ROBUST HIERARCHICAL BAYES ESTIMATION OF EXCHANGEABLE MEANS

by

James O. Berger
Department of Statistics
Purdue University
West Lafayette, IN, USA

and

Jean-Francois Angers
Department de Math et Info
Université de Sherbrooke
Sherbrooke, Quebec, CANADA

Technical Report # 88–33C

Department of Statistics
Purdue University

July 1989

ROBUST HIERARCHICAL BAYES ESTIMATION OF EXCHANGEABLE MEANS*

by

James O. Berger
Purdue University,

and

Jean-Francois Angers
Universite de Sherbrooke

ABSTRACT

Estimation of the mean of a multivaraite normal distribution is considered. The components of the mean vector, $\boldsymbol{\theta}$, are assumed to be exchangeable; this is modeled in a hierarchical fashion with independent Cauchy distributions as the first stage prior. The resulting generalized Bayes estimator is calculated and shown to be robust with respect to the presence of outlying means. Alternative estimators that have similar behavior but are cheaper to compute are also derived.

RÉSUMÉ

Dans cet article, nous étudierons l'estimation de la moyenne d'une loi normale multivariée. Nous assumerons que les composantes du vecteur moyenne, $\boldsymbol{\theta}$, sont échangeables. Cette information a priori sera représenté par un modéle hiérarchique avec des lois Cauchy indépendantes comme distribution a priori de premier niveau. L'estimateur de Bayes généralisé correspondant à ce modéle sera calculé et nous montrerons quíl est robuste par rapport à la présence de valeurs abérrantes dans le vecteur observation. D'autres estimateurs possédant cette propriété mais plus économique à calculer sont aussi développés.

Key words and phrases: Robustn estimation, Bayes estimator, Hierarchical Bayes, Normal-Cauchy convolution, Monte Carlo simulation.

AMS 1985 subjects classifications: Primary 62F35; Second 62E15; 62E25.

---

*Research supported by National Science Foundation (U.S.) Grants DMS–8702620 and DMS–8717799 and by NSERC Grant 002–007.

1. INTRODUCTION

1.1 Background

Let $Y = (Y_1, \ldots, Y_p)^t$ have a p-variate normal distribution with mean vector $\theta = (\theta_1, \ldots, \theta_p)^t$ and covariance matrix $\sigma^2 I_p$, where σ^2 is assumed to be known. The components of θ are believed to be exchangeable, and hence "shrinkage" estimation of them is desired. Most shrinkage estimators that have been developed are extremely sensitive to outlying means, in the sense that they collapse back to Y itself when there are outliers. In this paper, a hierarchical Bayes shrinkage estimator is developed which is robust with respect to outlying means.

The assumption that the components of θ are to be exchangeable can be modeled in a hierarchical Bayesian fashion, with a two stage prior. We consider a first stage prior of the form $\pi_1(\theta \mid \mu, A) = \prod_{k=1}^{p} \pi_1(\theta_k \mid \mu, A)$, where the hyperparameters μ and A represent the location and scale parameters of π_1. The prior distribution on the hyperparameters μ and A will be denoted by $\pi_2(\mu, A) = \pi_{2,1}(\mu \mid A)\pi_{2,2}(A)$. If subjective information about the location of the θ_i is available, it can be modeled through $\pi_{2,1}$. If no prior location information about θ is available, $\pi_{2,1}$ will be a noninformative prior. Note that one of the main advantages of the hierarchical Bayesian approach to the problem is the possibility of including subjective information in the exchangeable shrinkage estimator. Typically, $\pi_{2,2}$ is chosen to be noninformative.

The familiar hierarchical Bayes estimator in this situation (cf. Lindley and Smith (1972) or Berger (1985)) is derived from the choices

$$\pi_1(\theta \mid \mu, A) = N_p(\mu 1_p, A^2 I_p),$$

$$\pi_{2,1}(\mu) = N(\mu_0, \tau^2),$$

$$\pi_{2,2}(A) = A \text{ or } \pi_{2,2}(A) = 1,$$

where $1_p = (1, 1, \ldots, 1)^t$, N_p stands for a multivariate normal distribution with the indicated mean and covariance matrix, and μ_0 and τ^2 are assumed to be known (reflecting subjective information about the location of θ). Note that the given prior for A corresponds to a uniform prior for the variance, A^2, and the standard deviation, A, respectively. The estimator corresponding to this model is, for $j = 1, \ldots, p$,

$$\begin{aligned} \tilde{\theta}_j^{H_N} &= y_j - E^{\pi_{2,2}(A|y)}\left[\frac{\sigma^2}{\sigma^2 + A^2}\right](y_j - \bar{y}) \\ &\quad - E^{\pi_{2,2}(A|y)}\left[\frac{\sigma^2}{p\tau^2 + \sigma^2 + A^2}\right](\bar{y} - \mu_0) \end{aligned} \tag{1}$$

where, defining $S_Y^2 = \sum_{k=1}^p (y_k - \bar{y})^2$,

$$\pi_{2,2}(A \mid y) \propto \frac{\exp\left\{-\frac{1}{2}\left[\frac{S_Y^2}{\sigma^2+A^2} + \frac{p(\bar{y}-\mu_0)^2}{p\tau^2+\sigma^2+A^2}\right]\right\}}{\tau^{-1}(\sigma^2 + A^2)^{(p-1)/2}(p\tau^2 + \sigma^2 + A^2)^{1/2}}\pi_{2,2}(A).$$

When subjective location information about θ is not available, the common noninformative prior for μ used is $\pi_{2,1}(\mu) \equiv 1$. The resulting estimator can be obtained from equation (1) by letting τ^2 go to ∞. It is given (cf. Lindley (1971), Berger (1985)) for $j = 1, \ldots, p$, by

$$\hat{\theta}_j^{H_N} = y_j - E^{\pi_{2,2}(A|y)}\left[\frac{\sigma^2}{\sigma^2 + A^2}\right](y_j - \bar{y}), \tag{2}$$

where

$$\pi_{2,2}(A \mid y) \propto (\sigma^2 + A^2)^{-(p-1)/2} \exp\left\{-\frac{S_Y^2}{2(\sigma^2 + A^2)}\right\}\pi_{2,2}(A).$$

An empirical Bayes approximation to $\hat{\theta}^{H_N}$ was developed in Morris (1983a) and is given, for $j = 1, \ldots, p$, by

$$\hat{\theta}_j^{EB_N} = y_j - \min\left\{\frac{(p-3)}{(p-1)}, \frac{(p-3)\sigma^2}{S_Y^2}\right\}(y_j - \bar{y}). \tag{3}$$

1.2 Robustness with Respect to Partial Misspecification of the Prior

The estimators $\hat{\theta}^{H_N}$, $\hat{\theta}^{EB_N}$ and $\tilde{\theta}^{H_N}$ are usually called robust because they collapse to $\delta_0(Y) = Y$ when the data is not compatible with the prior specification. For instance, if one of the θ_i (and hence one of the Y_i) is outlying (which is "incompatible" with the prior specification of first stage normality), S_Y^2 will be large and $\widehat{\theta}^{H_N}$, $\widehat{\theta}^{EB_N}$ and $\widetilde{\theta}^{H_N}$ will collapse to $\delta_0(Y)$. Although this is "safe" behavior because of the conservative nature of $\delta_0(Y)$, it is not appealing to use an estimator which ignores the prior information because part of it is wrong. It is of considerable interest to develop estimators which discard only the part of the prior information which is not supported by the observations. An estimator with such behavior will be called robust with respect to partial misspecification of the prior.

Note that an estimator designed to accomodate the problem of robustness with respect to outlying means was developed in Dey and Berger (1983) and in Berger and Dey (1985) based on an idea of Stein (1981). It was created for the situation where each θ_j is thought to have known prior mean μ, and is given by

$$\widehat{\theta}_j^T = \left(1 - \frac{(l^* - 2)\sigma^2 \min\{1, z_{(l^*)}/|y_j - \mu|\}}{\sum_{k=1}^p (y_k - \mu)^2 \wedge z_{(l^*)}^2}\right)^+ (y_j - \mu) + \mu, \tag{4}$$

for $j = 1, \ldots, p$, where $z_{(l^*)}$ is the $(l^*)^{th}$ order statistic of $(|y_1 - \mu|, \ldots, |y_p - \mu|)$ and l^* is the value of l which maximizes

$$\frac{(l-2)^2}{\sum_{k=1}^p (y_k - \mu)^2 \wedge z_{(l)}^2}. \tag{5}$$

Later, we will consider a generalization of $\widehat{\theta}^T$ to the case in which μ is not known.

1.3 Summary of Results

A common way to achieve robustness with respect to outliers is to use flat tailed priors (cf. Box and Tiao (1968, 1973), Dawid (1973), and O'Hagan (1979)). Therefore, for the first stage prior on the θ_j we will consider independent Cauchy priors with median μ and quartiles $\mu \pm A$, i.e. $\pi_1(\theta_j \mid \mu, A) = C(\mu, A)$ for $j = 1, \ldots, p$. The second stage prior, $\pi_2(\mu, A)$, is left arbitrary. The generalized Bayes estimator, denoted by $\widehat{\theta}^{H_C}$, corresponding to this prior will be developed and its behavior when there are outlying means will be investigated. In the development of $\widehat{\theta}^{H_C}$, a fast normal-Cauchy convolution formula will be obtained. In section 3, alternative robust estimators will be developed.

In the last section, some numerical examples will be considered. First, $\widehat{\theta}^{H_C}$ will be compared with the hierarchical normal estimator $\widehat{\theta}^{H_N}$ (ref. eq. 2), and with the alternative robust estimators, when one component of the observation vector is an outlier. Another way to measure the robustness of an estimator is to compute its Bayes risk with respect to some plausible priors and compare it with the Bayes risk of estimators which are known to be robust. We will compute the Bayes risks of $\widehat{\theta}^{H_C}$ for $C(0, [0.67574]A)$ and $N(0, A^2)$ priors (which have the same quartiles) and compare them to the Bayes risks of $\widehat{\theta}^{H_N}$ and the modification of $\widehat{\theta}^{T}$.

Similar work has been done in Gaver (1985), and Gaver and O'Muircheartaigh (1987) for Poisson event rates.

2. THE ROBUST HIERARCHICAL BAYES ESTIMATOR

2.1 The Posterior Mean

If $\pi(\theta \mid \mu, A) = \prod_{k=1}^{p} \pi_1(\theta_j \mid \mu, A)$ where, for $j = 1, \ldots, p$, $\pi_1(\theta_j \mid \mu, A)$ is $C(\mu, A)$, then the posterior mean of θ_j can be written as

$$\widehat{\theta}_j^{H_C} = \frac{\int_0^\infty \int_{-\infty}^{\infty} \{\widehat{\theta}_{j|\mu,A} \prod_{k=1}^{p} m(y_k \mid \mu, A)\} \pi_2(\mu, A) d\mu dA}{\int_0^\infty \int_{-\infty}^{\infty} \{\prod_{k=1}^{p} m(y_k \mid \mu, A)\} \pi_2(\mu, A) d\mu dA}, \tag{6}$$

where

$$m(y_k \mid \mu, A) = \int_{-\infty}^{\infty} f(y_k \mid \theta_k) \pi_1(\theta_k \mid \mu, A) d\theta_k$$

$$\widehat{\theta}_{j|\mu,A} = \int_{-\infty}^{\infty} \theta_j \pi_1(\theta_j \mid \mu, A, y_j) d\theta_j,$$

$$f(y_k \mid \theta_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_k - \theta_k)^2\right\},$$

$$\pi_1(\theta_k \mid \mu, A, y_k) = \frac{f(y_k \mid \theta_k)\pi_1(\theta_k \mid \mu, A)}{m(y_k \mid \mu, A)},$$
$$\pi_1(\theta_k \mid \mu, A) = \frac{A}{\pi\left[A^2 + (\theta_k - \mu)^2\right]}.$$

The advantage of using equation 6 to compute $\widehat{\theta}^{Hc}$ is that it is effectively only a three dimensional numerical integral, even though the total number of parameters is $p+2$. However, the inner integral is a product of $p+1$ simple integrals. Consequently, an accurate and inexpensive way to compute these normal-Cauchy convolutions will be needed if we want $\widehat{\theta}^{Hc}$ to be calculable in practice.

2.2 Normal-Cauchy Convolution

For the moment, assume that μ and A are given.

Theorem 1 *Suppose that $Y_j \sim N(\theta_j, \sigma^2)$ independently for $j = 1, \ldots, p$ and that $\theta_j \sim \mathcal{C}(\mu, A)$ independently for $j = 1, \ldots, p$ where σ^2, μ and A are known. Then*

$$\begin{aligned} m(y_j \mid \mu, A) &= \text{marginal of } y_j \text{ given } \mu \text{ and } A \\ &= \frac{\sqrt{2}}{\pi\sigma}\Re\phi^*(t_j), \qquad (7) \\ \widehat{\theta}_{j|\mu,A} &= \text{posterior mean of } \theta_j \text{ given } \mu \text{ and } A \text{ and } y_j \\ &= \mu + A\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)}, \qquad (8) \\ V_{j,j|\mu,A} &= \text{posterior variance of } \theta_j \text{ given } \mu \text{ and } A \text{ and } y_j \\ &= \frac{A\sigma}{\sqrt{2}}\frac{1}{\Re\phi^*(t_j)} - A^2\left[1 + \left(\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)}\right)^2\right]; \qquad (9) \end{aligned}$$

here $\Re\phi^*(t_j)$ and $\Im\phi^*(t_j)$ denote the real and imaginary parts of the complex function

$$\begin{aligned} \phi^*(t_j) &= \exp(t_j^2)\left[\frac{\sqrt{\pi}}{2} - \int_0^{t_j} \exp(-z^2)dz\right], \qquad (10) \\ t_j &= \frac{A - i(y_j - \mu)}{\sqrt{2\sigma^2}}, \end{aligned}$$

where $i = \sqrt{-1}$ and the integral in equation (10) is a contour integral.

Proof: Given in appendix A.

Note that the posterior covariance of θ_j and θ_k, given μ and A, is equal to 0 for all $j \neq k$ since the y_j's are independent as are the θ_j a priori.

In the previous theorem, all the quantities of interest ($m(y_j \mid \mu, A)$, $\widehat{\theta}_{j|\mu,A}$, $V_{j,j|\mu,A}$) have been expressed as functions of the complex function $\phi^*(\cdot)$. A fast method of evaluating $\phi^*(\cdot)$ using complex continued fractions is given in appendix B.

2.3 The Hierchical Bayes Estimator

From equation (6), it is clear that $\widehat{\theta}_j^{Hc}$ is obtained by integrating $\widehat{\theta}_{j|\mu,A}$ with respect to the measure

$$\pi_2(\mu, A \mid y) = \frac{\{\prod_{k=1}^p m(y_k \mid \mu, A)\}\pi_2(\mu, A)}{\int_0^\infty \int_{-\infty}^\infty \{\prod_{k=1}^p m(y_k \mid \mu, A)\}\pi_2(\mu, A) d\mu dA}. \tag{11}$$

If this technique is also applied to the other quantities of interest (marginal of y, posterior variance), and Theorem 1 is utilized, the following is obtained.

Theorem 2 *Let $Y_j \sim N(\theta_j, \sigma^2)$ (σ^2 known) and let $\theta_j \sim C(\mu, A)$ (μ, A unknown), $j = 1, \ldots, p$. If $\pi_2(\mu, A)$ is the prior density of (μ, A) and if*

$$\rho^* = \int_0^\infty \int_{-\infty}^\infty A^m |\mu|^n \{\prod_{k=1}^p \Re\phi^*(t_k)\}\pi_2(\mu, A) d\mu dA < \infty$$

for all $m \geq 0$ and $n \geq 0$ such that $m + n \leq 2$, then

$$\begin{aligned}
m(y) &= \left(\frac{2}{\pi^2\sigma^2}\right)^{p/2} \int_0^\infty \int_{-\infty}^\infty \left[\prod_{k=1}^p \Re\phi^*(t_k)\right] \pi_2(\mu, A) d\mu dA, \\
\widehat{\theta}_j^{Hc} &= \text{posterior mean of } \theta_j \\
&= E^{\pi_2(\mu,A)|y)}\left[\mu + A\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)}\right], \qquad (12) \\
V_{j,j}^{Hc} &= \text{posterior variance of } \theta_j \\
&= E^{\pi_2(\mu,A|y)}\left[\frac{A\sigma}{\sqrt{2}}\frac{1}{\Re\phi^*(t_j)} + (\mu^2 - A^2) + 2\mu A\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)}\right] - \left(\widehat{\theta}_j^{Hc}\right)^2,
\end{aligned}$$

$$
\begin{aligned}
V_{j,k}^{Hc} &= \textit{posterior covariance of } \theta_j \textit{ and } \theta_k \\
&= E^{\pi_2(\mu,A)|y)}\left[\mu^2 + 2\mu A\left(\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)} + \frac{\Im\phi^*(t_k)}{\Re\phi^*(t_k)}\right)\right. \\
&\qquad \left. + A^2\frac{(\Im\phi^*(t_j)\Im\phi^*(t_k))}{(\Re\phi^*(t_j)\Re\phi^*(t_k))}\right] - \widehat{\theta}_j^{Hc}\widehat{\theta}_k^{Hc},
\end{aligned}
$$

where

$$
\pi_2(\mu, A \mid y) = \left(\frac{2}{\pi^2\sigma^2}\right)^{p/2} \frac{[\prod_{k=1}^{p} \Re\phi^*(t_k)]\,\pi_2(\mu, A)}{m(y)}.
$$

Proof: Straightfoward.

Proposition 1 *Let* $\pi_2(\mu, A) \equiv 1$. *If* $p \geq 5$ *then* $\rho^* < \infty$.

Proof: Given in the appendix A.

2.4 Calculation

To calculate the three dimensional numerical integrals in Theorem 2, several techniques (IMSL subroutines, Gaussian quadrature, Monte-Carlo, etc.) can be applied. The Monte-Carlo method with importance sampling function was chosen over the others because it seemed no more expensive and the precision of the result obtained by this method is easier to control. Also, in section 4.2 the frequentist Bayes risk of $\widehat{\theta}^{Hc}$ will be computed, and to perform the additional integrations over the sample space Monte-Carlo is now definitely cost effective.

To compute $\widehat{\theta}^{Hc}$ for the noninformative prior $\pi_2(\mu, A) = 1$, one can use the following scheme. If $(\mu^{(l)}, A^{(l)})$, $l = 1, \ldots, n$, are i.i.d. with density $g(\cdot,\cdot)$ (the importance sampling function), an estimate of the posterior expectation of $h(\theta_j)$ is

$$
\begin{aligned}
\widehat{h(\theta_j)} &= \int_0^\infty \int_{-\infty}^\infty \widehat{h}(\theta_{j|\mu,A})\pi_2(\mu, A \mid y)d\mu dA \\
&\cong \frac{\sum_{l=1}^n \widehat{h}(\theta_{j|\mu^{(l)},A^{(l)}})\left[\prod_{k=1}^p \Re\phi^*(t_k^{(l)})\pi_2(\mu^{(l)}, A^{(l)})\right]/g(\mu^{(l)}, A^{(l)})}{\sum_{l=1}^n \left[\prod_{k=1}^p \Re\phi^*(t_k^{(l)})\pi_2(\mu^{(l)}, A^{(l)})\right]/g(\mu^{(l)}, A^{(l)})}, \qquad (13)
\end{aligned}
$$

where

$$\widehat{h}(\theta_{j|\mu^{(l)},A^{(l)}}) = \int_{-\infty}^{\infty} h(\theta_j)\pi_1(\theta_j \mid \mu^{(l)}, A^{(l)}, y_j)d\theta_j,$$

$$t_j^{(l)} = \frac{A^{(l)} - i(y_j - \mu^{(l)})}{\sqrt{2\sigma^2}}.$$

The importance sampling function, $g(\mu, A)$, should be as similar to the posterior of (μ, A), $\pi_2(\mu, A \mid y)$ (ref. eq. 11), as possible. Therefore, $g(\mu, A)$ will be chosen such that its location and scale parameters are approximatively equivalent to those of $\pi_2(\mu, A \mid y)$. Let $(\widehat{\mu}_1, \widehat{A}_1)$ and $\begin{pmatrix} v_\mu & c_{\mu,A} \\ c_{\mu,A} & v_A \end{pmatrix}$ be estimates of the location parameter and the covariance matrix of $\pi_2(\mu, A \mid y)$; these will be defined shortly. If one wrote $g(\mu, A) = g_1(\mu \mid A)g_2(A)$ and assumed normality, the mean and variance of g_1 would be equal to $\widehat{\mu}_1 - \frac{c_{\mu,A}}{v_A}(A - \widehat{A}_1)$ and $v_\mu - \frac{(c_{\mu,A})^2}{v_A}$, respectively, and those of g_2 would be equal to $\widehat{A_1}$ and v_A. However, since heavy tails are desired (large values for $\pi_2(\mu, A \mid y)/g(\mu, A)$ are to be avoided in importance sampling), $g(\mu, A)$ will be chosen to be

$$g(\mu, A) = C\left(\widehat{\mu}_1 - \frac{c_{\mu,A}}{v_A}(A - \widehat{A}_1), [0.67574]\left(v_\mu - \frac{(c_{\mu,A})^2}{v_A}\right)^{1/2}\right)$$
$$\times C\left(\widehat{A}_1, [0.67574](v_A)^{1/2}\right) I_{[0,\infty)}(A).$$

(Note that the scale parameters have been multiplied by 0.67574 in order to match the normal and Cauchy densities at their 1^{st} and 3^{rd} quartiles.)

The estimates $(\widehat{\mu}_1, \widehat{A}_1)$ and $\begin{pmatrix} v_\mu & c_{\mu,A} \\ c_{\mu,A} & v_A \end{pmatrix}$ are the result of a Monte-Carlo prerun. The importance sampling function used in this prerun is $g_0(\mu, A) = g_{0,1}(\mu)g_{0,2}(A)$, where $g_{0,1}(\mu) = C(\widehat{\mu}_0, 1)$, $g_{0,2}(A) = C(\widehat{A}_0, 1)I_{[0,\infty)}(A)$, and where $\widehat{\mu}_0$ and $\widehat{A}_0$ are the sample median and $\left[\left(\frac{IQ}{1.35148}\right)^2 - \sigma^2\right]^{1/2}$, respectively, for the data $(y_1, \ldots, y_p)$; here IQ is the sample interquartile distance. Using the above scheme, the quantities in Theorem 2 can be computed in 13 cpu seconds when $p = 5$ on a CDC65000 computer. For $p = 10$, the same calculations take about 21 cpu seconds.

2.5 Behavior of $\widehat{\theta}^{Hc}$ in the Presence of Outliers

To study the behavior of $\widehat{\theta}^{Hc}$ in the presence of outliers, we will define the distance, d_l, of an observation, y_l, from the rest of the observation vector

$$y_{(-l)} = (y_1, \ldots, y_{l-1}, y_{l+1}, \ldots, y_p)^t$$

as being

$$d_l = \min_{k \neq l}\{|y_k - y_l|\}.$$

Intuitively, y_l will be an outlier (corresponding to an outlying mean) if d_l is large. To begin, we consider the behavior of the posterior density of μ and A.

Lemma 1 *Let $\pi_2(\mu, A \mid y)$ (ref. eq. 11) be the posterior of (μ, A). Suppose that*

$$\int_0^\infty \int_{-\infty}^\infty A^m |\mu|^n \left\{ \prod_{k \neq l} \Re\phi^*(t_k) \right\} \pi_2(\mu, A) d\mu dA < \infty \qquad (14)$$

for $m \geq -1$, $n \geq 0$ and $m + n \leq 1$, then

$$\lim_{d_l \to \infty} \pi_2(\mu, A \mid y) = \pi_2^*(\mu, A \mid y_{(-l)}), \qquad (15)$$

where

$$\pi_2^*(\mu, A \mid y_{(-l)}) = \frac{A\{\prod_{k=1, k\neq l}^p \Re\phi^*(t_k)\}\pi_2(\mu, A)}{\int_0^\infty \int_{-\infty}^\infty A\{\prod_{k=1, k\neq l}^p \Re\phi^*(t_k)\}\pi_2(\mu, A) d\mu dA}.$$

Note: As in the proof of Proposition 1, it can be shown that eq. 14 is satisfied if $\pi_2(\mu, A) = 1$ and $p \geq 5$.

Proof: Given in the appendix A.

Note interestingly, that $\pi_2^*(\mu, A \mid y_{(-l)})$ is not the posterior that would have resulted from consideration of the problem with the l^{th} coordinate omitted; there is an additional multiplicative factor of A.

Theorem 3 *Let $Y_j \sim N(\theta_j, \sigma^2)$ (σ^2 known) and let $\theta_j \sim \mathcal{C}(\mu, A)$, where $j = 1, \ldots, l-1, l+1, \ldots, p$, and assume that eq. 14 holds for $m \geq -1$, $n \geq 0$ and $m + n \leq 2$. If $d_l \to \infty$ (i.e. y_j is an outlier), then*

$$\widehat{\theta}_l^{Hc} \to y_l, \tag{16}$$

$$\widehat{\theta}_j^{Hc} \to E^{\pi_2^*(\mu,A|\mathbf{y}_{(-l)})}\left[\widehat{\theta}_{j|\mu,A}\right]. \tag{17}$$

Proof: Similar to that of Lemma 1.

Thus, $\widehat{\theta}^{Hc}$ behaves (for large d_l) as desired; the outlying y_l does not eliminate the possibility of Bayesian shrinkage of the other coordinates. Note that, as in the proof of Proposition 1, it can be shown that the condition of Theorem 3 is satisfied by $\pi_2(\mu, A) = 1$ if $p \geq 6$.

3. ALTERNATIVE ESTIMATORS

This section will be concerned with the development of two alternative estimators that have behavior similar to that of $\widehat{\theta}^{Hc}$ in the presence of outliers, but require less numerical calculation. The first will be obtained by taking a hierarchical version of a robust estimator developed in Berger (1985). The second estimator is an *ad hoc* modification of the truncated estimator already discussed in section 1.4.

3.1 Development of $\widehat{\theta}^{GS}$

An alternative first stage prior for the θ_j's, which has flat tails and allows for closed form convolution with a normal likelihood, was developed in Berger (1985) (similar to ones developed in Strawderman (1971) and Berger (1980)). It can most easily be represented hierarchically as $\theta_j \sim N(\mu, B(\lambda_j))$, $j = 1, \ldots, p$, where $B(\lambda_j) = 0.5\lambda_j^{-1}(\sigma^2 + A^2) - \sigma^2$

and λ_j has prior density $\xi(\lambda_j) = 0.5\lambda_j^{-1/2}I_{(0,1)}(\lambda_j)$. We still view μ and A as being hyperparameters with prior $\pi_2(\mu, A)$.

Given μ and A, the posterior mean of θ_j is given by

$$\widehat{\theta}^{GS}_{j|\mu,A} = y_j - \frac{2\sigma^2}{(\sigma^2 + A^2)}\left(\frac{1}{\|y_j\|^2} - \frac{1}{(e^{\|y_j\|^2} - 1)}\right)(y_j - \mu), \tag{18}$$

the posterior variances by

$$V^{GS}_{j,j|\mu,A} = \sigma^2 - \frac{2\sigma^4}{(\sigma^2 + A^2)}\left\{\frac{1}{(e^{\|y_j\|^2} - 1)}\left[\frac{2\|y_j\|^2}{(1 - e^{-\|y_j\|^2})} - 1\right] - \frac{1}{\|y_j\|^2}\right\},$$

and the marginal of y_j by

$$m(y_j \mid \mu, A) = \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{\sigma^2 + A^2}}\left[\frac{1 - e^{-\|y_j\|^2}}{\|y_j\|^2}\right], \tag{19}$$

where $\|y_j\|^2 = (y_j - \mu)^2/(\sigma^2 + A^2)$ (cf. Berger (1985)). As previoulsy mentioned, the posterior covariance of θ_j and θ_k, given μ and A, is equal to 0 for all $j \neq k$. These formulas lead to the hierarchical Bayes estimator

$$\widehat{\theta}^{GS}_j = E^{\pi_2(\mu,A|y)}\left[\widehat{\theta}^{GS}_{j|\mu,A}\right], \tag{20}$$

where

$$\pi_2(\mu, A \mid y) = \frac{[\prod_{k=1}^p m(y_k \mid \mu, A)]\,\pi_2(\mu, A)}{m(y)}, \tag{21}$$

$$m(y) = \frac{1}{(4\pi)^{p/2}}\int_0^\infty\int_{-\infty}^\infty \frac{1}{(\sigma^2 + A^2)^{p/2}}\prod_{k=1}^p\left[\frac{1 - e^{-\|y_k\|^2}}{\|y_k\|^2}\right]\pi_2(\mu, A)d\mu dA.$$

The posterior variances and covariances are given by

$$V^{GS}_{j,j} = E^{\pi_2(\mu,A|y)}\left[V^{GS}_{j,j|\mu,A}\right] + E^{\pi_2(\mu,A|y)}\left[(\widehat{\theta}^{GS}_{j|\mu,A})^2\right] - \left(\widehat{\theta}^{GS}_j\right)^2,$$

$$V^{GS}_{j,k} = 4E^{\pi_2(\mu,A|y)}\left[\frac{\sigma^4}{(\sigma^2 + A^2)^2}\left(\frac{1}{\|y_j\|^2} - \frac{1}{(e^{\|y_j\|^2} - 1)}\right)(y_j - \mu) \times \left(\frac{1}{\|y_k\|^2} - \frac{1}{(e^{\|y_k\|^2} - 1)}\right)(y_k - \mu)\right] - (y_j - \widehat{\theta}^{GS}_j)(y_k - \widehat{\theta}^{GS}_k).$$

Note that since $\widehat{\theta}^{GS}_{j|\mu,A}$ (ref. eq. 18) is given in closed form, $\widehat{\theta}^{GS}_{j}$ will be about 50% cheaper to compute than $\widehat{\theta}^{Hc}_{j}$. Existence of these expectations for noninformative choices of $\pi_2(\mu, A)$ is guaranteed by the following proposition.

Proposition 2 *If $\pi_2(\mu, A) = A$ with $p \geq 5$, or $\pi_2(\mu, A) = 1$ with $p \geq 4$, and the y_j are not all equal, then $m(y)$, the $\widehat{\theta}^{GS}_{j}$ and the $V^{GS}_{j,k}$ all exist and are finite.*

Proof: Given in the appendix A.

To study the behavior of $\widehat{\theta}^{GS}$ in the presence of an outlier, it is first necessary to consider the behavior of the marginal of y_j, given μ and A, as $\|y_j\|^2$ goes to infinity.

Theorem 4 *Let $\pi_2(\mu, A \mid y)$ be given by eq. 21 and suppose that*

$$\int_0^\infty \int_{-\infty}^\infty \frac{[\mu^2 + \sigma^2 + A^2]}{\sqrt{\sigma^2 + A^2}} \left\{ \prod_{k \neq l} m(y_k \mid \mu, A) \right\} \pi_2(\mu, A) d\mu dA < \infty. \tag{22}$$

If $d_l \to \infty$ (i.e. y_l is an outlier, using the notation in section 2.5), then

$$\begin{aligned} \pi_2(\mu, A \mid y) &\to \pi_2^*(\mu, A \mid y_{(-l)}), \\ \widehat{\theta}^{GS}_{l} &\to y_l, \\ \widehat{\theta}^{GS}_{j} &\to E^{\pi_2^*(\mu, A|y_{(-l)})}\left[\widehat{\theta}^{GS}_{j|\mu,A}\right] \text{ for } j \neq l, \end{aligned} \tag{23}$$

where

$$\pi_2^*(\mu, A \mid y_{(-l)}) = \frac{(\sigma^2 + A^2)^{1/2} \left[\prod_{k \neq l} m(y_k \mid \mu, A)\right] \pi_2(\mu, A)}{\int_0^\infty \int_{-\infty}^\infty (\sigma^2 + A^2)^{1/2} \left[\prod_{k \neq l} m(y_k \mid \mu, A)\right] \pi_2(\mu, A) d\mu dA}.$$

Proof: Given in the appendix A.

Note: As in the proof of Proposition 2, it can be shown that the noninformative prior $\pi_2(\mu, A) = A$ satisfies eq. 22 providing $p \geq 7$ and that there are at least 3 distinct y_j; the same can be said for $\pi_2(\mu, A) = 1$ when $p \geq 6$.

Thus $\widehat{\theta}^{GS}$ also does not allow an outlier to prevent Bayesian shrinkage of the other coordinates.

3.2 The Truncated Estimator

If the θ_j's are thought to be independent realizations from a common symmetric prior distribution having known median μ, it was suggested in Dey and Berger (1983) (based on an estimator from Stein(1981)) to estimate θ_j by the truncated estimator defined in equation 4. When the prior median, μ, of θ is unknown, one can use a robust estimator to estimate μ and modify $\widehat{\theta_j^T}$ to account for this estimation. A sensible robust estimator for μ is the α-trimmed mean

$$\overline{y}_\alpha = \frac{1}{p-2g} \sum_{k=g+1}^{p-g} y_{(k)},$$

where 2α is the proportion of observations trimmed, g is the integer part of αp and $y_{(k)}$ is the k^{th} order statistic of $(y_1, \ldots, y_p)$. The choice of $\alpha = \alpha(l) \equiv (p - l + 1)/(2p)$ seems natural, since $\overline{y}_\alpha$ will then trim approximatively $(p - l)$ observations, as is implicitely done in equation 4. A natural modification of $\widehat{\theta_j^T}$ is then:

$$\widehat{\theta_j^{T_*}} = \left(1 - \frac{(l^* - 3)\sigma^2 \min\{1, z_{(l^*)}/|y_j - \overline{y}_{\alpha^*}|\}}{\sum_{k=1}^{p} (y_k - \overline{y}_{\alpha^*})^2 \wedge z_{(l^*)}^2}\right)^+ (y_j - \overline{y}_{\alpha^*}) + \overline{y}_{\alpha^*}, \tag{24}$$

where $z_{(k)}$ is the k^{th} order statistic of $(|y_1 - \overline{y}_{\alpha^*}|, \ldots, |y_p - \overline{y}_{\alpha^*}|)$, $\alpha^* = \alpha(l^*)$, and l^* is the value of l $(l \geq 4)$ which maximizes

$$\frac{(l-3)^2}{\sum_{k=1}^{p} (y_k - \overline{y}_{\alpha(l)})^2 \wedge z_{(l)}^2}.$$

(We have replaced $(l - 2)$ in equations 4 and 5 by $(l - 3)$, because of the familiar "loss of a dimension" when shrinking towards an estimated common mean.) As for the previous alternative estimator, it can be shown that, as $d_l \to \infty$, the outlier, y_l , does not seriously affect the shrinkage of $\widehat{\theta_j^{T_*}}$ for $j \neq l$.

4. NUMERICAL COMPARISON OF ESTIMATORS

4.1 Behavior in the Presence of Outliers

For illustrative purposes, the behavior of $\widehat{\theta}^{H_C}$ in (12), $\widehat{\theta}^{H_N}$ in (2), $\widehat{\theta}^{GS}$ in (20), and $\widehat{\theta}^{T_*}$ in (24) will be compared in an example when the sample contains an outlier. Altough the first stage prior for $\widehat{\theta}^{H_C}$, $\widehat{\theta}^{H_N}$ and $\widehat{\theta}^{GS}$ differ, we selected the same noninformative prior $\pi_{2,2}(A) = 1$, at the second stage for each estimator. For the comparison, we generated a sample of size 10 where each observation, Y_j, is $N(\theta_j, 1)$ and the θ_j's were generated according to a $N(0,1)$ distribution. Using this scheme, the sample turned out to be $(-0.068, 0.969, 1.329, -0.512, 0.071, 2.892, 1.944, 0.671, -0.018, 0.008)$. To study the effect of an outlier on the estimators, we added an integer k to the last observation, $y_{10} = 0.008$. The value of k was varied from 0 to 12 by steps of 2.

In the first series of graphs (figures 1 and 2) we compare $\widehat{\theta}^{H_C}$ (dotted square), $\widehat{\theta}^{H_N}$ (full diamond), $\widehat{\theta}^{GS}$ (square), $\widehat{\theta}^{T_*}$ (diamond) and y (full square). Looking at these graphs, we can see that all the estimators are nearly equal for small k ($k = 0, 2$), and that $\widehat{\theta}^{H_N}$ collapses back to y when k gets large while the others stabilize away from y. The only exception to this behavior is for $y_2 = 0.969$, where $\widehat{\theta}_2^{T_*}$ seems to collapse to y_2, but it may be caused by the fact that, for large k, $\alpha^* = 1/p$ and $\bar{y}_{1/p} = 0.973$. We can also notice that $\widehat{\theta}^{H_C}$, $\widehat{\theta}^{GS}$ and $\widehat{\theta}^{T_*}$ behave in a similar fashion except, possibly, for y_2, y_8 and y_9. For the outlier θ_{10} ($y_{10} = k + 0.008$), one can see that all the estimators are essentially equivalent.

In figure 3 and 4, we compare the posterior variance for each of the components of $\widehat{\theta}^{H_C}$ (dotted square), $\widehat{\theta}^{H_N}$ (full diamond) and $\widehat{\theta}^{GS}$ (square) under their respective priors. (We did not compute the posterior variance of $\widehat{\theta}^{T_*}$, since this estimator is not based on a specific prior. Note also that a smaller posterior variance does not necessarily mean that the estimator is better, since the computations are for different priors; the goal is simply to compare the measures of accuracy produced in each scenario.) Generally, the posterior variance is an increasing function of k. For $\widehat{\theta}^{H_C}$ and $\widehat{\theta}^{GS}$ it seems to stabilize at a value less than one, while for $\widehat{\theta}^{H_N}$ it is steadily increasing towards one. We can also notice that, when k is large, the posterior variances of $\widehat{\theta}^{H_C}$ and $\widehat{\theta}^{GS}$ are usually almost equivalent. If we look at the graph for the outlier, θ_{10} ($y_{10} = k + 0.008$), we can see that, for sufficiently large k, the posterior variances of the three estimators are essentially the same.

4.2 Bayes Risk Comparison

Also of interest is comparison of the overall performance of $\widehat{\theta}^{H_C}$, $\widehat{\theta}^{H_N}$ and $\widehat{\theta}^{T_*}$, in terms of frequentist Bayes risk with respect to the squared error loss. Here, also, the same noninformative prior, $\pi_{2,2}(A) = 1$, was used for $\widehat{\theta}^{H_N}$ and $\widehat{\theta}^{H_C}$ at the second stage. We did not include $\widehat{\theta}^{GS}$ in this study. Two cases are considered: in the first the true prior is $C(\mu, [0.67574]A)$, and in the second it is $N(\mu, A^2)$. These priors were chosen because they have the same quartiles and correspond to the first stage of the hierarchical priors used in the development of $\widehat{\theta}^{H_C}$ and $\widehat{\theta}^{H_N}$, respectively. Without loss of generality, μ will be taken to be zero. However, A will be considered as a parameter and the frequentist Bayes risk comparison will be done as a function of A. In order to compare results, it is helpful to also consider the relative savings loss (RSL), introduced in Efron and Morris (1971), defined as

$$\text{RSL}(\pi, \widehat{\theta}) = \frac{r(\pi, \widehat{\theta}) - r(\pi, \widehat{\theta^\pi})}{r(\pi, y) - r(\pi, \widehat{\theta^\pi})},$$

where $\widehat{\theta^\pi}$ is the Bayes rule if π is the true prior and $r(\pi, \hat{\theta})$ is the Bayes risk of $\hat{\theta}$. The RSL were computed by using the following proposition (whose proof is standard).

Proposition 3 *Under squared error loss,*

$$r(\pi, \widehat{\theta}) - r(\pi, \widehat{\theta^\pi}) = \int \|\widehat{\theta} - \widehat{\theta^\pi}\|^2 m(y) dy,$$

where $\|\widehat{\theta} - \widehat{\theta^\pi}\|^2 = \sum_{k=1}^{p} (\widehat{\theta}_k - \widehat{\theta^\pi_k})^2$.

The integrals involved were evaluated by Monte-Carlo analysis. (Note that we did not need to use an importance sampling function here, since the θ_j's and the y_j's can be generated directly from their distribution.) The average accuracy in the calculation of the RSL was about 0.05 for both priors when $p = 5$.

Figure 5 shows the difference in the actual Bayes risk between $\widehat{\theta}^{H_C}$ (dotted square), $\widehat{\theta}^{H_N}$ (full diamond), $\widehat{\theta}^{T\bullet}$ (square) and the Bayes rule, as a function of the scale parameter A, when the dimension of θ is equal to 5. Figure 6 gives the RSL's of $\widehat{\theta}^{H_C}$ (dotted square), $\widehat{\theta}^{H_N}$ (full diamond) and $\widehat{\theta}^{T\bullet}$ (square) as a function of A. Also graphed in figure 6 (diamond) is the maximum improvement in the Bayes risk that one can achieve over y (namely $[p - r(\pi, \widehat{\theta^{\pi}})]/p$). Because of the cost and the complexity of the simulation, only the case $p = 5$ has been studied.

From figure 5, it is clear that, for small A, $\widehat{\theta}^{H_N}$ does better than $\widehat{\theta}^{H_C}$ under either prior. However, in the Cauchy case, $\widehat{\theta}^{H_C}$ performs better than $\widehat{\theta}^{H_N}$ for intermediate and larger values of A. Under the normal prior, $\widehat{\theta}^{H_C}$ outperforms $\widehat{\theta}^{H_N}$ only for larger values of A. In both cases, $\widehat{\theta}^{H_C}$ and $\widehat{\theta}^{H_N}$ do better than $\widehat{\theta}^{T\bullet}$ when A is small; but for intermediate and larger values, $\widehat{\theta}^{T\bullet}$ performs better than $\widehat{\theta}^{H_N}$ when the true prior is Cauchy.

Looking at figure 6, one can see that the maximum possible improvement in the Bayes risk is almost negligible for large A. Consequently, the comparison of the RSL's is of less interest for large A. When the true prior is Cauchy, the RSL of $\widehat{\theta}^{H_C}$ is the smallest except for small values of A. In the normal case, $\widehat{\theta}^{H_N}$ has the smallest RSL, but for intermediate values of A the RSL of $\widehat{\theta}^{H_C}$ and $\widehat{\theta}^{H_N}$ are quite similar. When the true prior is Cauchy, $\widehat{\theta}^{T\bullet}$ seems to be a good compromise between $\widehat{\theta}^{H_C}$ and $\widehat{\theta}^{H_N}$ in terms of RSL but it is not so clear for the normal case.

4.3 Cost Comparisons

The calculations needed to compute $\widehat{\theta}^{H_C}$, $\widehat{\theta}^{H_N}$ and $\widehat{\theta}^{GS}$ were performed on a CDC6500 computer at Purdue University. (Note that, to compute $\widehat{\theta}^{T_*}$, only a pocket calculator is required.) When $p = 10$, it takes on average about 2.2 CPU seconds to compute one component of $\widehat{\theta}^{H_C}$, 0.4 CPU seconds for one of $\widehat{\theta}^{H_N}$ and 0.9 CPU seconds for one of $\widehat{\theta}^{GS}$.

4.4 Conclusion

The previous figures indicate that $\widehat{\theta}^{H_C}$ is a better estimator to use than $\widehat{\theta}^{H_N}$, unless one feels quite confident in a first stage normality assumption. Moreover, $\widehat{\theta}^{H_C}$ has a better or similar Bayes risk and RSL than $\widehat{\theta}^{H_N}$ when the prior is normal or Cauchy. Recall, however, that $\widehat{\theta}^{H_C}$ is more expensive to compute than the others.

Note that, despite its *ad hoc* nature, $\widehat{\theta}^{T_*}$ seems to be reasonable; indeed, for large p we founded to be quite satisfactory (unreported study). However, $\widehat{\theta}^{T_*}$ is not very good when the dimension is small. Moreover, as previously mentioned, it is difficult to provide error estimates for $\widehat{\theta}^{T_*}$ at the component level, since it has no posterior variance.

In this article the variance, σ^2, of the observations was assumed known. If σ^2 is unknown, one typically has available an estimate $\hat{\sigma}^2$, independent of Y. One, then, has only to add a level of integration (integration with respect to $h(\hat{\sigma}^2 \mid \sigma^2)\pi_3(\sigma^2)$, where $h(\hat{\sigma}^2 \mid \sigma^2)$ is the likelihood of σ^2 and π_3 is a prior on σ^2) to all expressions in the article. The approximation of simply replacing σ^2 by $\hat{\sigma}^2$ in all expressions is probably satisfactory when the number of degrees of freedom is large or for estimation purpose alone, but is less satisfactory for calculation of the posterior variance.

ACKNOWLEDGMENTS

The authors are grateful to Professor Herman Rubin for many valuable suggestions.

APPENDICES

A. Proofs of theorems and lemmas

Proof of Theorem 1

Let $f_Y(\cdot)$ denote the density of $N(0,\sigma^2)$, $\pi(\cdot)$ the density of $C(0,A)$, and φ_Y and φ_μ be their characteristic functions.

Proof of equation 7:

$$\begin{aligned}
m(y_j \mid \mu, A) &= \int_{-\infty}^{\infty} f_Y(\theta_j - y_j)\pi(\mu-\theta_j)d\theta_j \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{-iz(\mu-y_j)\}\varphi_Y(z)\varphi_\mu(z)dz \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left\{-\frac{\sigma^2 z^2}{2} - A|z| + i(y_j-\mu)z\right\}dz \\
&= \frac{1}{\pi}\int_{0}^{\infty} \cos[(y_j-\mu)z]\exp\left\{-\frac{\sigma^2 z^2}{2} - Az\right\}dz \\
&= \frac{\sqrt{2}}{\pi\sigma}\Re\phi^*(t_j).
\end{aligned}$$

Proof of equation 8:

The posterior mean of θ_j, $\widehat{\theta}_{j|\mu,A}$, is given by

$$\begin{aligned}
\widehat{\theta}_{j|\mu,A} &= \int_{-\infty}^{\infty} \theta_j \pi(\theta_j \mid \mu, A, y_j)d\theta_j \\
&= \frac{1}{m(y_j \mid \mu, A)}\int_{-\infty}^{\infty} \theta_j f_Y(\theta_j - y_j)\pi(\mu-\theta_j)d\theta_j \\
&= \frac{1}{m(y_j \mid \mu, A)}\left[\int_{-\infty}^{\infty} \eta f_Y(\eta)\pi([\mu-y_j]-\eta)d\eta \right.\\
&\qquad \left. + \, y_j\int_{-\infty}^{\infty} f_Y(\eta)\pi([\mu-y_j]-\eta)d\eta\right] \\
&= y_j + \frac{1}{m(y_j \mid \mu, A)}\int_{-\infty}^{\infty} \eta_j f_Y(\eta)\pi([\mu-y_j]-\eta)d\eta. \qquad \text{(A.1)}
\end{aligned}$$

Using the same technique as in the proof of eq. 7, we have

$$\int_{-\infty}^{\infty} \eta f_Y(\eta)\pi([\mu-y_j]-\eta)d\eta$$

$$= \frac{i\sigma^2}{2\pi}\int_{-\infty}^{\infty} z\exp\{-iz(\mu - y_j)\}\phi_Y(z)\phi_\mu(z)dz$$
$$= -\frac{\sigma^2}{\pi}\int_0^{\infty} z\sin[(y_j-\mu)z]\exp\left\{-\frac{\sigma^2 z^2}{2} - Az\right\}dz$$
$$= \frac{2}{\pi}\Im t_j\phi^*(t_j)$$
$$= \frac{\sqrt{2}}{\pi\sigma}[(\mu - y_j)\Re\phi^*(t_j) + A\Im\phi^*(t_j)].$$

Using this in eq. A.1, we obtain

$$\begin{aligned}\widehat{\theta}_{j|\mu,A} &= y_j + \frac{1}{\Re\phi^*(t_j)}[(\mu - y_j)\Re\phi^*(t_j) + A\Im\phi^*(t_j)]\\ &= \mu + A\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)}.\end{aligned}$$

Proof of equation 9:

By definition, the posterior variance of θ_j is given by

$$\begin{aligned}V_{j,j|\mu,A} &= \int_{-\infty}^{\infty}(\theta_j - \widehat{\theta}_{j|\mu,A})^2\pi(\theta_j \mid \mu, A, y_j)d\theta_j\\ &= \frac{1}{m(y_j \mid \mu, A)}\int_{-\infty}^{\infty}\theta_j^2 f_Y(\theta_j - y_j)\pi(\mu - \theta_j)d\theta_j - (\widehat{\theta}_{j|\mu,A})^2.\end{aligned}$$

Using the same technique as before, one can show that the last integral is equal to

$$\begin{aligned}&\int_{-\infty}^{\infty}\theta_j^2 f_Y(\theta_j - y_j)\pi(\mu - \theta_j)d\theta_j\\ &\quad= \int_{-\infty}^{\infty}\eta^2 f_Y(\eta)\pi([\mu - y_j] - \eta)d\eta + 2y_j\int_{-\infty}^{\infty}\eta f_Y(\eta)\pi([\mu - y_j] - \eta)d\eta\\ &\qquad+ y_j^2\int_{-\infty}^{\infty}f_Y(\eta)\pi([\mu - y_j] - \eta)d\eta\\ &\quad= \frac{\sqrt{2}}{\pi\sigma}\left[\left\{\frac{A\sigma}{\sqrt{2}} + [(y_j-\mu)^2 - A^2]\Re\phi^*(t_j) - 2A(y_j - \mu)\Im\phi^*(t_j)\right\}\right.\\ &\qquad\left. + 2y_j\{(\mu - y_j)\Re\phi^*(t_j) + A\Im\phi^*(t_j)\} + y_j^2\Re\phi^*(t_j)\right]\\ &\quad= \frac{\sqrt{2}}{\pi\sigma}\left[\frac{A\sigma}{\sqrt{2}} + (\mu^2 - A^2)\Re\phi^*(t_j) + 2A\mu\Im\phi^*(t_j)\right].\end{aligned}$$

Hence the posterior variance is given by

$$V_{j,j|\mu,A} = \frac{1}{\Re\phi^*(t_j)}\left[\frac{A\sigma}{\sqrt{2}} + (\mu^2 - A^2)\Re\phi^*(t_j) + 2A\mu\Im\phi^*(t_j)\right]$$

$$- \left(\mu + A\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)}\right)^2$$

$$= \frac{A\sigma}{\sqrt{2}}\frac{1}{\Re\phi^*(t_j)} - A^2\left[1 + \left(\frac{\Im\phi^*(t_j)}{\Re\phi^*(t_j)}\right)^2\right]$$

Q.E.D.

Proof of Proposition 1

Note from eq. 7 that

$$\rho^* = \int_0^\infty \int_{-\infty}^\infty A^m|\mu|^n \left\{\prod_{k=1}^p \frac{1}{2\sqrt{\pi}} \int_{-\infty}^\infty \frac{A}{[A^2 + (\theta_k - \mu)^2]} \exp(-\frac{1}{2\sigma^2}(y_k - \theta_k)^2) d\theta_k\right\} d\mu dA$$

Part 1. The integral for $1 \leq A < \infty$.

Claim: $\frac{1}{1 + (x - \xi)^2} \leq \frac{(1 + |x|)^2}{(1 + \xi^2)}$ (A.2)

Proof of the claim:

This can be rewritten as

$$[2 + |x|]\,\xi^2 - 2(1 + |x|)^2\xi + \left[|x|(1 + |x|)^2 + 2 + |x|\right] \geq 0$$

for $x > 0$ and

$$[2 + |x|]\,\xi^2 + 2(1 + |x|)^2\xi + \left[|x|(1 + |x|)^2 + 2 + |x|\right] \geq 0$$

for $x < 0$. It is easy to check that, as a polynomial in ξ, this has only imaginary roots, from which the claim is immediate.

Continuing with part 1, note that (A.2) yields

$$\frac{1}{A^2 + (\theta_k - \mu)^2} = \frac{1}{A^2\left(1 + [\frac{\theta_k}{A} - \frac{\mu}{A}]^2\right)}$$

$$\leq \frac{(1 + |\theta_k|/A)^2}{A^2\left(1 + [\mu/A]^2\right)}$$

$$\leq \frac{(1 + |\theta_k|)^2}{A^2 + \mu^2}$$

(using $A \geq 1$ here). Furthermore

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(1+|\theta_k|)^2\exp(-\frac{1}{2\sigma^2}(y_k-\theta_k)^2)d\theta_k \leq 2(1+\sigma^2+y_k^2),$$

so that

$$\int_1^{\infty}\int_{-\infty}^{\infty}A^m|\mu|^n\left\{\prod_{k=1}^{p}\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty}\frac{A}{[A^2+(\theta_k-\mu)^2]}\exp(-\frac{1}{2\sigma^2}(y_k-\theta_k)^2)d\theta_k\right\}d\mu dA$$
$$\leq (2\sigma^2)^{p/2}\left[\prod_{k=1}^{p}(1+\sigma^2+y_k^2)\right]\int_1^{\infty}\int_{-\infty}^{\infty}\frac{A^{m+p}|\mu|^n}{(A^2+\mu^2)^p}d\mu dA.$$

Now

$$\int_1^{\infty}\int_{-\infty}^{\infty}\frac{A^{m+p}|\mu|^n}{(A^2+\mu^2)^p}d\mu dA = 2\int_1^{\infty}\int_0^{\infty}\frac{A^{n+m-p}\,[\mu/A]^n}{(1+[\mu/A]^2)^p}d\mu dA$$
$$= 2\left(\int_1^{\infty}A^{n+m+1-p}dA\right)\left(\int_0^{\infty}\frac{\xi^2}{(1+\xi^2)^p}d\xi\right).$$

These integrals are finite if $p > (n+1)/2$ and $p > n+m+2$, which will be true for $p \geq 5$.

Part 2. The integral for $0 < A \leq 1$.

Note first that, by changing variables,

$$\int_{-\infty}^{\infty}\frac{A}{[A^2+(\theta_k-\mu)^2]}\exp(-\frac{1}{2\sigma^2}(y_k-\theta_k)^2)d\theta_k$$
$$= \int_{-\infty}^{\infty}\frac{A}{[A^2+x_k^2]}\exp(-\frac{1}{2\sigma^2}(\mu-[y_k-x_k])^2)dx_k.$$

Thus

$$\prod_{k=1}^{p}\int_{-\infty}^{\infty}\frac{A}{[A^2+(\theta_k-\mu)^2]}\exp(-\frac{1}{2\sigma^2}(y_k-\theta_k)^2)d\theta_k$$
$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\prod_{k=1}^{p}\frac{A}{[A^2+x_k^2]}\right\}\exp(-\frac{1}{2\sigma^2}\sum_{k=1}^{p}(\mu-[y_k-x_k])^2)dx_1\cdots dx_p$$
$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\prod_{k=1}^{p}\frac{A}{[A^2+x_k^2]}\right\}\exp(-\frac{p}{2\sigma^2}(\mu-[\bar{y}-\bar{x}])^2)\exp(-\frac{1}{2\sigma^2}S^2)dx_1\cdots dx_p, \quad (A.3)$$

where $S^2 = \sum_{k=1}^{p}([y_k-x_k]-[\bar{y}-\bar{x}])^2$.

Note next that, for $0 \le n \le 2$,

$$\int_{-\infty}^{\infty} |\mu|^n \exp(-\frac{p}{2\sigma^2}(\mu - [\bar{y} - \bar{x}])^2)d\mu \le k_1 + k_2|\bar{y} - \bar{x}|^n$$

$$\le k_3 + k_4 \sum_{k=1}^{p} x_k^2 \qquad (A.4)$$

(since $\bar{y}$ is just a constant).

Using A.3 and A.4 and freely interchanging integrals (since the integrand for ρ^* is nonnegative) yields

$$\int_0^1 \int_{-\infty}^{\infty} A^m |\mu|^n \left\{ \prod_{k=1}^{p} \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{A}{[A^2 + (\theta_k - \mu)^2]} \exp(-\frac{1}{2\sigma^2}(\theta_k - \mu)^2)d\theta_k \right\} d\mu dA$$
$$\le \int_0^1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} A^m \left(k_3 + k_4 \sum_{k=1}^{p} x_k^2 \right) \left\{ \prod_{k=1}^{p} \frac{A}{[A^2 + x_k^2]} \right\} \exp(-\frac{1}{2}S^2)dx_1 \ldots dx_p dA.$$

Now

$$k_3 \int_0^1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} A^m \left\{ \prod_{k=1}^{p} \frac{A}{[A^2 + x_k^2]} \right\} \exp(-\frac{1}{2}S^2)dx_1 \ldots dx_p dA$$
$$\le k_3 \int_0^1 A^m \left\{ \prod_{k=1}^{p} \int_{-\infty}^{\infty} \frac{A}{[A^2 + x_k^2]} dx_k \right\} dA$$
$$\le k_3' \int_0^1 A^m dA$$
$$< \infty.$$

Finally

$$k_4 \int_0^1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} A^m \left(\sum_{k=1}^{p} x_k^2 \right) \left\{ \prod_{k=1}^{p} \frac{A}{[A^2 + x_k^2]} \right\} \exp(-\frac{1}{2}S^2)dx_1 \ldots dx_p dA$$
$$\le k_4 \sum_{k=1}^{p} \int_0^1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} A^{m+1} \left\{ \prod_{j \ne k} \frac{A}{[A^2 + x_j^2]} \right\} \exp(-\frac{1}{2}S^2)dx_1 \ldots dx_p dA.$$

and since $\int_{-\infty}^{\infty} \exp(-\frac{1}{2}S^2)dx_k < k_5$, an identical argument verifies that this integral also is finite. This completes the proof. Q.E.D.

Proof of lemma 1

Claim 1:

$$y_k^2 \Re\phi^*(t_k) \le \frac{\pi\sigma}{\sqrt{2}} \left(A + \frac{[\mu^2 + 1]}{A} \right).$$

Proof of claim 1: Using eq. 7 and making a change of variables gives

$$y_k^2 \Re\phi^*(t_k) = \frac{\pi\sigma}{\sqrt{2}} E^Z \left[\frac{A y_k^2}{A^2 + (Z + y_k - \mu)^2} \right],$$

where Z is $N(0,1)$. Maximization over y_k yields

$$\frac{A y_k^2}{A^2 + (z + y_k - \mu)^2} \leq A + \frac{(z-\mu)^2}{A} \tag{A.5}$$

so that

$$y_k^2 \Re\phi^*(t_k) \leq \frac{\pi\sigma}{\sqrt{2}} \mathrm{E}^Z \left[A + \frac{(Z-\mu)^2}{A} \right],$$

from which the claim follows.

Claim 2:

$$\lim_{|y_l| \to \infty} y_l^2 \int_0^\infty \int_{-\infty}^\infty \left\{ \prod_{k=1}^p \Re\phi^*(t_k) \right\} \pi_2(\mu, A) d\mu dA$$
$$= \frac{\pi\sigma}{\sqrt{2}} \int_0^\infty \int_{-\infty}^\infty A \left\{ \prod_{k \neq l} \Re\phi^*(t_k) \right\} \pi_2(\mu, A) d\mu dA.$$

Proof of claim 2: Note that

$$\begin{aligned} \lim_{|y_l| \to \infty} y_l^2 \Re\phi^*(t_l) &= \lim_{|y_l| \to \infty} \frac{\pi\sigma}{\sqrt{2}} E^Z \left[\frac{A y_l^2}{A^2 + (Z + y_l - \mu)^2} \right] \\ &= \frac{\pi\sigma}{\sqrt{2}} A, \end{aligned} \tag{A.6}$$

by the Dominated Convergence Theorem (using A.5).

Claim 1 and A.2 imply that the Dominated Convergence Theorem can also be applied to the sequence $\{y_l^2 \Re\phi^*(t_l)\}$, so that

$$\lim_{|y_l| \to \infty} \int_0^\infty \int_{-\infty}^\infty \left\{ \prod_{k \neq l} \Re\phi^*(t_k) \right\} \left[y_l^2 \Re\phi^*(t_l) \right] \pi_2(\mu, A) d\mu dA$$
$$= \int_0^\infty \int_{-\infty}^\infty \left\{ \prod_{k \neq l} \Re\phi^*(t_k) \right\} \left[\lim_{|y_l| \to \infty} y_l^2 \Re\phi^*(t_l) \right] \pi_2(\mu, A) d\mu dA$$
$$= \frac{\pi\sigma}{\sqrt{2}} \int_0^\infty \int_{-\infty}^\infty A \left\{ \prod_{k \neq l} \Re\phi^*(t_k) \right\} \pi_2(\mu, A) d\mu dA.$$

Claim 3: Eq. 15 holds.

Proof of claim 3: This follows directly from A.6 and Claim 2, noting that $|y_l| \to \infty$ is equivalent to $d_l \to \infty$ (by translation) and that

$$\pi_2(\mu, A \mid y) = \frac{\{\prod_{k=1}^{p} \Re\phi^*(t_k)\}\, \pi_2(\mu, A)}{\int_0^\infty \int_{-\infty}^\infty \{\prod_{k=1}^{p} \Re\phi^*(t_k)\}\, \pi_2(\mu, A) d\mu dA}$$
$$= \frac{\left\{\prod_{k\neq l} \Re\phi^*(t_k)\right\} \left[y_l^2 \Re\phi^*(t_l)\right] \pi_2(\mu, A)}{y_l^2 \int_0^\infty \int_{-\infty}^\infty \{\prod_{k=1}^{p} \Re\phi^*(t_k)\}\, \pi_2(\mu, A) d\mu dA}.$$

This complete the proof of the lemma. Q.E.D.

Proof of Proposition 2

Let $y_{(1)}, \ldots, y_{(r)}$ the distinct values of the order statistics for $y_1, \ldots, y_p$. (Equality of the y_i is allowed, as long as $r \geq 2$.) Define

$$z_0 = -\infty, \quad z_r = +\infty, \quad z_i = \frac{1}{2}(y_{(i)} + y_{(i+1)}) \text{ for } i = 1, \ldots, r-1.$$

Clearly

$$(4\pi)^{-\frac{p}{2}} m(y) = \int_0^\infty (\sigma^2 + A^2)^{-\frac{p}{2}} \left\{ \sum_{i=0}^{r} \int_{z_i}^{z_{i+1}} \prod_{k=1}^{p} \left[\frac{1 - e^{-\|y_k\|^2}}{\|y_k\|^2} \right] d\mu \right\} \pi_{2,2}(A) dA.$$

It is easy to show that $(1 - e^{-x})/x$ is a decreasing function of x, with a maximum value of 1. Hence, in the interval (z_i, z_{i+1}),

$$\prod_{k=1}^{p} \left[\frac{1 - e^{-\|y_k\|^2}}{\|y_k\|^2} \right] \leq \frac{1 - e^{-\|y^*\|^2}}{\|y^*\|^2}$$
$$\leq \frac{\sigma^2 + A^2}{(y^* - \mu)^2},$$

where y^* is any one of the $y_{(j)}$ other than $y_{(i+1)}$. (It is here that we use the condition that the y_i are not all equal.) Since the interval (z_i, z_{i+1}) does not contain y^*, it follows that

$$\int_{z_i}^{z_{i+1}} \prod_{k=1}^{p} \left[\frac{1 - e^{-\|y_k\|^2}}{\|y_k\|^2} \right] d\mu \leq (\sigma^2 + A^2) \int_{z_i}^{z_{i+1}} (y^* - \mu)^{-2} d\mu$$
$$= k_i(\sigma^2 + A^2) \quad (k_i < \infty).$$

Thus

$$(4\pi)^{-\frac{p}{2}}m(y)=\int_0^\infty(\sigma^2+A^2)^{-\frac{p}{2}}A\sum_{i=0}^{r}k_i(\sigma^2+A^2)\pi_{2,2}(A)dA,$$

which is clearly finite if $p\geq 5$ for $\pi_{2,2}(A)=A$, and if $p\geq 4$ for $\pi_{2,2}(A)=1$. This complete the proof for $m(y)$.

The proofs that the $\widehat{\theta}_j^{GS}$ and the $V_{j,k}^{GS}$ are finite are identical, using the easy to prove facts that, for all μ and A

$$|\widehat{\theta}_{j|\mu A}^{GS}-y_j|\leq k(\sigma^2+A^2)^{-\frac{1}{2}}$$

and

$$|V_{j,j|\mu A}^{GS}-\sigma^2|\leq k(\sigma^2+A^2)^{-\frac{1}{2}}.$$ Q.E.D.

Proof of Theorem 4

Partitioning the real line into $(-\infty,\mu-\sqrt{\sigma^2+A^2})$, $[\mu-\sqrt{\sigma^2+A^2},\mu+\sqrt{\sigma^2+A^2}]$ and $(\mu+\sqrt{\sigma^2+A^2},\infty)$, one can show that

$$\frac{\left[1-e^{-\|y_l\|^2}\right]y_l^2}{\|y_l\|^2}\leq 2\left[\mu^2+\sigma^2+A^2\right].$$

Also,

$$\lim_{|y_l|\to\infty}\frac{\left[1-e^{-\|y_l\|^2}\right]y_l^2}{\|y_l\|^2}=\sigma^2+A^2.$$

Thus, for any y_l, the function $\left[y_l^2\prod_{k=1}^{p}m(y_k\mid\mu,A)\right]$ is bounded by

$$\left\{\prod_{k\neq l}m(y_k\mid\mu,A)\right\}\frac{\left[\mu^2+\sigma^2+A^2\right]}{\sqrt{\pi}\sqrt{\sigma^2+A^2}},$$

which is integrable by eq. 22 in Theorem 4. Hence, the Dominated Convergence Theorem yields eq. 23 in Theorem 4.

The limiting results involving the $\widehat{\theta}_j^{GS}$ are proved in the same way, using the fact that

$$|\widehat{\theta}_{j|\mu A}^{GS}-y_j|\leq k(\sigma^2+A^2)^{-\frac{1}{2}}.$$ Q.E.D.

B. Calculation of $\phi^*(t)$

As defined in chapter 2, $\phi^*(t)$ is given by

$$\phi^*(t) = e^{t^2}\left[\frac{\sqrt{\pi}}{2} - \int_0^t e^{-z^2} dz\right], \tag{B.1}$$

where $t = a - ib$.

Using Wall (1948) and Jones and Thron (1980), $\phi^*(t)$ can be evaluated using the following expressions:

$$\begin{aligned}
\phi^*(t) &= \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{2}t+}\frac{1}{\sqrt{2}t+}\frac{2}{\sqrt{2}t+}\frac{3}{\sqrt{2}t+}\cdots\right] && \text{(B.2)}\\
&= \frac{\sqrt{\pi}}{2}e^{t^2} - \left[\frac{t}{1-}\frac{t^2}{3/2+}\frac{t^2}{5/2-}\frac{(3/2)t^2}{7/2+}\frac{2t^2}{9/2-}\cdots\right] && \text{(B.3)}\\
&= \frac{\sqrt{\pi}}{2}e^{t^2} - t\left[\frac{1}{1+}\frac{(it)^2}{3/2-}\frac{(it)^2}{5/2+}\frac{(3/2)(it)^2}{7/2-}\frac{2(it)^2}{9/2+}\cdots\right] && \text{(B.4)}
\end{aligned}$$

Based on some empirical evidence, $\phi^*(t)$ was computed according to the following scheme:

- if $a \geq 1$, use equation B.2;
- if $a < 1$ and $|b| < 1$, use equation B.3;
- if $a < 1$ and $|b| \geq 1$, use equation B.4.

Using this scheme, the average number of terms needed to reach a precision of 1×10^{-5} in $|\phi^*(t)|$ was 7.14 and it cost less than \$0.02 of computer time for each value of $\phi^*(t)$. Using IMSL subroutines to compute $\phi^*(t)$ to the same precision costs around \$0.06 per value computed. Since the function $\phi^*(t)$ is involved in their inner integrals of equation 6, it pays to use the complex continued fractions.

References

[1] Berger, J. (1980), "A Robust Generalized Bayes Estimator and Confidence Region for a Multivariate Normal Mean", The Annals of Statistics, 8, 716-761.

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

[3] Berger, J. and Dey, D.K. (1985), "Truncation of Shrinkage Estimators of Normals Means in the Nonsymmetric Case", Multivariate Analysis VI, P.R. Krishnaiah (Ed.), North-Holland, Amsterdam.

[4] Box, G.E.P. and Tiao, G. C. (1968), "A Bayesian Approach to some Outlier Problems", Biometrika, 51, 153-167.

[5] Box, G.E.P. and Tiao, G. C. (1973), Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Massachusetts.

[6] Dawid, A.P. (1973), "Posterior Expectations for Large Observations", Biometrika 60, 664-667.

[7] Dey, D.K. and Berger, J. (1983), "On Truncation of Shrinkage Estimators in Simultaneous Estimation of Normal Means", Journal of the American Statistical Association, 78, 865-869.

[8] Efron, B. and Morris, C. (1971), "Limiting the Risk of Bayes and Empirical Bayes – Part I: The Bayes Case", Journal of the American Statistical Association, 66, 807-815.

[9] Gaver D.P., (1985), "Discrepancy-Tolerant Hierarchical Poisson Event-Rate Analyses", Technical Report, Operations Research Dept., Naval Postgraduate School, Monterey.

[10] Gaver D.P. and O'Muircheartaigh, I.G. (1987), "Robust Empirical Bayes Analyses of Event Rates", Technometrics, 29, 1-15.

[11] Jones, W.B. and Thron, W.J. (1980), Continued Fractions: Analytic Theory and Applications, Encyclopedia of Mathematics and its Applications, Vol 11, Addison-Wesley, Reading, Massachusetts.

[12] Lindley, D.V. (1971), "The Estimation of Many Parameters", Foundations of Statistical Inference, V.P. Godambe and D.A. Sprott (Eds.), Holt, Rinehart and Winston, Toronto.

[13] Lindley D.V. and Smith, A.F.M. (1972) "Bayes Estimates for the Linear Model", Journal of the Royal Statistical Society, B, 34, 1-41.

[14] Magnus, W., Oberhettinger, F. and Soni, R.P. (1966), Formulas and Theorems for the Special Functions of Mathematical Physics, Springer-Verlag, New York.

[15] Morris, C. (1983), "Parametric Empirical Bayes Inference: Theory and Applications", Journal of the American Statistical Association, 78, 47-65.

[16] O'Hagan, A. (1979), "On Outlier Rejection Phenomena in Bayes Inference", Journal of the Royal Statistical Society B41, 358-367.

[17] Stein, C. (1981), "Estimation of the Parameters of a Multivariate Normal Distribution: I. Estimation of the Means", The Annals of Statistics, 9, 1135-1151.

[18] Wall, H.S. (1948), Analytic Theory of Continued Fractions, D. Van Nostrand Co., New York.

Dép. de Mathématiques et d'Informatique

Université de Sherbrooke

Sherbrooke (Québec) J1K 2R1

Statistics Department

Purdue University

West Lafayette, IN 47906

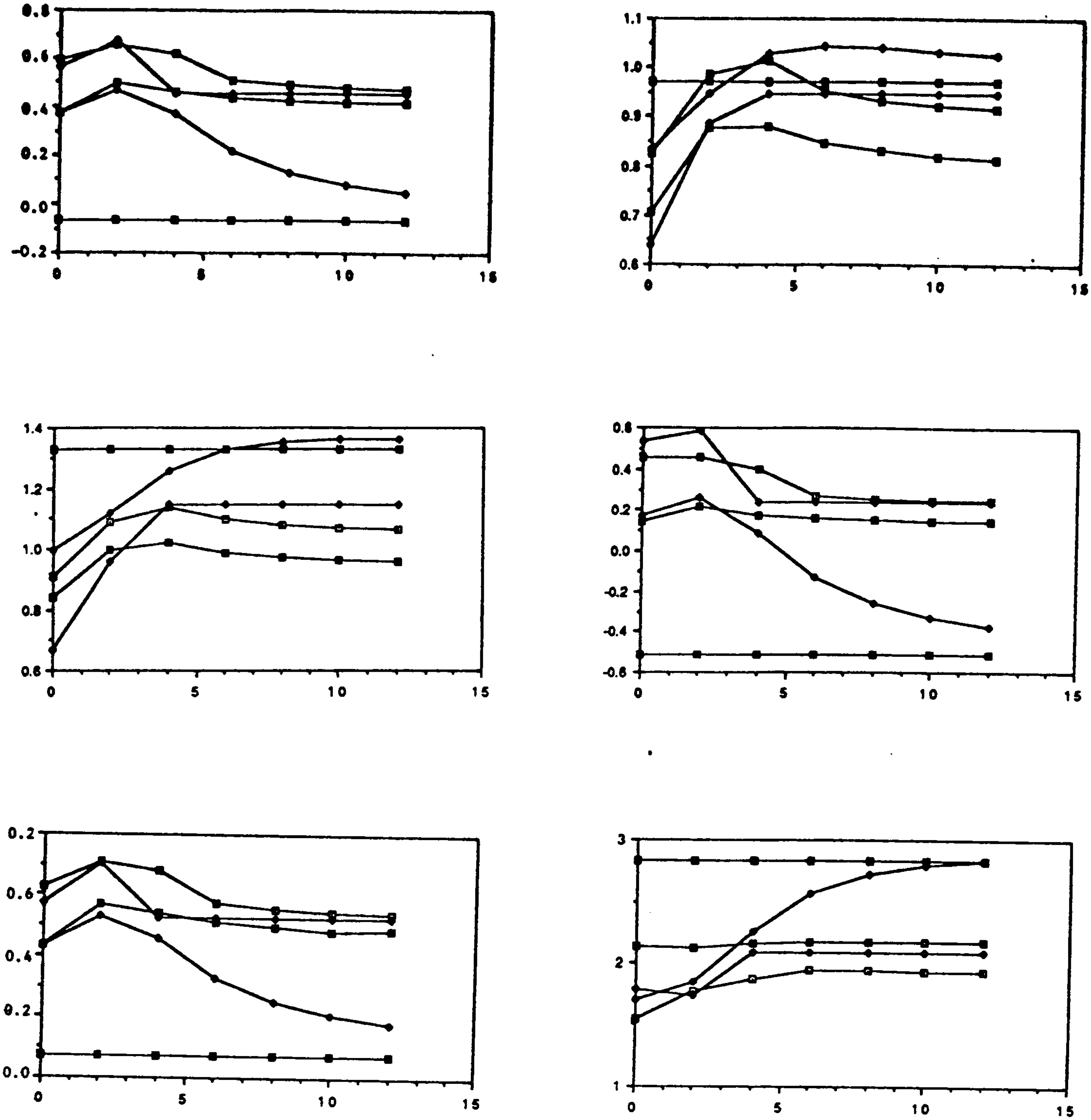

Figure 1 : Comparison of the behavior of $\hat{\theta}_j^{H_C}$ (dotted square), $\hat{\theta}_j^{H_N}$ (full diamond), $\hat{\theta}_j^{GS}$ (square), $\hat{\theta}_j^{T*}$ (diamond) and y_j (full square) when the sample contains an outlier (for j = 1, ...6).

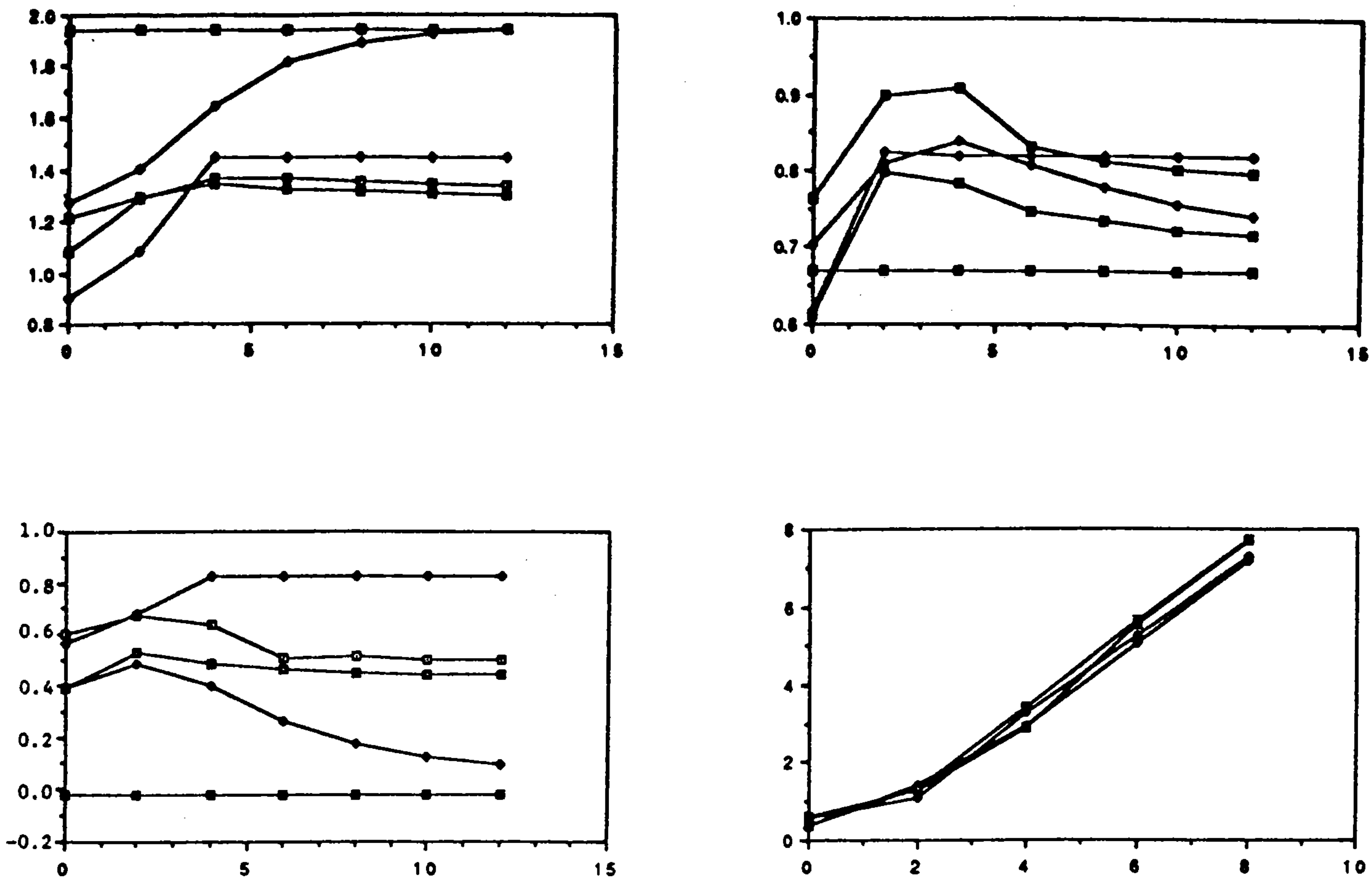

Figure 2 : Comparison of the behavior of $\hat{\theta}_j^{H_C}$ (dotted square), $\hat{\theta}_j^{H_N}$ (full diamond), $\hat{\theta}_j^{G_S}$ (square), $\hat{\theta}_j^{T*}$ (diamond) and y_j (full square) when the sample contains an outlier (for j = 7, ...10).

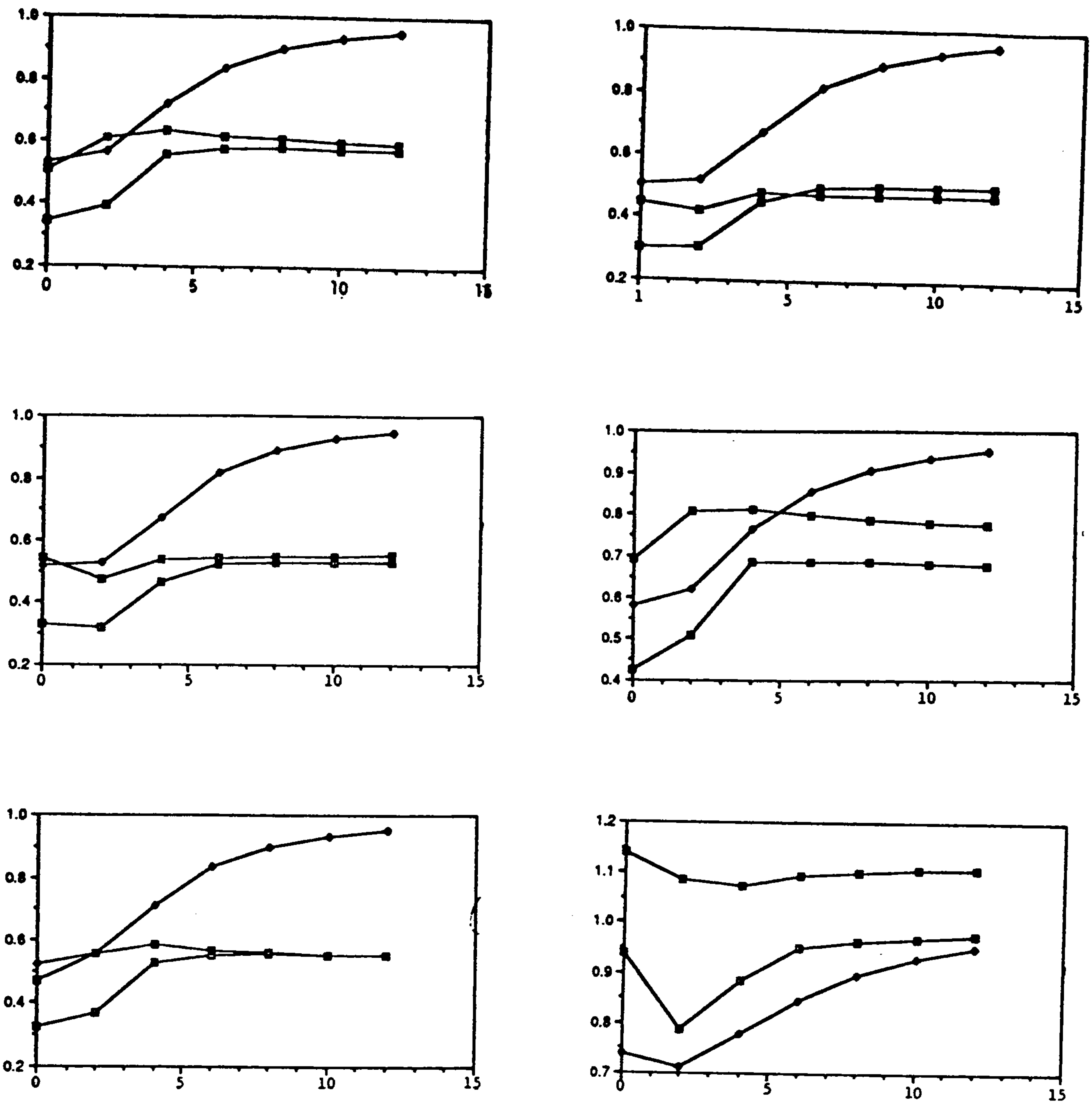

Figure 3 : Comparison of the posterior variance of $\hat{\theta}^{HC}$ (dotted square), $\hat{\theta}^{HN}$ (full diamond) and $\hat{\theta}^{GS}$ (square) when the sample contains an outlier (for j = 1, ...6).

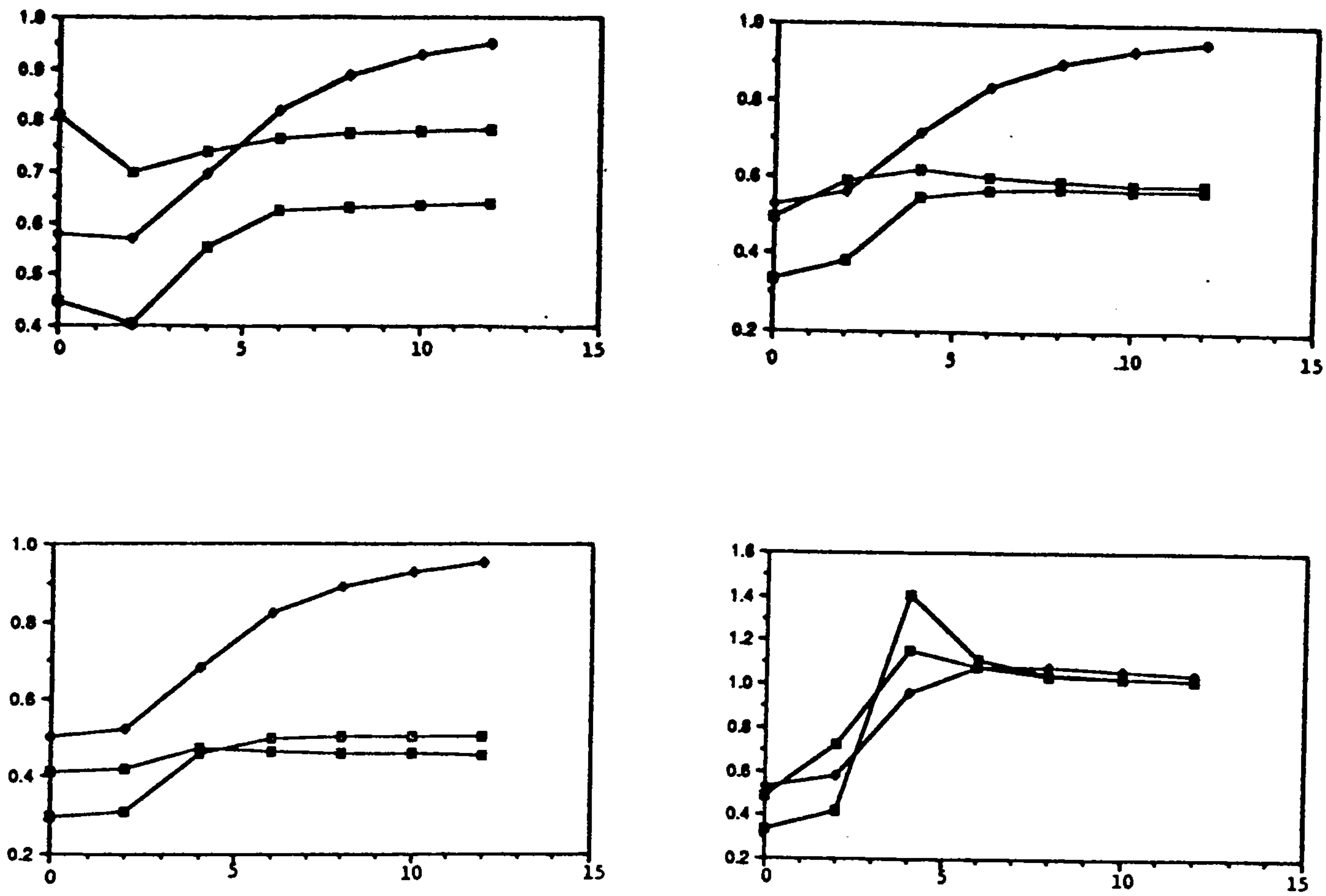

Figure 4 : Comparison of the posterior variance of $\hat{\theta}^{HC}$ (dotted square), $\hat{\theta}^{HN}$ (full diamond) and $\hat{\theta}^{GS}$ (square) when the sample contains an outlier (for j = 7, ...10).

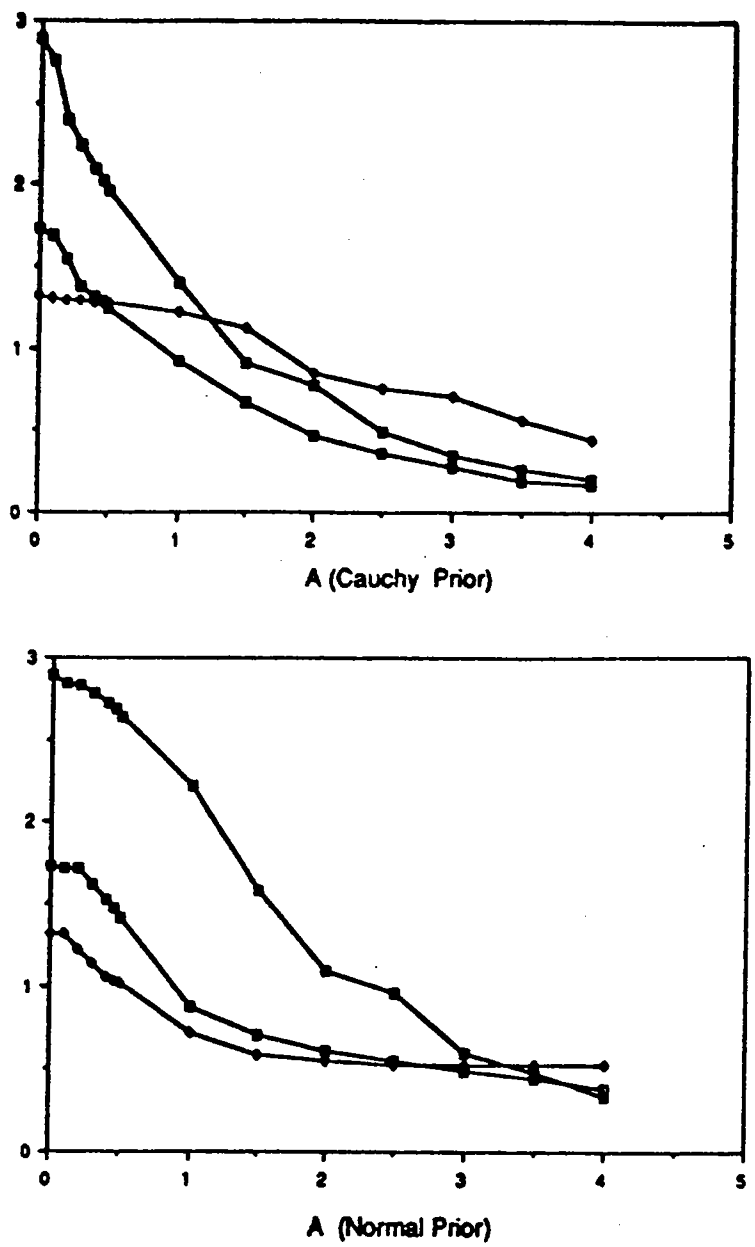

Figure 5: Difference in the Bayes risk of $\hat{\theta}^{H_C}$, $\hat{\theta}^{H_N}$, $\hat{\theta}^{T*}$ and the Bayes rule (p=5).

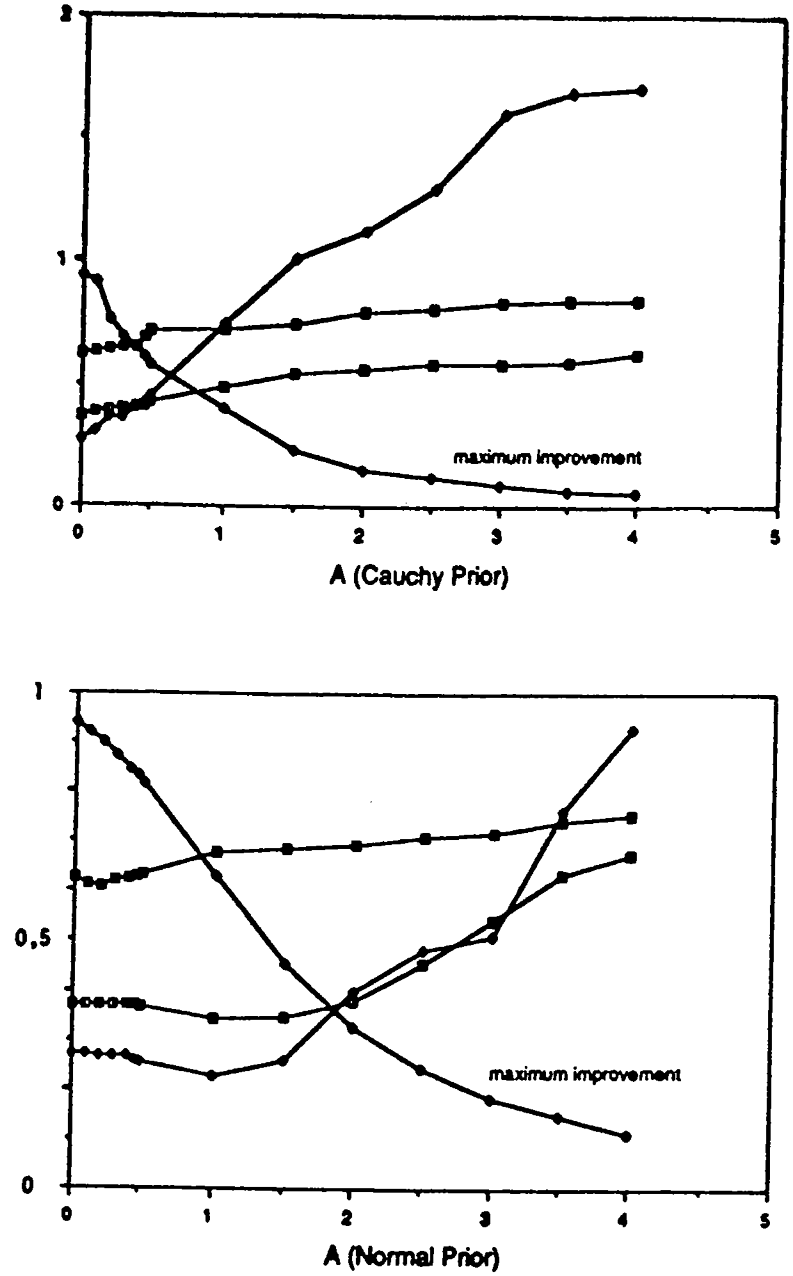

Figure 6: RSL Comparison between $\hat{\theta}^{H_C}$, $\hat{\theta}^{H_N}$ and $\hat{\theta}^{T\cdot}$ (p=5)