily risky, in which case we would like to move up towards $\Lambda_2 \underset{\sim}{y}$. It actually turns out, as we shall see in the next section, that if we want to be near Bayes for the ML prior but still want to guard against priors with heavier tails, then the appropriate action is in fact to move up somewhat towards $\Lambda_2 \underset{\sim}{y}$ (for $\underset{\sim}{y}$ near $\underset{\sim}{0}$). Secondly, again note the clear connection of the ML rule to the Stein estimate. Of course, it is known that if the coordinates of $\underset{\sim}{\theta}$ are considered iid normal, then the ML rule is a Stein estimate. The new finding is that even if the coordinate means are not iid, the ML rule coincides with a Stein rule for moderate (or large, if $n_2 = \infty$) $\underset{\sim}{y}'\underset{\sim}{y}$. That the ML rule of (3.6) has good frequentist risk properties whenever sufficiently flat priors are included in Γ is an attractive feature of the ML rule. We have the following simple theorem.

Theorem 3.3. Let $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, I)$ and $\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, \Sigma)$, where $\Sigma \ge n_1 I$. Then the type-II ML rule is frequentist minimax for every $p \ge 3$.

Proof: The ML rule $\underset{\sim}{\hat{\theta}}_m$ can be written in the form $\underset{\sim}{\hat{\theta}}_m = \left(1 - \frac{r(\underset{\sim}{y}'\underset{\sim}{y})}{\underset{\sim}{y}'\underset{\sim}{y}}\right)\underset{\sim}{y}$ where

$$\begin{aligned} r(\underset{\sim}{y}'\underset{\sim}{y}) &= \frac{\underset{\sim}{y}'\underset{\sim}{y}}{n_1+1} \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} \le n_1+1 \\ &= 1 \qquad \text{if } \underset{\sim}{y}'\underset{\sim}{y} > n_1+1. \end{aligned} \tag{3.7}$$

It now follows from standard minimaxity results (see, for example, Strawderman (1971)) that the ML rule is minimax for $p \ge 3$. More will be said about the risk properties of the ML rule in section 5. Meanwhile, we now focus attention on the construction of restricted ML Bayes or restricted conditional minimax rules.

4. Applications of the Lehmann-Wald theory: restricted Bayes and restricted minimax rules. As mentioned in the preceding section, minimax rules often try to protect against priors that are deemed unlikely after observing the data. A more rational approach to selecting one Bayes action might be the restricted minimax principle where we minimize $\sup_{\pi\in\Gamma_0} r_{\underset{\sim}{y}}(\pi,\delta)$ where Γ_0 is a subclass of the likely priors. For example, each prior π in the original class Γ gives one value for the marginal likelihood $m_\pi(\underset{\sim}{y}) = \int \ell_{\underset{\sim}{y}}(\theta)d\pi(\underset{\sim}{\theta})$ of $\underset{\sim}{y}$ where $\ell_{\underset{\sim}{y}}(\underset{\sim}{\theta})$ is the usual likelihood function for $\underset{\sim}{\theta}$. A natural choice for Γ_0 may be

$$\Gamma_0 = \{\pi \in \Gamma\colon\ d(m_\pi, m_{\hat{\pi}}) \le \varepsilon\}, \tag{4.1}$$

for some metric d, where $\hat{\pi}$ is the type-II ML prior and $\varepsilon > 0$ is a suitable (usually small) fixed number. Now, even though derivation of the ML prior in our setup is relatively easy as long as Σ_1 and Σ_2 are multiples of the identity matrix, it seems rather hard to identify all priors π (or equivalently, all matrices Σ in the range $(k\lambda)^{-1}I \le \Sigma \le \lambda^{-1}I$) satisfying the inequality in (4.1) for any metric d. An alternative but related formulation, that is mathematically tractable, is to try to strike a balance between the minimax and the ML Bayes rule by minimizing $\sup_\pi r_{\underset{\sim}{y}}(\pi,\delta)$ subject to $r_{\underset{\sim}{y}}(\hat{\pi},\delta) \le (1+\varepsilon) r_{\underset{\sim}{y}}(\hat{\pi},\hat{\delta})$, where $\hat{\pi}$ is the ML prior and $\hat{\delta}$ the ML Bayes rule. The $(1+\varepsilon)$-constraint automatically forces good Bayesian behavior with respect to the ML prior, and the minimax component attempts to guard against other priors. This, of course, is the restricted minimax problem. See Bickel (1980) for some results in a frequentist setup.

Formally, an "equivalent" formulation of this restricted minimax problem is to minimize $r_{\underset{\sim}{y}}(\hat{\pi},\delta)$ subject to $\sup_\pi r_{\underset{\sim}{y}}(\pi,\delta) \le (1+\varepsilon)\overline{V}$, where $\overline{V}$ is the minimax risk. This is usually known as the restricted-Bayes problem. The restricted minimax and the restricted Bayes problems are both equivalent to minimizing $\lambda r_{\underset{\sim}{y}}(\hat{\pi}, \delta) + (1-\lambda) \sup_{\pi} r_{\underset{\sim}{y}}(\pi, \delta)$ for some $0 \leq \lambda \leq 1$ (see Hodges and Lehmann (1952)). However, depending on emphasis (i.e., whether ε-Bayes for small $\varepsilon > 0$ or ε-minimaxity for small $\varepsilon > 0$ is considered more important), it may be advantageous to derive the two rules separately. Part of the reason that this is more advantageous is that if one starts with an ε in either of the two problems, then identification of the corresponding λ is not immediate.

The mathematics involved in the derivation of the restricted-Bayes and the restricted minimax rules are similar. So we will derive only the restricted-Bayes rule. The results in Hodges and Lehmann (1952) are useful for this purpose. However, before we go into explicit derivation of the restricted-Bayes rule, we like to state and prove a trichotomy theorem that would be useful later. It shows that for certain δ, $\sup_{\Lambda} r(\Lambda, \delta)$ is achieved at one of three values of Λ. The prior mean $\underset{\sim}{\mu}$ is assumed to be zero, if it is not, the results are valid with $\underset{\sim}{y}$ replaced by $\underset{\sim}{y} - \underset{\sim}{\mu}$.

Theorem 4.1. Let $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, I)$, $\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, \Sigma)$, and let $n_1 I \leq \Sigma \leq n_2 I$, where n_1, n_2 are known. Let $\Lambda^n \underset{\sim}{v}$ denote the Bayes action corresponding to the prior $\Sigma = nI$, where $n_1 \leq n \leq n_2$. Here, $\Lambda^n = (I + \Sigma^{-1})^{-1} = \frac{n}{n+1} I$.

(a) If $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} \leq 1$, then for any n, $\sup_{\Lambda} r(\Lambda, \Lambda^n \underset{\sim}{v})$ (defined in (2.9)) is attained at $\Lambda = \Lambda_2$.

(b) If $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} > 1$, then $\sup_{\Lambda} r(\Lambda, \Lambda^n \underset{\sim}{v})$ is attained at $\Lambda = \Lambda_2$ whenever $\Lambda^n < \Lambda^*$ (Λ^* is defined in (2.7)), at $\Lambda = \Lambda_1^*$ whenever $\Lambda^n > \Lambda^*$ (Λ_1^* maximizes tr Λ among Λ satisfying $\Lambda \underset{\sim}{v} = \Lambda_1 \underset{\sim}{v}$), and at $\Lambda = \Lambda_2$ and Λ_1^* if $\Lambda^n = \Lambda^*$. Moreover, for every n, the trichotomy $\Lambda^n < \Lambda^*$ or $\Lambda^n > \Lambda^*$ or $\Lambda^n = \Lambda^*$ holds.

The variance-covariance matrix of $\underset{\sim}{Y}$ could be taken as a multiple of the identity matrix without ruining the above trichotomy result about the maximum risk. Before we give a proof of the theorem, we like to give a visual geometric description of the above result. Under the set up of the above theorem, the set of all Bayes actions $\Lambda \underset{\sim}{v}$ (where $\Lambda_1 \leq \Lambda \leq \Lambda_2$) form a p-dimensional sphere (see Theorem 2.1 in DasGupta and Studden (1988a)). For ease in visualizing, let us consider the case $p = 2$. Then the above sphere is just a circle; moreover, it is centered at $\overline{\Lambda} \underset{\sim}{v}$ and the center is obviously on the line through the origin joining $\Lambda_1 \underset{\sim}{v}$ and $\Lambda_2 \underset{\sim}{v}$, both of which are on the boundary of this circle. The conditional minimax action is $\Lambda^* \underset{\sim}{v}$ if $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} > 1$. This is also on the line joining $\Lambda_1 \underset{\sim}{v}$ and $\Lambda_2 \underset{\sim}{v}$. The trichotomy result says that for any action on this line (and inside the circle), the maximum risk is attained at $\Lambda = \Lambda_2$ if the action is closer to the origin than $\Lambda^* \underset{\sim}{v}$, at $\Lambda = \Lambda_1^*$ if the action is further from the origin than $\Lambda^* \underset{\sim}{v}$, and at Λ_2 and Λ_1^* if the action coincides with $\Lambda^* \underset{\sim}{v}$. The minimax action $\Lambda^* \underset{\sim}{v}$ thus acts as a 'splitting point' in this situation.

Proof of Theorem 4.1: Since Λ^n and Λ^* are both proportional to the identity matrix, it is clear that the trichotomy $\Lambda^n < \Lambda^*$ or $\Lambda^n > \Lambda^*$ or $\Lambda^n = \Lambda^*$ holds.

For simplicity of notation, we will give the proof in the case $n_1 = 1$; also notation will be simplified by denoting n_2 as m. First observe that for every $n (1 \leq n \leq m)$, $\Lambda^n \underset{\sim}{v}$ can be written as

$$\Lambda^n \underset{\sim}{v} = r \frac{\underset{\sim}{v}}{2} + (1-r) \frac{m}{m+1} \underset{\sim}{v}, \tag{4.2}$$

for some $0 \leq r \leq 1$. Also, routine algebra using the definition of Λ^* gives

$$\Lambda^* \underset{\sim}{v} = \left(1 - \frac{m+3}{4(m+1)} + \frac{1}{2\underset{\sim}{v}'\underset{\sim}{v}}\right) \underset{\sim}{v}, \tag{4.3}$$

and thus, $\Lambda^n = r \frac{I}{2} + (1-r) \frac{m}{m+1} I < \Lambda^*$

$$\Leftrightarrow r > \frac{1}{2} - \frac{m+1}{(m-1)\underset{\sim}{v}'\underset{\sim}{v}}. \tag{4.4}$$

We thus have to show that if (4.4) holds, then

$$\begin{aligned} &\|r \frac{\underset{\sim}{v}}{2} + (1-r) \frac{m}{m+1} \underset{\sim}{v} - \Lambda \underset{\sim}{v}\|^2 + \operatorname{tr}\Lambda \\ &\leq \|r \frac{\underset{\sim}{v}}{2} + (1-r) \frac{m}{m+1} \underset{\sim}{v} - \Lambda_2 \underset{\sim}{v}\|^2 + \operatorname{tr}\Lambda_2, \end{aligned} \tag{4.5}$$

for $\Lambda_1 \leq \Lambda \leq \Lambda_2$.

Since $\Lambda_1 = \frac{1}{2} I$ and $\Lambda_2 = \frac{m}{m+1} I$, using the representation $\Lambda = \overline{\Lambda} + C$, where $-\frac{m-1}{4(m+1)} I \leq C \leq \frac{m-1}{4(m+1)} I$, one can rewrite Λ in the more convenient form

$$\Lambda = \frac{m}{m+1} I - \frac{m-1}{4(m+1)} B, \tag{4.6}$$

where $0 \le B \le 2I$. (4.5) then reduces to showing

$$\frac{(m-1)\underset{\sim}{v}'B^2\underset{\sim}{v}}{4(m+1)} - \frac{r(m-1)}{m+1}\underset{\sim}{v}'B\underset{\sim}{v} - \text{ tr}B \le 0. \tag{4.7}$$

Using the facts that $\underset{\sim}{v}'B^2\underset{\sim}{v} \le 2\underset{\sim}{v}'B\underset{\sim}{v}$, $\frac{\underset{\sim}{v}'B\underset{\sim}{v}}{\underset{\sim}{v}'\underset{\sim}{v}} \le$ tr$B$ and (4.4), (4.7) follows. This proves that as long as we take an action closer to the origin than the action $\Lambda^*\underset{\sim}{v}$, then irrespective of the value of $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v}$, the maximum risk of that action is always attained at $\Lambda = \Lambda_2$. The proof for the cases $r < \frac{1}{2} - \frac{m+1}{(m-1)\underset{\sim}{v}'\underset{\sim}{v}}$, $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} \le 1$ and $r < \frac{1}{2} - \frac{m+1}{(m-1)\underset{\sim}{v}'\underset{\sim}{v}}$, $\underset{\sim}{v}'(\Lambda_2 - \Lambda_1)\underset{\sim}{v} > 1$ are very similar and are omitted. The case $r = \frac{1}{2} - \frac{m+1}{(m-1)\underset{\sim}{v}'\underset{\sim}{v}}$ (i.e., $\Lambda^n = \Lambda^*$) has already been treated in Theorem 2.4.

We are now in a position to describe the restricted-ML Bayes rule, i.e., the rule that minimizes the risk for the type-II ML prior subject to being ε-minimax over all priors in Γ. We will do this only for the case when Σ_0, Σ_1 Σ_2 are multiples of the identity matrix. This is because the trichotomy result of Theorem 4.1 is crucial for this purpose and we do not have an analog of Theorem 4.1 in more general cases. The following theorem is stated for the case $\Sigma_0 = I$, and $\underset{\sim}{\mu} = \underset{\sim}{0}$.

Theorem 4.2. Let $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, I)$, $\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, \Sigma)$, where $n_1 I \le \Sigma \le n_2 I$ for known constants n_1, n_2. The action that minimizes $||\delta - \Lambda_m\underset{\sim}{v}||^2 +$ trΛ_m subject to $\sup_{\Lambda}\{||\delta - \Lambda\underset{\sim}{v}||^2 + \text{ tr}\Lambda\} \le (1+\varepsilon)\overline{V}$, where $\overline{V}$ is the minimax risk and Λ_m is the type-II mle of Λ (and is defined in (3.5)), and $\varepsilon > 0$ is a fixed real number, is either the action $\Lambda_m\underset{\sim}{v}$ itself or is the action on the line segment joining $\Lambda_1\underset{\sim}{v}$ and $\Lambda_2\underset{\sim}{v}$ that is closest to $\Lambda_m\underset{\sim}{v}$ in L_2 norm among all actions on this line that satisfy the $(1+\varepsilon)$-constraint.

Proof: We will give a descriptive proof of this theorem because of ease in understanding. First note that if $\varepsilon > 0$ is sufficiently large, then the type-II ML action would itself satisfy the $(1+\varepsilon)$-constraint and would therefore be the restricted Bayes solution too. Now consider the case when $\Lambda_m\underset{\sim}{v}$ does not satisfy the $(1+\varepsilon)$-constraint. Recall that the ML Bayes action $\Lambda_m\underset{\sim}{v}$ is on the same line through the origin as the minimax action (it is helpful to consider the case $p = 2$ and think geometrically at this stage). Now, if $\Lambda_m\underset{\sim}{v}$ is closer to the origin than the minimax action, then we can move along the line joining $\Lambda_1\underset{\sim}{v}$ and $\Lambda_2\underset{\sim}{v}$, starting at the minimax action and move towards the ML Bayes action, and stop as soon as equality is attained in the $(1+\varepsilon)$-constraint. In fact, because the trichotomy theorem implies that all actions on the line which are closer to the origin than the minimax action have their maximum risk attained at $\Lambda = \Lambda_2$, we immediately have that movement should be stopped as soon as we reach the action $r\Lambda_1 \underset{\sim}{v} + (1-r)\Lambda_2 \underset{\sim}{v} = t\underset{\sim}{v}$ (say) such that

$$||t\underset{\sim}{v} - \Lambda_2 \underset{\sim}{v}||^2 + \text{ tr}\Lambda_2 = (1+\varepsilon)\overline{V}. \tag{4.8}$$

This action $t\underset{\sim}{v}$, by construction, is the action on the line closest to the ML action $\Lambda_m \underset{\sim}{v}$ subject to the $(1+\varepsilon)$-constraint being satisfied. The reason it is the restricted Bayes action is that it can be written as a convex combination of $\Lambda_m \underset{\sim}{v}$ and $\Lambda_2 \underset{\sim}{v}$, and is thus Bayes with respect to a prior of the form (or equivalently, a Λ of the form)

$$\Lambda = (1-\lambda)\Lambda_m + \lambda\Lambda_2$$

for a suitable $\lambda(0 < \lambda \leq 1)$. Since the trichotomy theorem again implies that the maximum risk of $t\underset{\sim}{v}$ is attained at Λ_2, standard theorems imply that $t\underset{\sim}{v}$ is the restricted-Bayes action (see Hodges and Lehmann (1952)).

The argument for the case when $\Lambda_m \underset{\sim}{v}$ is farther from the origin than the minimax action is practically the same, except now we have to move away from the origin, starting at the minimax action, and stop as soon as equality is attained in the $(1+\varepsilon)$-constraint. The other difference is that the maximum risk of the restricted-Bayes action is now attained at Λ_1^* instead of Λ_2. However, the restricted-Bayes action still is a convex combination of $\Lambda_1^* \underset{\sim}{v}$ and $\Lambda_m \underset{\sim}{v}$ (recall $\Lambda_1^* \underset{\sim}{v} = \Lambda_1 \underset{\sim}{v}$) and the Hodges-Lehmann theorem again applies.

Theorem 4.2 gives a very convenient recipe for identifying the restricted-Bayes action. We now work out an example giving an exact expression for the restricted-Bayes action in a special case.

Example 2. Let $\Sigma_0 = I$, $\underset{\sim}{\mu} = \underset{\sim}{0}$, $n_1 = 1$, $n_2 = \infty$ (i.e., $\Sigma \geq I$; this should be considered as an assumption that prior information is at least as uncertain as sample information). Then the ML Bayes and the minimax actions are respectively of the form (here $\underset{\sim}{v}$ is actually the same as $\underset{\sim}{y}$)

$$\begin{aligned} \Lambda_m \underset{\sim}{v} &= \frac{\underset{\sim}{v}}{2} \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} \leq 2 \\ &= \left(1 - \frac{1}{\underset{\sim}{v}'\underset{\sim}{v}}\right)\underset{\sim}{v} \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} > 2, \end{aligned} \tag{4.9}$$

and

$$\begin{aligned}\hat{\theta} &= v \qquad \text{if } v'v \le 2 \\ &= (\frac{3}{4} + \frac{1}{2v'v})v \qquad \text{if } v'v > 2.\end{aligned} \tag{4.10}$$

Let also $p = 5$ and $\varepsilon = .1$. Then for $v'v \le 2$, $\Lambda_m v = \frac{v}{2}$ satisfies the $(1+\varepsilon)$-constraint

$$\begin{aligned}&\sup_{\Lambda}\{||\Lambda_m v - \Lambda v||^2 + \text{ tr}\Lambda\} \le (1+\varepsilon)\overline{V} \\ \Leftrightarrow &||\frac{v}{2} - v||^2 \le \frac{p}{10}.\end{aligned} \tag{4.11}$$

Therefore, for $v'v \le 2$, the restricted Bayes action coincides with $\frac{v}{2} = \frac{y}{2}$. Next, for $2 < v'v \le 6$, the ML Bayes action $\Lambda_m v = (1 - \frac{1}{v'v})v$ is closer to the origin than the minimax action $\hat{\theta} = (\frac{3}{4} + \frac{1}{2v'v})v$. Thus for $2 < v'v \le 6$, $\Lambda_m v$ has its maximum risk at $\Lambda = \Lambda_2 = I$. Using this, it is routine to show that for $2 < v'v \le 6$, (4.11) is again satisfied and thus the restricted Bayes action again coincides with $\Lambda_m v = (1 - \frac{1}{v'v})v$. Notice in this case,

$$\overline{V} = (\frac{1}{2v'v} - \frac{1}{4})^2 + p. \tag{4.12}$$

For $v'v > 6$, $\Lambda_m v$ is farther from the origin than $\hat{\theta}$. Thus for $v'v > 6$, $\Lambda_m v$ has its maximum risk at $\Lambda_1^* = I - \frac{vv'}{2v'v}$. Using this, it follows that for $6 < v'v \le 9.0765$ (approximately), $\Lambda_m v$ satisfies (4.11). For $v'v > 9.0765$, (4.11) does not hold and therefore the restricted Bayes action is such that equality holds in (4.11). It follows that the action equals

$$(\frac{1}{2} + \sqrt{.06875(\frac{2}{v'v} - 1)^2 + \frac{1}{v'v}}) \cdot v. \tag{4.13}$$

We therefore have that for $\Sigma_0 = I$, $\Sigma \ge \Sigma_0$, $\mu = 0$, $\varepsilon = .1$, and $p = 5$, the restricted Bayes action is given by

$$\begin{aligned}\hat{\theta}_R &= \frac{v}{2} \qquad \text{if } v'v \le 2 \\ &= (1 - \frac{1}{v'v})v \qquad \text{if } 2 < v'v \le 9.0765 \\ &= (\frac{1}{2} + \sqrt{.06875(\frac{2}{v'v} - 1)^2 + \frac{1}{v'v}}) \cdot v \qquad \text{if } v'v > 9.0765.\end{aligned} \tag{4.14}$$

Note that as $v'v \to \infty$, $\hat{\theta}_R$ converges to the regular Bayes estimate with respect to the $N(0, \frac{2+\sqrt{1+\varepsilon}}{2-\sqrt{1+\varepsilon}}I)$. Therefore, $\hat{\theta}_R$ cannot be frequentist minimax. However, it can obviously be again written in the usual form

$$\hat{\theta}_R = (1 - \frac{r_R(v'v)}{v'v})v.$$

In fact,

$$r_R(s) = \frac{s}{2} \qquad \text{if } s \le 2$$
$$= 1 \qquad \text{if } 2 < s \le 9.0765 \tag{4.15}$$
$$= s\left(\frac{1}{2} - \sqrt{.06875\left(\frac{2}{s} - 1\right)^2 + \frac{1}{s}}\right) \qquad \text{if } s > 9.0765.$$

$r_R(s)$ is easily seen to be nondecreasing in s; however, it is clearly not bounded above by $2(p-2) = 6$ (in fact, $r_R(s)$ is unbounded). However, it in fact is bounded above by 6 for $s \le 30.64$. This indicates that while $\hat{\underset{\sim}{\theta}}_R$ cannot be frequentist minimax, it may have good risk properties for even moderately large values of $||\underset{\sim}{\theta}||$. More will be said about this in section 5.

Before closing this example, we will like to work out how much improvement in risk at Λ_m the restricted Bayes action provides (in comparison to the minimax action $\hat{\theta}$) by sacrificing $100\varepsilon\%$ in maximum risk (again in comparison to the minimax action). Towards this end, we will consider the quantity

$$I(\underset{\sim}{v}'\underset{\sim}{v}) = \frac{r(\Lambda_m, \hat{\underset{\sim}{\theta}}) - r(\Lambda_m, \hat{\underset{\sim}{\theta}}_R)}{r(\Lambda_m, \hat{\underset{\sim}{\theta}}) - r(\Lambda_m, \Lambda_m \underset{\sim}{v})}; \tag{4.16}$$

thus $I(\underset{\sim}{v}'\underset{\sim}{v})$ represents the relative improvement in deficiency by using $\hat{\underset{\sim}{\theta}}_R$ instead of $\hat{\underset{\sim}{\theta}}$ as an alternative to $\Lambda_m \underset{\sim}{v}$, the best action for $\Lambda = \Lambda_m$. $I(\underset{\sim}{v}'\underset{\sim}{v})$ is conceptually related to the Relative Savings Loss (RSL), a quantity proposed in Efron and Morris (1971). However, unlike the RSL, large values of $I(\underset{\sim}{v}'\underset{\sim}{v})$ would be preferred.

Straightforward computation using the definitions of $\Lambda_m \underset{\sim}{v}$, $\hat{\underset{\sim}{\theta}}$, and $\hat{\underset{\sim}{\theta}}_R$ (see (4.9), (4.10), and (4.14)) gives

$$I(\underset{\sim}{v}'\underset{\sim}{v}) = 1 \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} \le 9.0765$$
$$= 1 - \frac{\left(\frac{1}{\underset{\sim}{v}'\underset{\sim}{v}} + \sqrt{.06875\left(\frac{2}{\underset{\sim}{v}'\underset{\sim}{v}} - 1\right)^2 + \frac{1}{\underset{\sim}{v}'\underset{\sim}{v}}} - \frac{1}{2}\right)^2}{\left(\frac{3}{2\underset{\sim}{v}'\underset{\sim}{v}} - \frac{1}{4}\right)^2} \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} > 9.0765. \tag{4.17}$$

Thus, for $\underset{\sim}{v}'\underset{\sim}{v} \le 9.0765$, there is a 100% improvement in deficiency in exchange for a 10% increase in maximum risk. The improvement in deficiency decreases monotonically as $\underset{\sim}{v}'\underset{\sim}{v}$ increases and converges to .0952 (i.e., 9.52%) as $\underset{\sim}{v}'\underset{\sim}{v} \to \infty$. Thus the asymptotic improvement is slightly smaller than the increase in maximum risk. But, the improvement is significantly larger than 10% for quite large values of $\underset{\sim}{v}'\underset{\sim}{v}$. A plot of $I(\underset{\sim}{v}'\underset{\sim}{v})$ is given in Figure 5. Here are some values of $I(\underset{\sim}{v}'\underset{\sim}{v})$ for a few selected values of $\underset{\sim}{v}'\underset{\sim}{v}$:

Table 1: Relative improvement in risk at Λ_m

$\underset{\sim}{v}'\underset{\sim}{v}$	9.5	12.05	16.75	21	30	49
$I(\underset{\sim}{v}'\underset{\sim}{v})$	.99	.80	.57	.47	.35	.25.

We now turn to derivation of the restricted minimax rule. A theorem analogous to Theorem 4.2 is again valid; this and the trichotomy theorem again aid in writing down exact expressions for the restricted minimax rule. The analysis is very similar to that for the restricted Bayes rule. Therefore, all details will be omitted.

Theorem 4.3. Let $\underset{\sim}{Y} \sim N(\underset{\sim}{\theta}, I)$, $\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, \Sigma)$, where $n_1 I \leq \Sigma \leq n_2 I$ for known constants n_1, n_2. The action that minimizes $\sup_{\Lambda}\{||\delta - \Lambda\underset{\sim}{v}||^2 + \text{ tr}\Lambda\}$ subject to $||\delta - \Lambda_m\underset{\sim}{v}||^2 + \text{ tr}\Lambda_m \leq (1+\varepsilon)\text{ tr}\Lambda_m$ is either the unrestricted minimax action $\hat{\underset{\sim}{\theta}}$ itself or is the action on the line segment joining $\Lambda_1\underset{\sim}{v}$ and $\Lambda_2\underset{\sim}{v}$ that is closest to the minimax action among all actions on this line that satisfy the $(1+\varepsilon)$-constraint.

Proof: It is easily seen that the restricted minimax action must lie on the line joining $\Lambda_1\underset{\sim}{v}$ and $\Lambda_2\underset{\sim}{v}$. Once this simplification is reached, the rest of the proof is virtually a repetition of the argument given in Theorem 4.2. The details are omitted.

Example 3. Again we will give an exact expression for the restricted minimax rule in a specific example. The set up considered is $\Sigma_0 = I$, $\underset{\sim}{\mu} = \underset{\sim}{0}$, $n_1 = 1$, $n_2 = \infty$, $p = 10$, and $\varepsilon = .1$ (the reason for taking a larger p in this example is to give an idea of what happens in relatively high dimensional problems). We are thus sacrificing 10% in terms of risk at the ML prior. Direct calculations similar to those in Example 2 yield that the restricted minimax action $\hat{\underset{\sim}{\theta}}_M$ is given as

$$\begin{aligned} \hat{\underset{\sim}{\theta}}_M &= \underset{\sim}{v} \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} \leq 2 \\ &= \left(\frac{3}{4} + \frac{1}{2\underset{\sim}{v}'\underset{\sim}{v}}\right)\underset{\sim}{v} \qquad \text{if } 2 < \underset{\sim}{v}'\underset{\sim}{v} \leq 18.8 \\ &= \left(1 - \frac{1}{\underset{\sim}{v}'\underset{\sim}{v}} - \sqrt{\frac{\cdot 1}{\underset{\sim}{v}'\underset{\sim}{v}}\left(5.5 - \frac{1}{\underset{\sim}{v}'\underset{\sim}{v}}\right)}\right) \cdot \underset{\sim}{v} \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} > 18.8. \end{aligned} \tag{4.18}$$

Thus the restricted minimax action coincides with the minimax action for small and moderate values of $\underset{\sim}{v}'\underset{\sim}{v}$ and converges to the ML-Bayes estimate as $\underset{\sim}{v}'\underset{\sim}{v} \to \infty$. As usual, $\hat{\underset{\sim}{\theta}}_M$

can be written in the form

$$\hat{\theta}_M = (1 - \frac{r_M(\underset{\sim}{v}'\underset{\sim}{v})}{\underset{\sim}{v}'\underset{\sim}{v}}) \cdot \underset{\sim}{v}$$

with

$$\begin{aligned} r_M(\underset{\sim}{v}'\underset{\sim}{v}) &= 0 \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} \le 2 \\ &= \frac{\underset{\sim}{v}'\underset{\sim}{v}}{4} - \frac{1}{2} \qquad \text{if } 2 < \underset{\sim}{v}'\underset{\sim}{v} \le 18.8 \\ &= 1 + \sqrt{\cdot 1(5.5\underset{\sim}{v}'\underset{\sim}{v} - 1)} \qquad \text{if } \underset{\sim}{v}'\underset{\sim}{v} > 18.8. \end{aligned} \tag{4.19}$$

Again $r_M(\underset{\sim}{v}'\underset{\sim}{v})$ is nondecreasing in $\underset{\sim}{v}'\underset{\sim}{v}$, but is not bounded above by $2(p-2) = 16$. However, $r_M(\cdot)$ is bounded above by 16 for $\underset{\sim}{v}'\underset{\sim}{v} \le 409.27$ (approximately), indicating that for quite large values of $||\underset{\sim}{\theta}||$, $\hat{\underset{\sim}{\theta}}_M$ may have good risk properties. We postpone such risk assessments till section 5.

We close this section with an evaluation of gains in maximum risk provided by $\hat{\underset{\sim}{\theta}}_M$ in comparison to the ML-Bayes rule $\Lambda_m \underset{\sim}{v}$ in exchange for a 10% increase in Bayes risk at Λ_m. In analogy with (4.16), we define

$$J(\underset{\sim}{v}'\underset{\sim}{v}) = \frac{\sup_{\Lambda} r(\Lambda, \Lambda_m \underset{\sim}{v}) - \sup_{\Lambda} r(\Lambda, \hat{\underset{\sim}{\theta}}_M)}{\sup_{\Lambda} r(\Lambda, \Lambda_m \underset{\sim}{v}) - \sup_{\Lambda} r(\Lambda, \hat{\underset{\sim}{\theta}})}. \tag{4.20}$$

Just like $I(\underset{\sim}{v}'\underset{\sim}{v})$, $J(\underset{\sim}{v}'\underset{\sim}{v})$ measures the relative improvement in deficiency by using $\hat{\underset{\sim}{\theta}}_M$ instead of $\Lambda_m \underset{\sim}{v}$ as an alternative to $\hat{\underset{\sim}{\theta}}$.

Since $\hat{\underset{\sim}{\theta}}_M = \hat{\underset{\sim}{\theta}}$ for $\underset{\sim}{v}'\underset{\sim}{v} \le 18.8$, one has that $J(\underset{\sim}{v}'\underset{\sim}{v}) = 1$ for $\underset{\sim}{v}'\underset{\sim}{v} \le 18.8$. Also, for $\underset{\sim}{v}'\underset{\sim}{v} > 18.8$, both $\Lambda_m \underset{\sim}{v}$ and $\hat{\underset{\sim}{\theta}}_M$ are farther from the origin than $\hat{\theta}$; therefore, by Theorem 4.1, $\Lambda_m \underset{\sim}{v}$, $\hat{\underset{\sim}{\theta}}_M$, and also $\hat{\underset{\sim}{\theta}}$ all have their maximum risks attained at $\Lambda = \Lambda_1^* = I - \frac{\underset{\sim}{v}\underset{\sim}{v}'}{2\underset{\sim}{v}'\underset{\sim}{v}}$. Then direct computation gives for $\underset{\sim}{v}'\underset{\sim}{v} > 18.8$,

$$J(\underset{\sim}{v}'\underset{\sim}{v}) = \frac{(\frac{1}{2} - \frac{1}{v'v})^2 - (\frac{1}{2} - \frac{1}{v'v} - \sqrt{\frac{\cdot 1}{v'v}(5.5 - \frac{1}{v'v})})^2}{(\frac{1}{2} - \frac{1}{\underset{\sim}{v}'\underset{\sim}{v}})^2 - (\frac{1}{4} + \frac{1}{2\underset{\sim}{v}'\underset{\sim}{v}})^2} \tag{4.21}$$

Thus, for $\underset{\sim}{v}'\underset{\sim}{v} < 18.8$ there is 100% improvement in deficiency in maximum risk for a 10% increase in Bayes risk; the improvement again decreases monotonically as $\underset{\sim}{v}'\underset{\sim}{v}$ increases and converges to 0 as $\underset{\sim}{v}'\underset{\sim}{v} \to \infty$. However, the convergence to zero seems to be rather slow; for example, even at $\underset{\sim}{v}'\underset{\sim}{v} = 100$, $J(\underset{\sim}{v}'\underset{\sim}{v})$ is more than .38! A few selected values of $J(\underset{\sim}{v}'\underset{\sim}{v})$ are given below. The function is plotted in Figure 6.

Table 2: Relative improvement in maximum risk

$v'v$	18.8	21	28	33.8	38	42	48.5	150	400
$J(v'v)$	1	.92	.77	.68	.64	.60	.56	.31	.19

The numbers indicate that in high dimensional problems a lot can be gained in terms of the maximum posterior risk in exchange for a small sacrifice in posterior risk for the ML prior. Similar results were found (though *not for posterior risks*) in Berger (1982), DasGupta and Rubin (1987a), DasGupta and Bose (1987b) etc.

5. Frequentist behavior and risk functions. As mentioned before, we believe that even if a procedure is constructed on the basis of conditional considerations, it must be judged in regard to its performance over repeated use. In other words, before a procedure can be seriously recommended for actual use, we must look into its risk behavior. Under an ordinary squared error loss, the risk of a procedure is just its mean squared error (MSE). In Figures 1 through 4, we show the MSE of 5 procedures; the following notation has been used: $\hat{\theta}_1 = \hat{\theta}$ (see (4.10)), $\hat{\theta}_2 =$ the positive-part Stein estimator, $\hat{\theta}_3 = \Lambda_m v$ (see (4.9)), $\hat{\theta}_4 = \hat{\theta}_R$ (see (4.14); this is considered only for $p = 5$), and $\hat{\theta}_5 = \hat{\theta}_M$ (see (4.18)); again recall that this corresponds to $p = 10$). Of these, $\hat{\theta}_2$ and $\hat{\theta}_3$ are known to be frequentist minimax (see Theorem (3.3)).

All four figures show that the risk function of $\hat{\theta}_1$ increases at a rather high rate, but has satisfactory risks for small or even moderate values of $||\theta||$ ($||\theta||$ has been labeled R in the figures). For example, if $p = 10$ and one takes a $N(0, 5I)$ prior for θ, then $\theta'\theta$ is smaller than 46 with a probability of approximately .9. At $||\theta|| = \sqrt{46} = 6.78$, $\hat{\theta}_1$ has a risk of about 8, 20% less than that of the usual estimate Y. The figures also show that quite soon the risks of the positive-part Stein estimator ($\hat{\theta}_2$) and the type-II ML Bayes rule ($\hat{\theta}_3$) become virtually identical. The risk behavior of $\hat{\theta}_5$ (see Figure 4) is especially encouraging. Up to $||\theta|| = 20$, it has a risk smaller than that of Y and for $20 \geq ||\theta|| \geq 7$, has a risk function virtually identical to that of the positive-part Stein rule. Actually, it is possible that the restricted conditional minimax estimators for small ε are frequentist minimax, but we have not checked the risks for other values of p and ε ($\hat{\theta}_5$ corresponds to $p = 10$ and $\varepsilon = \cdot 1$). $\hat{\theta}_4$ has similar risks than $\hat{\theta}_1$, which may be anticipated because $\hat{\theta}_4$ was constructed in a way that forces it to be close to $\hat{\theta}_1$.

Finally, plots 5 and 6 show the improvement in deficiency (as discussed in section 4; see (4.17) and (4.21)) by using the estimators $\hat{\theta}_4$ and $\hat{\theta}_5$. Both plots indicate that a substantial amount can be gained in terms of one criterion by giving up a relatively small amount in terms of the other criterion. Figure 6 is especially encouraging.

Summarizing, $\hat{\theta}_3$ and $\hat{\theta}_5$ have very good risk properties; $\hat{\theta}_5$ has very good conditional properties too (it is nearly ML-Bayes and also significantly improves the deficiency of the ML-Bayes rule in terms of maximum posterior risk). The indication, we believe, is that a conditional Γ_0-minimax procedure where Γ_0 contains the sub-collection of "likely" priors in the original family Γ is likely to have good conditional as well as frequentist properties. The conditional Γ-minimax procedure as well as the restricted ML-Bayes procedure can be risky if somehow large $\underset{\sim}{\theta}$ are considered plausible.

6. Concluding remarks. In this article we have presented several *concrete* ways to select one Bayesian procedure for actual use in a conditional framework when we have a family of priors instead of a single prior. In particular, the heuristic concept of picking up the "likely" priors has been given a *concrete and workable* shape by consideration of 'neighborhoods' of the type-II ML prior. Considerable evidence has been presented that the Γ-minimax method may indeed work well if the unlikely priors are somehow eliminated. Undoubtedly, there are other ways of selecting the likely priors. We consider this article as a concrete first practical step towards the selection of a Bayesian rule. Selecting among the frequentist minimax rules in a normal location problem was considered in Berger (1982). The selection problems arise very naturally in the Bayes as well as the minimax set up because of a relatively large choice the practitioner is faced with. Our results stem from this practical issue.

Acknowledgment

We are grateful to Mr. Fu-Chuen Chang for help in the computing work of section 5. Thanks are also due to Jim Berger and Leon Gleser for helpful conversation on selection issues.

References

Berger, James (1982). Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. *Statistical Decision Theory and Related Topics III* **1**, 109–141, Academic Press, New York.

Berger, James and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14**, 461–486.

Berger, James and Sivaganesan, S. (1986). Ranges of posterior measures for priors with unimodal contaminations. Tech. Report #86-41, Dept. of Statistics, Purdue University.

Berger, James (1987). Robust Bayesian analysis: sensitivity to the prior. Tech. report #87-10, Dept. of Statistics, Purdue Univrsity.

Bickel, P. J. (1980). Minimax estimation of the mean of a normal distribution subject to doing well at a point. Technical Report, University of California, Berkeley.

DasGupta, Anirban and Berger, James (1986c). Estimation of multiple gamma scale-parameters: Bayes estimation subject to uniform domination. *Comm. in Statist., Theory and Methods*, Special issue on Stein-type multivariate estimation **15**, 2065–2086.

DasGupta, Anirban and Rubin, H. (1987a). Bayesian estimation subject to minimaxity of the mean of a multivariate normal distribution in the case of a common unknown variance. *Statistical Decision Theory and Related Topics IV*, S. S. Gupta and J. O. Berger eds., Volume 1, 325–345.

DasGupta, Anirban and Bose, A. (1987b). Γ-minimax and restricted risk Bayes estimation of multiple poisson means under ε-contaminations of the subjective prior. To appear in *Statistics and Decisions.*

DasGupta, Anirban and Studden, W. J. (1988a). Robust Bayesian analysis and optimal experimental designs in normal linear models with many parameters–I. Technical Report, Dept. of Statistics, Purdue University.

DasGupta, Anirban and Studden, W. J. (1988b). Robust Bayesian analysis and optimal experimental designs in normal linear models with many parameters–II. Technical Report, Dept. of Statistics, Purdue University.

DasGupta, Anirban and Studden, W. J. (1988c). Frequentist behavior of smallest volume robust Bayes confidence sets. Tech. report, Dept. of Statistics, Purdue University.

DeRobertis, L. and Hartigan, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9**, 235–244.

Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators–Part I: The Bayes case. *Jour. Amer. Statist. Assoc.* **66**, 807–815.

Good, I. J. (1965). *The Estimation of Probabilities.* M.I.T. Press, Cambridge.

Good, I. J. and Crook, J. F. (1987). The robustness and sensitivity of the mixed-Dirichlet Bayesian test for "independence" in contingency tables. *Ann. Statist.* **15**, 670–693.

Hodges, J. L. and Lehmann, E. L. (1952). The use of previous experience in reaching statistical decisions. *Ann. Math. Statist.* **23**, 392–407.

Kadane, J. and Chuang, D. T. (1978). Stable decision problems. *Ann. Statist.* **6**, 1095–1110.

Kempthorne, P. J. (1986). Dominating inadmissible procedures using compromise decision theory. *Stat. Dec. Theory and Related Topics IV*, S. S. Gupta and J. O. Berger eds., Vol 1.

Leamer, E. E. (1978). Specification searches: ad hoc inference with nonexperimental data. John Wiley, New York.

Leamer, E. E. (1982). Sets of posterior means with bounded variance prior. *Econometrica* **50**, 725–736.

Maritz, J. S. (1970). *Empirical Bayes Methods.* Methuen, London.

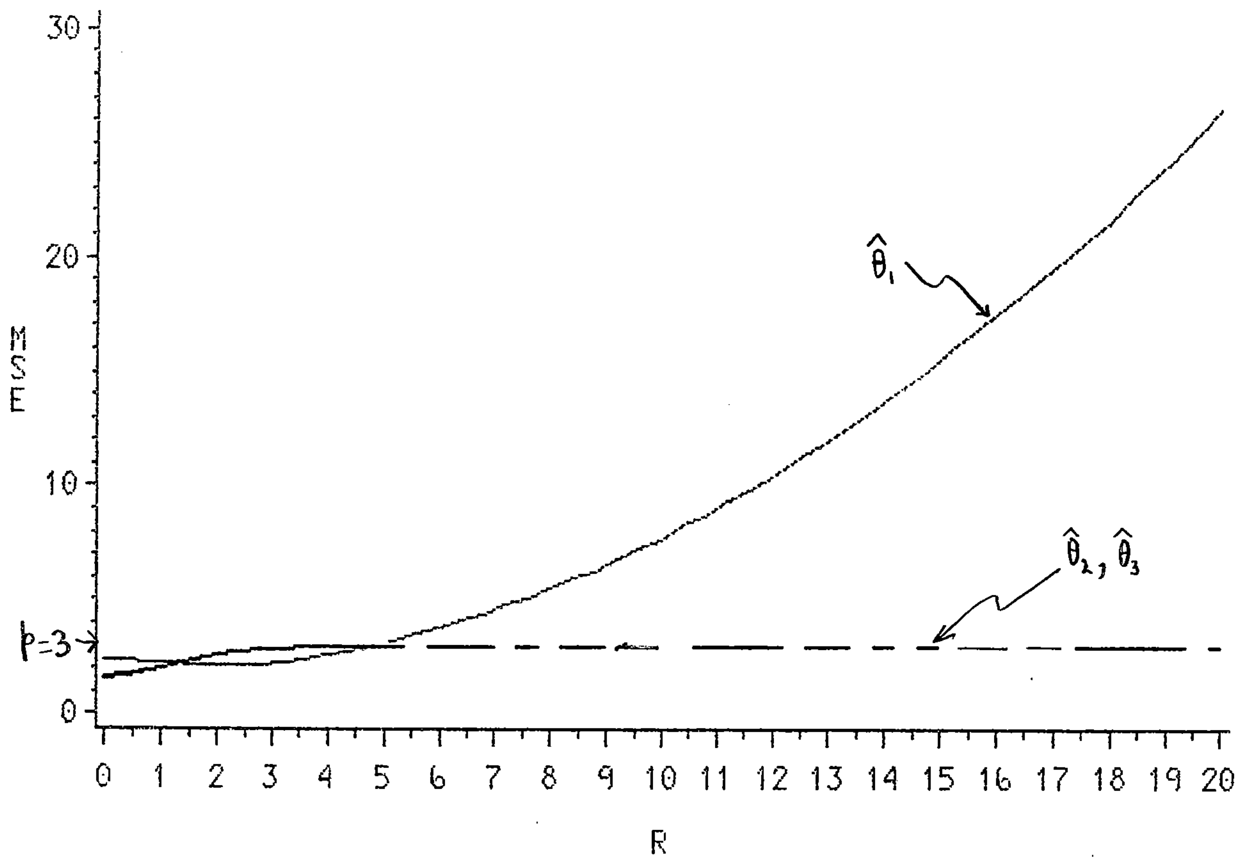
Morris, C. (1983). Parametric empirical Bayes inference: theory and applications (with

Discussion). *J. Amer. Statist. Assoc.* **78** 47–65.

Polasek, W. (1985). Sensitivity analysis for general and hierarchical linear regression models. Bayesian inference and Decision techniques with Applications, Eds. P. K. Goel and A. Zellner, North Holland, Amsterdam.

Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42**, 385–388.

FIG. 1: P=3 ESTIMATOR(1,2,3)

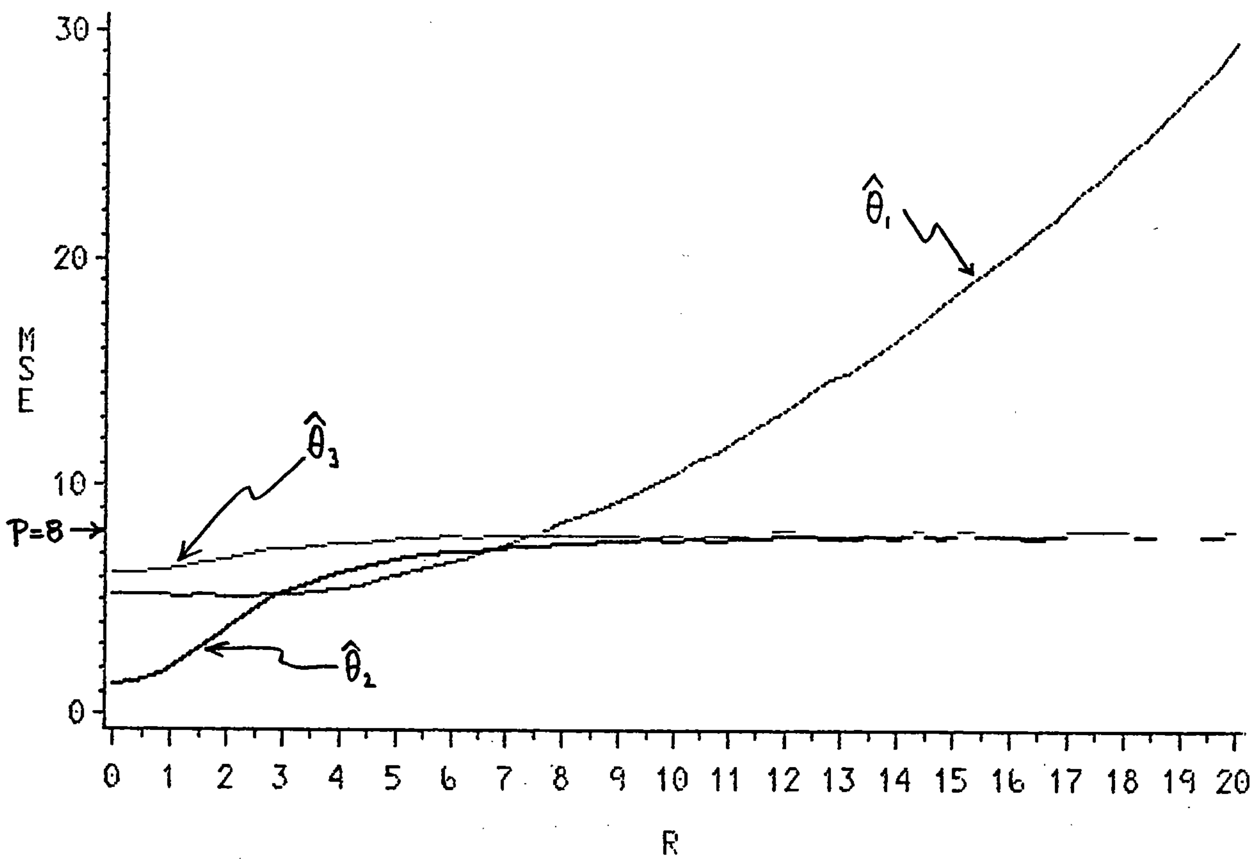


$\hat{\theta}_1$ = The Γ-minimax rule (defined in (4.10))

$\hat{\theta}_2$ = The positive part James-Stein rule

$\hat{\theta}_3$ = The type-2 ML Bayes rule (defined in (4.9))

FIG.2: P=8 ESTIMATOR(1,2,3)

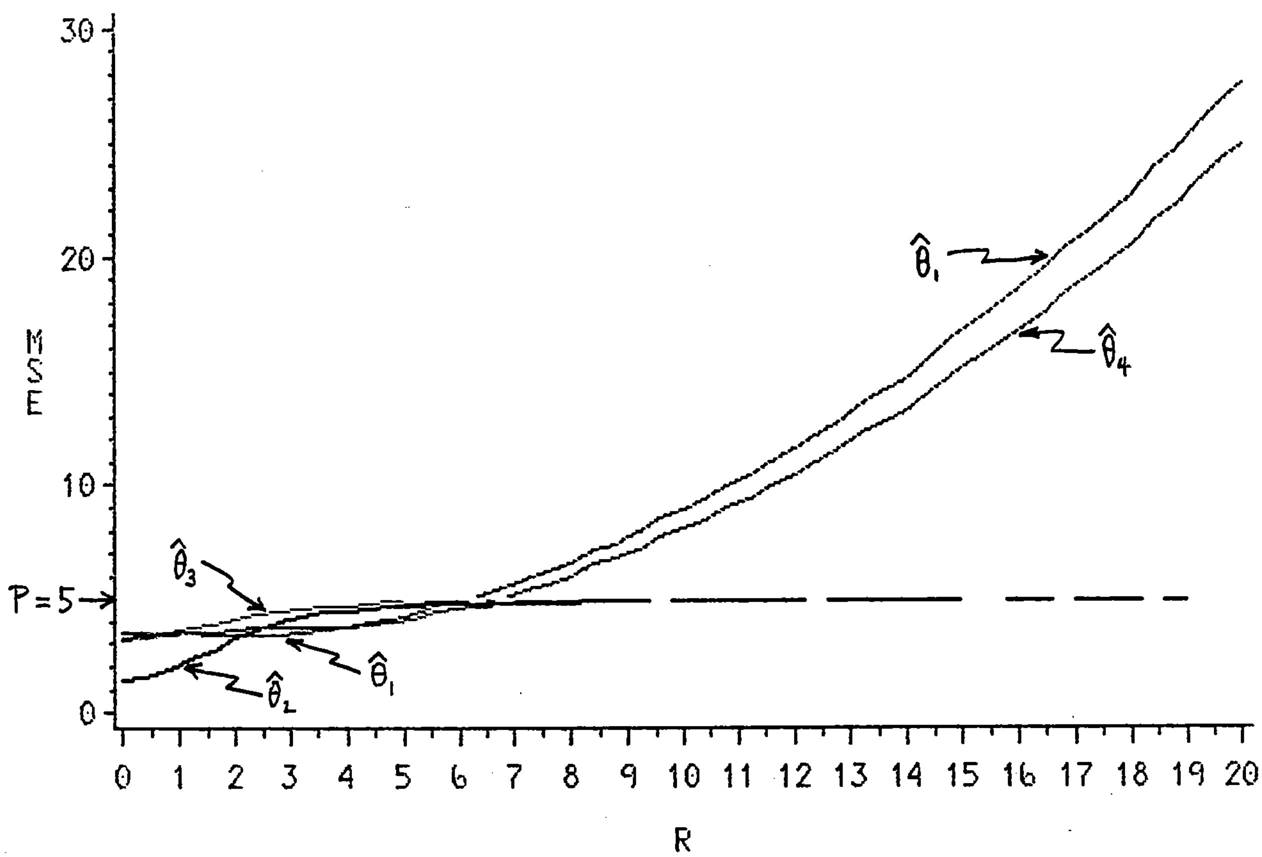


$\hat{\theta}_1$ = The Γ-minimax rule

$\hat{\theta}_2$ = The positive part James-Stein rule

$\hat{\theta}_3$ = The type-2 ML Bayes rule

FIG.3: P=5 ESTIMATOR(1,2,3,4)



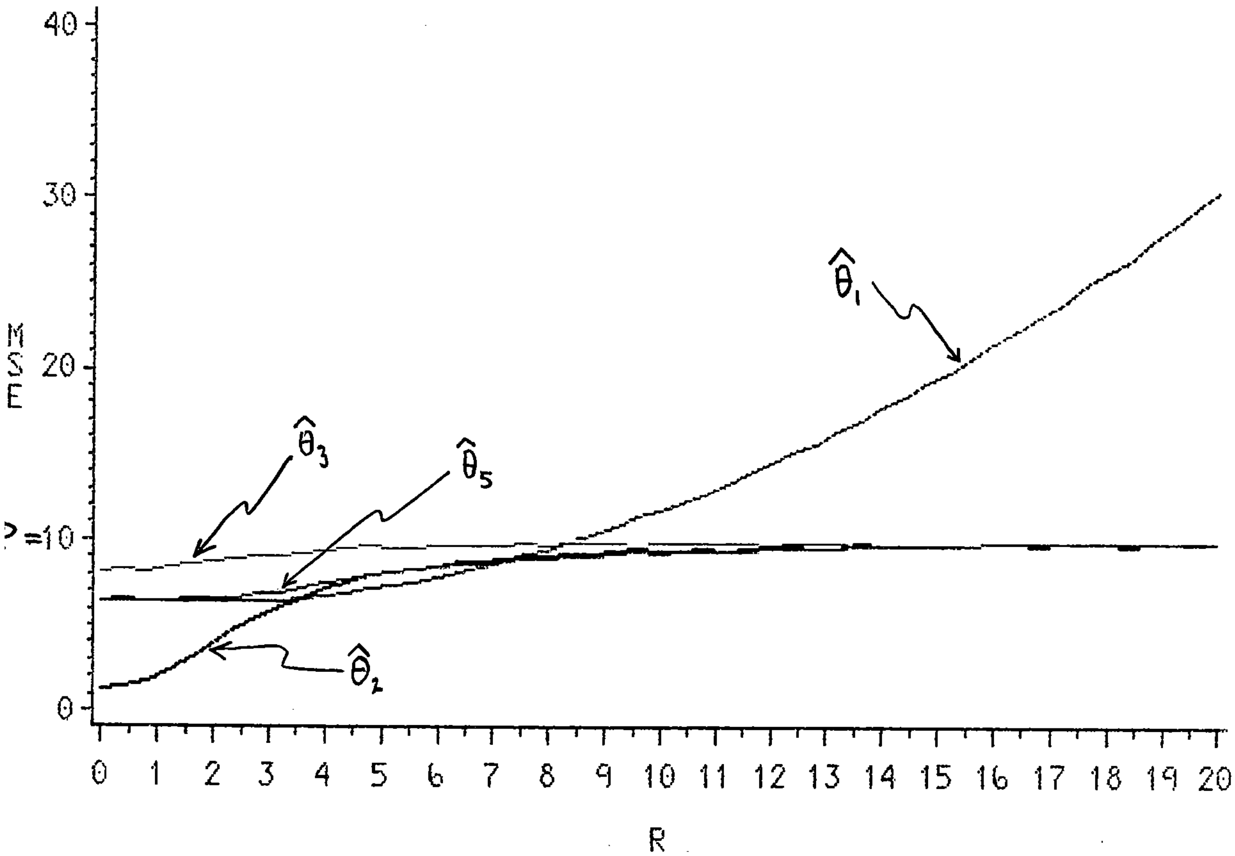
$\hat{\theta}_1$ = The Γ-minimax rule

$\hat{\theta}_2$ = The positive part James-Stein rule

$\hat{\theta}_3$ = The type-2 ML Bayes rule

$\hat{\theta}_4$ = The restricted Bayes rule (defined in (4.14))

FIG.4: P=10 ESTIMATOR(1,2,3,5)



$\hat{\theta}_1$ = The Γ-minimax rule

$\hat{\theta}_2$ = The positive part James-Stein rule

$\hat{\theta}_3$ = The type-2 ML Bayes rule

$\hat{\theta}_4$ = The restricted minimax rule (defined in (4.18))

FIG.5

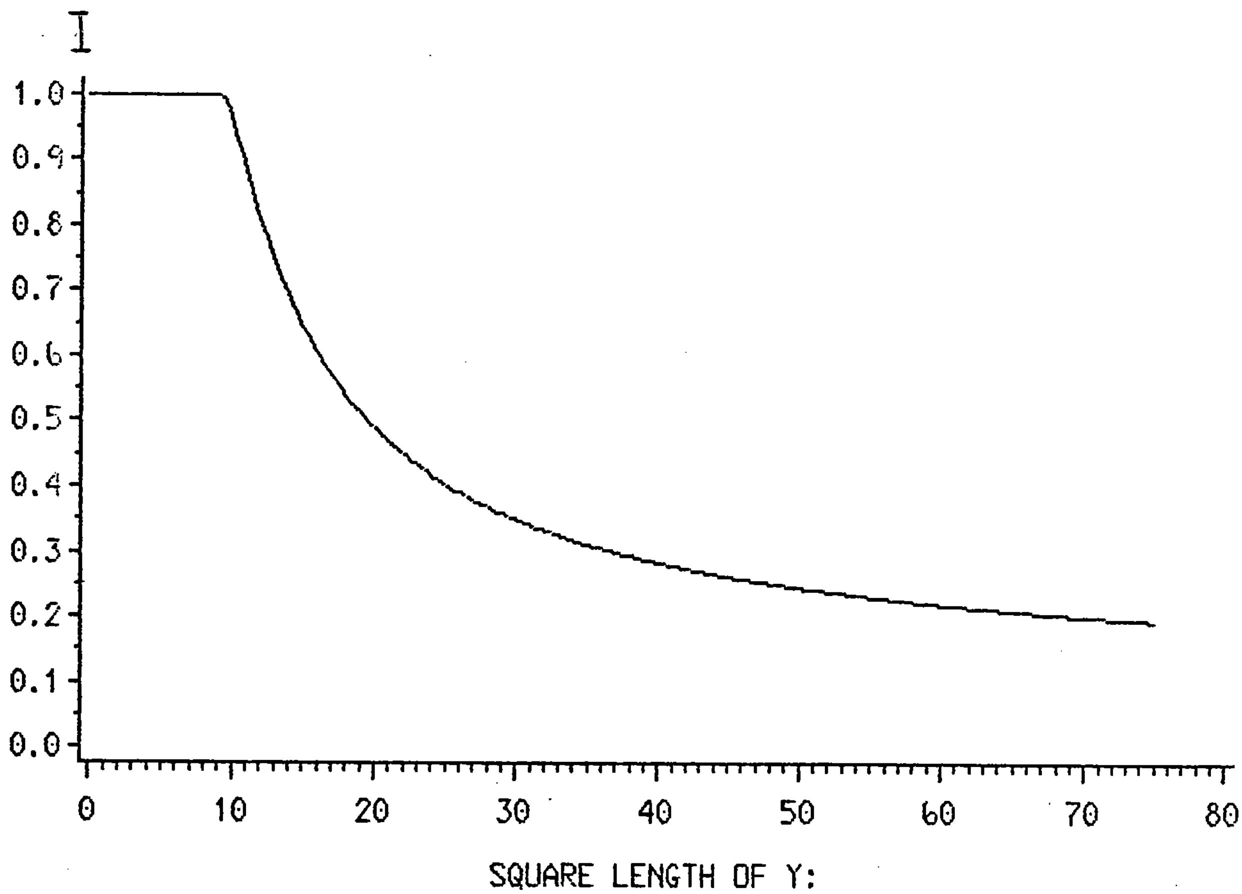
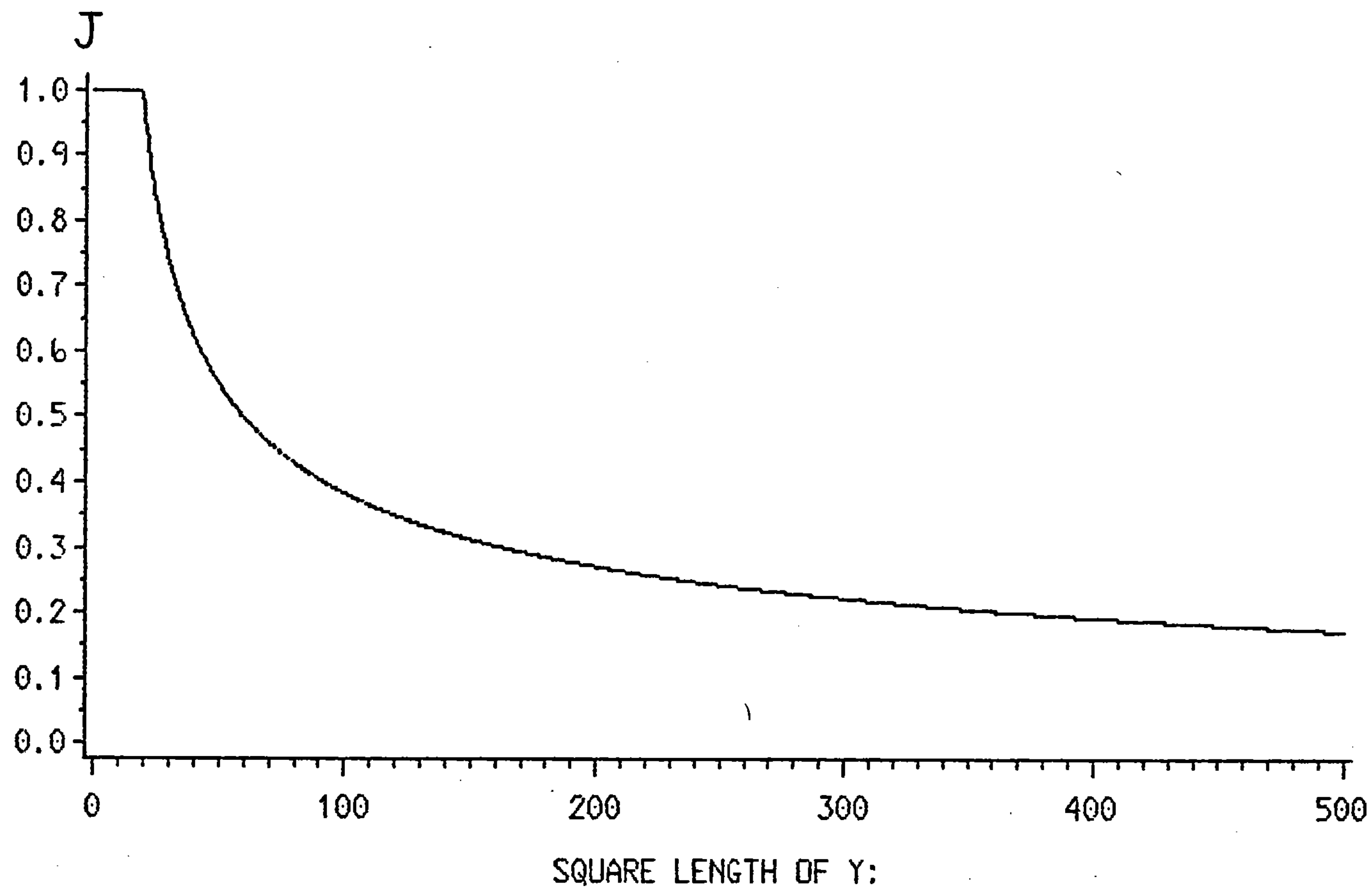

I
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
70
80
SQUARE LENGTH OF Y:

FIG.6


J
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500
SQUARE LENGTH OF Y: