The Annals of Statistics
1987, Vol. 15, No. 2, 817–829

SEQUENTIAL SHRINKAGE ESTIMATION

BY MALAY GHOSH,[1] DAVID M. NICKERSON[1] AND PRANAB K. SEN[2]

University of Florida, University of Georgia and University of North Carolina at Chapel Hill

The paper develops a class of James–Stein estimators that dominates the sample mean under sequential sampling schemes of Ghosh, Sinha and Mukhopadhyay (1976). Asymptotic risk expansions of the sample mean and James–Stein estimators are provided up to the second-order term. Also, a Monte Carlo study is undertaken to compare the risks of these estimators.

1. Introduction. Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a sequence of iid $N(\boldsymbol{\theta}, \sigma^2\mathbf{V})$ variables, where $\boldsymbol{\theta} \in R^p$ is unknown, and $\mathbf{V}$ (positive definite) is known. Based on $\mathbf{X}_1, \dots, \mathbf{X}_n$, if the estimator $\boldsymbol{\delta}_n = \boldsymbol{\delta}_n(\mathbf{X}_1, \dots, \mathbf{X}_n)$ is used for estimating $\boldsymbol{\theta}$, suppose that the loss incurred is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\delta}_n) = (\boldsymbol{\delta}_n - \boldsymbol{\theta})'\mathbf{Q}(\boldsymbol{\delta}_n - \boldsymbol{\theta}) + cn, \tag{1.1}$$

where $\mathbf{Q}$ denotes a known positive definite matrix, and c (> 0) denotes the known cost per unit sample. In particular, using $\overline{\mathbf{X}}_n = n^{-1}\sum_{i=1}^n \mathbf{X}_i$ as an estimator of $\boldsymbol{\theta}$, the expected loss (risk) is given by

$$R(\boldsymbol{\theta}, \sigma^2, \overline{\mathbf{X}}_n) = n^{-1}\sigma^2 \operatorname{tr}(\mathbf{QV}) + cn. \tag{1.2}$$

If σ^2 is known, the above risk is minimized at $n = n^* = (\operatorname{tr}\mathbf{QV}/c)^{1/2}\sigma$. For simplicity, we shall, henceforth assume n^* to be a positive integer. However, for unknown σ^2, there does not exist any fixed sample size that minimizes (1.2) simultaneously for all σ^2. In this case, motivated by the optimal fixed sample size n^* (when σ^2 is known), the following sequential procedure is proposed for determining the sample size:

$$N = \inf\left\{n \geq m \colon n \geq \left(\tilde{s}_n^2/c\right)^{1/2}(\operatorname{tr}\mathbf{QV})^{1/2}\right\}, \tag{1.3}$$

where m (≥ 2) denotes the initial sample size and for every $n \geq 2$,

$$\tilde{s}_n^2 = \{(n-1)p\}^{-1}\sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}}_n)'\mathbf{V}^{-1}(\mathbf{X}_i - \overline{\mathbf{X}}_n). \tag{1.4}$$

Note that for *known* σ^2, the fixed sample rule with sample size n^* and the corresponding estimator $\overline{\mathbf{X}}_{n^*}$ is also a sequential minimax rule under the loss (1.1) [see, e.g., Kiefer (1957)]. Thus, (1.3) can be viewed as an empirical minimax stopping rule.

Received May 1985; revised September 1986.
[1]Research partially supported by NSF Grant DMS-8218091 and partially supported by NSF Grant DMS-8600666.
[2]Research partially supported by ONR Contract N00014-83-K-0387.
AMS 1980 *subject classifications.* 62J07, 62L12.
Key words and phrases. Multivariate, sequential estimation, James–Stein estimators, asymptotic risk expansions, Monte Carlo.

Ghosh, Sinha and Mukhopadhyay (1976) and Woodroofe (1977) proposed similar stopping rules in the more general case when the $\mathbf{X}_i$'s have an unknown dispersion matrix $\boldsymbol{\Sigma}$. Such stopping rules were multivariate analogs of a univariate procedure introduced in Robbins (1959). In all these papers, the estimator $\overline{\mathbf{X}}_N$ was proposed for $\boldsymbol{\theta}$, and asymptotic (as $c \to 0$) properties of N and $\overline{\mathbf{X}}_N$ were studied.

It is known, however, that for *fixed sample* sizes, James–Stein estimators dominate the sample mean under the loss (1.1). This is shown in James and Stein (1961) when $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$ (the identity matrix of order p) and σ^2 is known, in Berger (1976) when σ^2 is known and in Efron and Morris (1976) when $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$ but σ^2 is unknown. For a general class of estimators dominating the sample mean when the sample size is fixed, one may refer to Baranchik (1970), Strawderman (1971), Efron and Morris (1976) and Berger (1976).

Ghosh and Sen (1983) developed James–Stein estimators dominating the sample mean under two-stage sampling schemes. However, a two-stage procedure needs on an average more observations than n^*, the optimal fixed sample size if σ^2 were known.

Takada (1984) developed some sequential James–Stein estimators when $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$. However, as we shall see in Section 4, at the nth stage of the experiment, Takada's estimator uses only p out of the $(n-1)p$ degrees of freedom available for estimating σ^2, and thereby throws away some information.

The present paper removes both the objections raised in the preceding two paragraphs. First, it uses all the available degrees of freedom for estimating σ^2 at each stage of the experiment. Second, unlike a two-stage procedure, which employs s_m^2 in defining the stopping rule, the stopping rule (1.3) uses an updated estimate of the variance σ^2 at each stage of the experiment, and thereby demands less observations on an average than the corresponding two-stage procedure.

In the remainder of this section, we summarize the main results of this paper. The proofs are deferred to the subsequent sections. Consider the class of James–Stein estimators $\boldsymbol{\delta}_N^b(\mathbf{X}_1, \dots, \mathbf{X}_N)$, where

$$\begin{aligned}\boldsymbol{\delta}_n^b(\mathbf{X}_1, \dots, \mathbf{X}_n) = \overline{\mathbf{X}}_n - &\frac{b s_n^2}{n(\overline{\mathbf{X}}_n - \boldsymbol{\lambda})' \mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\overline{\mathbf{X}}_n - \boldsymbol{\lambda})} \\ &\times \mathbf{Q}^{-1}\mathbf{V}^{-1}(\overline{\mathbf{X}}_n - \boldsymbol{\lambda}),\end{aligned} \tag{1.5}$$

for every $n \ge 2$, $s_n^2 = p(p+2)^{-1}\tilde{s}_n^2$ $(n \ge 2)$, and $\boldsymbol{\lambda} \in R^p$ is the known point towards which we want to shrink. Very often $\boldsymbol{\lambda}$ is taken as the prior mean. For fixed samples, this estimator was proposed by Hudson (1974) and Berger (1976). The first main result of this paper is as follows.

THEOREM 1. *Under the stopping rule* (1.3), *and the loss* (1.1), $R(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\delta}_N^b) < R(\boldsymbol{\theta}, \sigma^2, \overline{\mathbf{X}}_N)$ *for every* $b \in (0, 2(p-2))$.

For fixed sample sizes, the optimal choice of b is $(p^2 - 4)p^{-1}$. Note that $(p^2-4)p^{-1}s_n^2 = (p-2)\tilde{s}_n^2$. Customarily, the numerator of the second term in the right-hand side of (1.5) is expressed as a multiple of $\tilde{s}_n^2$. We find it convenient to work with s_n^2 instead of $\tilde{s}_n^2$ for technical reasons. The important point is that unlike the fixed sample case, the optimal b in the sequential case depends on unknown parameters. The next two theorems provide asymptotic (as $c \to 0$) risk expansions of $\overline{\mathbf{X}}_N$ and $\boldsymbol{\delta}_N^b$. These results suggest that at least asymptotically, the optimal choice of b is indeed $(p^2 - 4)p^{-1}$.

Theorem 2. *Under the stopping rule* (1.3), *and the loss* (1.1),

$$R(\boldsymbol{\theta}, \sigma^2, \overline{\mathbf{X}}_N) = 2c^{1/2}\sigma(\operatorname{tr}\mathbf{QV})^{1/2} + c(2p)^{-1} + o(c), \tag{1.6}$$

as $c \to 0$ when $m \geq 2$.

Theorem 3. *Under the stopping rule* (1.3), *and the loss* (1.1), *as $c \to 0$,*

$$\begin{aligned} R(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\delta}_N^b) &= 2c^{1/2}\sigma(\operatorname{tr}\mathbf{QV})^{1/2} + c(2p)^{-1} \\ &\quad - cbp^2(p+2)^{-2}\big(2(p^2-4)p^{-1} - b\big)\sigma^2 \\ &\quad \times (\operatorname{tr}\mathbf{QV})^{-1}\{(\boldsymbol{\theta} - \boldsymbol{\lambda})'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\lambda})\}^{-1} + o(c), \end{aligned} \tag{1.7}$$

for $\boldsymbol{\theta} \neq \boldsymbol{\lambda}$ when $m \geq 3$;

$$\begin{aligned} R(\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\delta}_N^b) &= 2c^{1/2}\sigma(\operatorname{tr}\mathbf{QV})^{1/2} - c^{1/2}\sigma b p^2(p+2)^{-2}\big(2(p^2-4)p^{-1} - b\big) \\ &\quad (\operatorname{tr}\mathbf{QV})^{-1/2} E\left(\sum_{i=1}^{p} W_i a_i^{-1}\right)^{-1} + o(c^{1/2}), \end{aligned} \tag{1.8}$$

for $m \geq 2$, where the W_i's are iid χ_1^2 and the a_i's are the eigenvalues of $\mathbf{QV}$.

From (1.6)–(1.8), the asymptotic percentage risk improvement of $\boldsymbol{\delta}_N^b$ over $\overline{\mathbf{X}}_N$ is given by

$$\begin{aligned} &100\left\{\tfrac{1}{2}bp^2(p+2)^{-2}\big(2(p^2-4)p^{-1} - b\big)\sigma\right\} \\ &\quad \times (\operatorname{tr}\mathbf{QV})^{-3/2}\{(\boldsymbol{\theta} - \boldsymbol{\lambda})'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\lambda})\}^{-1} c^{1/2} + o(c^{1/2}), \end{aligned} \tag{1.9}$$

for $\boldsymbol{\theta} \neq \boldsymbol{\lambda}$, $0 < b < 2(p-2)$ and $m \geq 3$, while for $\boldsymbol{\theta} = \boldsymbol{\lambda}$, $0 < b < 2(p-2)$ and $m \geq 2$, the asymptotic percentage risk improvement is given by

$$\begin{aligned} &100\left\{\tfrac{1}{2}(\operatorname{tr}\mathbf{QV})^{-1}\right\} bp^2(p+2)^{-2}\big(2(p^2-4)p^{-1} - b\big) \\ &\quad \times E\left(\sum_{i=1}^{p} W_i a_i^{-1}\right)^{-1} + o(1). \end{aligned} \tag{1.10}$$

It is now easy to see that the dominant term in both (1.9) and (1.10) is maximized when $b = (p^2 - 4)p^{-1}$.

The layout of this paper is as follows. The proof of Theorem 1 is given in Section 2. The proof of Theorem 2 is omitted as it can be developed along the lines of Woodroofe (1977), Ghosh and Mukhopadhyay (1980) and Finster (1983).

The details appear in Ghosh, Nickerson and Sen (1985). The proof of Theorem 3 is briefly outlined in Section 3. Monte Carlo simulations comparing the risks of the sample mean, the proposed James–Stein estimators and Takada's estimators are given in Section 4. As anticipated, our estimators achieve bigger risk reduction than Takada's estimators. Finally, in Section 5, certain concluding remarks are made.

2. Proof of Theorem 1. First write

$$\begin{aligned} &R(\boldsymbol{\theta}, \sigma^2, \delta_N^b) - R(\boldsymbol{\theta}, \sigma^2, \overline{\mathbf{X}}_N) \\ &\quad = -2bE_{\boldsymbol{\theta}, \sigma^2}\left[\frac{s_N^2(\overline{\mathbf{X}}_N - \boldsymbol{\lambda})'\mathbf{V}^{-1}(\overline{\mathbf{X}}_N - \boldsymbol{\theta})}{N(\overline{\mathbf{X}}_N - \boldsymbol{\lambda})'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\overline{\mathbf{X}}_N - \boldsymbol{\lambda})}\right] \\ &\qquad + b^2E_{\boldsymbol{\theta}, \sigma^2}\left[s_N^4/\left(N^2\{(\overline{\mathbf{X}}_N - \boldsymbol{\lambda})'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\overline{\mathbf{X}}_N - \boldsymbol{\lambda})\}\right)\right]. \end{aligned} \tag{2.1}$$

Since $\mathbf{Q}$ and $\mathbf{V}$ are both positive definite, using the simultaneous diagonalization theorem, there exists a nonsingular $\mathbf{D}$ such that $\mathbf{D}\mathbf{Q}^{-1}\mathbf{D}' = \mathbf{I}_p$ and $\mathbf{D}\mathbf{V}\mathbf{D}' = \mathbf{A} = \mathrm{Diag}(a_1, \dots, a_p)$ with $a_i > 0$, $1 \le i \le p$. Use the transformation $\mathbf{Z}_i = \mathbf{D}(\mathbf{X}_i - \boldsymbol{\lambda})$, $i = 1, 2, \dots,$ so that the $\mathbf{Z}_i$'s are iid $N(\boldsymbol{\zeta}, \sigma^2\mathbf{A})$ with $\boldsymbol{\zeta} = \mathbf{D}(\boldsymbol{\theta} - \boldsymbol{\lambda})$. Write $\overline{\mathbf{Z}}_n = n^{-1}\sum_{i=1}^n \mathbf{Z}_i$, $n \ge 1$, and use the Helmert orthogonal transformation $\mathbf{Y}_2 = (\mathbf{Z}_1 - \mathbf{Z}_2)/\sqrt{2}$, $\mathbf{Y}_3 = (\mathbf{Z}_1 + \mathbf{Z}_2 - 2\mathbf{Z}_3)/\sqrt{6}, \dots$. Then, $\mathbf{Y}_i$'s are iid $N(\mathbf{0}, \sigma^2\mathbf{A})$ and $\overline{\mathbf{Z}}_n$ is distributed independently of $(\mathbf{Y}_2, \dots, \mathbf{Y}_n)$ for every $n \ge 2$.

Also, one has

$$(\overline{\mathbf{X}}_n - \boldsymbol{\lambda})'\mathbf{V}^{-1}(\overline{\mathbf{X}}_n - \boldsymbol{\theta}) = \overline{\mathbf{Z}}_n'\mathbf{A}^{-1}(\overline{\mathbf{Z}}_n - \boldsymbol{\zeta}), \qquad n \ge 1, \tag{2.2}$$

$$(\overline{\mathbf{X}}_n - \boldsymbol{\lambda})'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\overline{\mathbf{X}}_n - \boldsymbol{\lambda}) = \overline{\mathbf{Z}}_n'\mathbf{A}^{-2}\overline{\mathbf{Z}}_n, \qquad n \ge 1, \tag{2.3}$$

$$s_n^2 = \{(n-1)(p+2)\}^{-1}\sum_{i=2}^n \mathbf{Y}_i'\mathbf{A}^{-1}\mathbf{Y}_i, \qquad n \ge 2. \tag{2.4}$$

Then from (2.1), one gets

$$\begin{aligned} &R(\boldsymbol{\theta}, \sigma^2, \delta_N^b) - R(\boldsymbol{\theta}, \sigma^2, \overline{\mathbf{X}}_N) \\ &\quad = -2bE_{\boldsymbol{\zeta}, \sigma^2}\left[(s_N^2/N)(\overline{\mathbf{Z}}_N'\mathbf{A}^{-1}(\overline{\mathbf{Z}}_N - \boldsymbol{\zeta}))/(\overline{\mathbf{Z}}_N'\mathbf{A}^{-2}\overline{\mathbf{Z}}_N)\right] \\ &\qquad + b^2E_{\boldsymbol{\zeta}, \sigma^2}\left[(s_N^4/N^2)(\overline{\mathbf{Z}}_N'\mathbf{A}^{-2}\overline{\mathbf{Z}}_N)^{-1}\right]. \end{aligned} \tag{2.5}$$

Let $\mathscr{B}_n$ denote the σ-algebra generated by $\mathbf{Y}_2, \dots, \mathbf{Y}_n$. Then

$$\begin{aligned} &E_{\boldsymbol{\zeta}, \sigma^2}\left[(s_N^2/N)(\overline{\mathbf{Z}}_N'\mathbf{A}^{-1}(\overline{\mathbf{Z}}_N - \boldsymbol{\zeta}))/(\overline{\mathbf{Z}}_N'\mathbf{A}^{-2}\overline{\mathbf{Z}}_N)\right] \\ &\quad = \sum_{n=m}^{\infty} E_{\boldsymbol{\zeta}, \sigma^2}\left[n^{-1}s_n^2 I_{[N=n]} E_{\boldsymbol{\zeta}, \sigma^2}\{(\overline{\mathbf{Z}}_n'\mathbf{A}^{-1}(\overline{\mathbf{Z}}_n - \boldsymbol{\zeta}))/(\overline{\mathbf{Z}}_n'\mathbf{A}^{-2}\overline{\mathbf{Z}}_n) | \mathscr{B}_n\}\right] \\ &\quad = \sum_{n=m}^{\infty} E_{\boldsymbol{\zeta}, \sigma^2}\left[n^{-1}s_n^2 I_{[N=n]}\sum_{i=1}^p a_i^{-1}(\sigma^2/n)a_i E_{\boldsymbol{\zeta}, \sigma^2}\left\{\frac{\partial}{\partial \overline{Z}_{ni}}\left(\frac{\overline{Z}_{ni}}{\overline{\mathbf{Z}}_n'\mathbf{A}^{-2}\overline{\mathbf{Z}}_n}\right)\middle| \mathscr{B}_n\right\}\right] \\ &\quad = \sigma^2\sum_{n=m}^{\infty} E_{\boldsymbol{\zeta}, \sigma^2}\left[n^{-2}s_n^2 I_{[N=n]}(p-2)/(\overline{\mathbf{Z}}_n'\mathbf{A}^{-2}\overline{\mathbf{Z}}_n)\right]. \end{aligned} \tag{2.6}$$

For the first equality in (2.6), one uses the independence of $\bar{\mathbf{Z}}_n$ and $(\mathbf{Y}_2, \dots, \mathbf{Y}_n)$, while for the second equality, one uses Stein's identity [cf. Stein (1981)].

Next, note that for every n, $n\bar{\mathbf{Z}}_n'\mathbf{A}^{-2}\bar{\mathbf{Z}}_n/\sigma^2 \sim \sum_{i=1}^p W_{ni}/a_i$, where the W_{ni}'s are independent and $W_{ni} \sim \chi_1^2(n\zeta_i^2/2\sigma^2 a_i)$, $i = 1, \dots, p$. Hence, using the result [see, e.g., Cressie, Davis, Folks and Policello (1981)] that if $P(U > 0) = 1$, then $E(U^{-1}) = \int_0^\infty E(\exp(-tU))\,dt$, one gets

$$\begin{aligned} &E_{\zeta,\sigma^2}\left[n\bar{\mathbf{Z}}_n'\mathbf{A}^{-2}\bar{\mathbf{Z}}_n/\sigma^2\right]^{-1} \\ &\quad= \int_0^\infty \prod_{i=1}^p E\left[\exp(-tW_{ni}/a_i)\right] dt \\ &\quad= \int_0^\infty \prod_{i=1}^p \left\{\left(1 + 2ta_i^{-1}\right)^{-1/2} \exp\left(-nt\zeta_i^2/\left(\sigma^2 a_i(a_i + 2t)\right)\right)\right\} dt \\ &\quad= g_p(\Delta_n), \end{aligned} \tag{2.7}$$

where $\Delta_n = \Delta_n(\zeta, \sigma^2)$. In this section we need only that $g_p(\Delta_n)$ is nonincreasing in n. Now, using again the independence of $\bar{\mathbf{Z}}_n$ and $(\mathbf{Y}_2, \dots, \mathbf{Y}_n)$, one gets from (2.5)–(2.7),

$$\begin{aligned} &R(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\delta}_N^b) - R(\boldsymbol{\theta}, \sigma^2, \bar{\mathbf{X}}_N) \\ &\quad= -b \sum_{n=m}^\infty g_p(\Delta_n) E_{\sigma^2}\left[(s_n^2/n)\left(2(p-2) - bs_n^2\sigma^{-2}\right) I_{[N=n]}\right] \\ &\quad< -2b(p-2) \sum_{n=m}^\infty g_p(\Delta_n) E_{\sigma^2}\left[n^{-1}s_n^2\left(1 - s_n^2\sigma^{-2}\right) I_{[N=n]}\right], \end{aligned} \tag{2.8}$$

where in the last step one uses $0 < b < 2(p-2)$. Also, in the above, E_{σ^2} denotes expectation when the $\mathbf{Y}_i$'s are iid $N(\mathbf{0}, \sigma^2\mathbf{A})$. Accordingly, for proving the theorem, it suffices to show that

$$\sum_{n=m}^\infty g_p(\Delta_n) E_{\sigma^2}\left[n^{-1}s_n^2\left(1 - s_n^2/\sigma^2\right) I_{[N=n]}\right] \geq 0, \quad \text{for all } \sigma^2 > 0. \tag{2.9}$$

To prove (2.9), first observe that $\mathrm{tr}(\mathbf{QV}) = \mathrm{tr}(\mathbf{QD}^{-1}\mathbf{AD}'^{-1}) = \mathrm{tr}(\mathbf{A})$, since $\mathbf{DQ}^{-1}\mathbf{D}' = \mathbf{I}_p$. Let n_0 denote the *smallest* integer $\geq m$ such that $p(p+2)^{-1}cn_0^2/(\mathrm{tr}\,\mathbf{A}) \geq \sigma^2$. Then we write $p(p+2)^{-1}cn_0^2/(\mathrm{tr}\,\mathbf{A}) \geq \sigma^2$. Then we write

$$\begin{aligned} \text{lhs of (2.9)} = &\sum_{n=m}^{n_0-1} g_p(\Delta_n) E_{\sigma^2}\left[n^{-1}s_n^2\left(1 - s_n^2\sigma^{-2}\right) I_{[N=n]}\right] \\ &+ g_p(\Delta_{n_0}) E_{\sigma^2}\left[n_0^{-1}s_{n_0}^2\left(1 - s_{n_0}^2\sigma^{-2}\right) I_{[N \geq n_0]}\right] \\ &+ \sum_{n=n_0}^\infty \Big\{g_p(\Delta_{n+1}) E_{\sigma^2}\left[(n+1)^{-1}s_{n+1}^2\left(1 - s_{n+1}^2\sigma^{-2}\right) I_{[N \geq n+1]}\right] \\ &\qquad - g_p(\Delta_n) E_{\sigma^2}\left[n^{-1}s_n^2\left(1 - s_n^2\sigma^{-2}\right) I_{[N \geq n+1]}\right]\Big\}, \end{aligned} \tag{2.10}$$

where the first term in the rhs of (2.10) should be interpreted as zero if $n_0 = m$. The crux of the argument for proving (2.9) is as follows. Use the break up as given in (2.10). Since, $g_p(\Delta_n) \searrow$ in n, by definition, on the set $[N \geq n+1]$, $s_n^2 \geq p(p+2)^{-1}cn^2/(\operatorname{tr}\mathbf{A}) \geq \sigma^2$ when $n \geq n_0$. Thus, one gets

$$\text{third term in the rhs of (2.10)}$$

$$\geq \sum_{n=n_0}^{\infty} g_p(\Delta_{n+1})E_{\sigma^2}\Big[\big\{(n+1)^{-1}s_{n+1}^2\big(1-s_{n+1}^2\sigma^{-2}\big) - n^{-1}s_n^2\big(1-s_n^2\sigma^{-2}\big)\big\}I_{[N\geq n+1]}\Big]. \tag{2.11}$$

Use the representation $s_{n+1}^2 = ((n-1)s_n^2 + U_{n+1}^2)/n$, where

$$U_{n+1} = (p+2)^{-1}\mathbf{Y}'_{n+1}\mathbf{A}^{-1}\mathbf{Y}_{n+1} \sim \sigma^2\chi_p^2/(p+2)$$

independently of $\mathbf{Y}_2, \dots, \mathbf{Y}_n$, and prove by direct computations

$$\text{rhs of (2.11)} \geq 0. \tag{2.12}$$

Next use induction arguments to show that

$$\text{sum of the first two terms in the rhs of (2.10)} \geq 0. \tag{2.13}$$

To prove (2.12), note that $I_{[N\geq n+1]}$ is a $\mathscr{B}_n$ measurable function. Then, with probability 1,

$$\begin{aligned}
&E_{\sigma^2}\Big[(n+1)^{-1}s_{n+1}^2\big(1-s_{n+1}^2\sigma^{-2}\big)I_{[N\geq n+1]}|\mathscr{B}_n\Big]\\
&\quad= I_{[N\geq n+1]}\Big[n^{-1}s_n^2\big(1-s_n^2\sigma^{-2}\big) + \big\{n^{-1}-(n-1)^2(n+1)^{-1}n^{-1}\big\}s_n^4\sigma^{-2}\\
&\qquad\quad + s_n^2\big\{(n-1)(n+1)^{-1}n^{-1}\\
&\qquad\qquad -2(n-1)pn^{-2}(n+1)^{-1}(p+2)^{-1} - n^{-1}\big\}\\
&\qquad\quad + p\sigma^2(p+2)^{-1}\big\{(n+1)^{-1}n^{-1}-(n+1)^{-1}n^{-2}\big\}\Big]\\
&\quad= I_{[N\geq n+1]}n^{-1}s_n^2\big(1-s_n^2\sigma^{-2}\big)\\
&\qquad + I_{[N\geq n+1]}\left[\frac{s_n^4}{\sigma^2}\frac{3n-1}{n^2(n+1)} - 2\frac{s_n^2}{n^2(n+1)}\left(n+\frac{(n-1)p}{p+2}\right)\right.\\
&\qquad\qquad \left. + \frac{p\sigma^2}{p+2}\frac{n-1}{n^2(n+1)}\right].
\end{aligned} \tag{2.14}$$

Note that the multiple of $I_{[N\geq n+1]}$ in the extreme rhs of (2.14) is a convex function of s_n^2, where the minimum occurs at

$$s_n^2 = \sigma^2\Big[\big\{n+(n-1)p(p+2)^{-1}\big\}/(3n-1)\Big] < \sigma^2. \tag{2.15}$$

Hence, recalling that on the set $[N \geq n+1]$, $s_n^2 \geq \sigma^2$, it follows that

$$\begin{aligned} &\text{second term in the extreme right of (2.14)} \\ &\geq I_{[N\geq n+1]}\Big[(3n-1)\sigma^2 n^{-2}(n+1)^{-1} \\ &\qquad -2\sigma^2\big(n+(n-1)p(p+2)^{-1}\big)n^{-2}(n+1)^{-1} \\ &\qquad +\sigma^2(n-1)p(p+2)^{-1}n^{-2}(n+1)^{-1}\Big] \\ &= I_{[N\geq n+1]}\sigma^2 n^{-2}(n+1)^{-1}(n-1)\big(1-p(p+2)^{-1}\big) > 0. \end{aligned} \tag{2.16}$$

From (2.11), (2.14) and (2.16), one gets (2.12).

Next we prove (2.13). For $n_0 = m$, since $I_{[N\geq m]} = 1$ with probability 1 and

$$\begin{aligned} &E_{\sigma^2}\big[s_m^2(1-s_m^2\sigma^{-2})\big] \\ &\quad= \sigma^2\Big[p(p+2)^{-1} - ((m-1)p+2)p(m-1)^{-1}(p+2)^{-2}\Big] \\ &\quad= 2(m-2)p\sigma^2(m-1)^{-1}(p+2)^{-2} \geq 0, \end{aligned}$$

the result follows. For $n_0 > m (\geq 2)$, first note that for $n \leq n_0 - 1$, on the set $[N = n]$, $s_n^2 \leq p(p+2)^{-1}cn^2/(\operatorname{tr}\mathbf{A}) < \sigma^2$ so that using $g_p(\Delta_n) \searrow$ in n,

$$\begin{aligned} &\text{first term in the rhs of (2.10)} \\ &\geq g_p(\Delta_{n_0}) \sum_{n=m}^{n_0-1} E_{\sigma^2}\big[n^{-1}s_n^2(1-s_n^2\sigma^{-2})I_{[N=n]}\big]. \end{aligned} \tag{2.17}$$

Hence, (2.13) follows if

$$\begin{aligned} &\sum_{n=m}^{n_0-1} E_{\sigma^2}\big[n^{-1}s_n^2(1-s_n^2\sigma^{-2})I_{[N=n]}\big] \\ &\qquad + E_{\sigma^2}\Big[n_0^{-1}s_{n_0}^2\big(1-s_{n_0}^2\sigma^{-2}\big)I_{[N\geq n_0]}\Big] \geq 0 \end{aligned} \tag{2.18}$$

holds true. To prove (2.18), first use $s_{n_0}^2 = ((n_0-2)s_{n_0-1}^2 + U_{n_0})/(n_0-1)$, and get with probability 1,

$$\begin{aligned} &E_{\sigma^2}\Big[n_0^{-1}s_{n_0}^2\big(1-s_{n_0}^2\sigma^{-2}\big)I_{[N\geq n_0]}|\mathscr{B}_{n_0-1}\Big] \\ &\quad= \Big[\Big\{(n_0-2)s_{n_0-1}^2 + p(p+2)^{-1}\sigma^2\Big\}n_0^{-1}(n_0-1)^{-1} \\ &\qquad -\Big\{(n_0-2)^2 s_{n_0-1}^4 + 2(n_0-2)s_{n_0-1}^2 p\sigma^2(p+2)^{-1} \\ &\qquad +p\sigma^4(p+2)^{-1}\Big\}n_0^{-1}(n_0-1)^{-2}\sigma^{-2}\Big]I_{[N\geq n_0]} \\ &\quad= \Big\{a_{n_0}s_{n_0-1}^2 - b_{n_0}s_{n_0-1}^4\sigma^{-2} + \sigma^2 c_{n_0}\Big\}I_{[N\geq n_0]}, \end{aligned} \tag{2.19}$$

where

$$
\begin{aligned}
a_{n_0} &= (n_0 - 2)\left(n_0 - 1 - 2p(p+2)^{-1}\right)n_0^{-1}(n_0 - 1)^{-2}, \\
b_{n_0} &= (n_0 - 2)^2 n_0^{-1}(n_0 - 1)^{-2}, \qquad (2.20)\\
c_{n_0} &= (n_0 - 2)p n_0^{-1}(n_0 - 1)^{-2}(p+2)^{-1}.
\end{aligned}
$$

Let

$$g_{n_0} = b_{n_0} - a_{n_0} = (p-2)(n_0 - 2)n_0^{-1}(n_0 - 1)^{-2}(p+2)^{-1} \quad (> 0),$$

so that

$$c_{n_0} - g_{n_0} = 2(n_0 - 2)n_0^{-1}(n_0 - 1)^{-2}(p+2)^{-1} > 0.$$

Also, let

$$d_{n_0} = 2b_{n_0} - a_{n_0} = (n_0 - 2)\left(n_0 - 3 + 2p(p+2)^{-1}\right)n_0^{-1}(n_0 - 1)^{-2},$$

so that $(n_0 - 1)d_{n_0} \in (0, 1)$.

Now rewrite

extreme right of (2.19)

$$
\begin{aligned}
&= E_{\sigma^2}\left[\left\{d_{n_0}\left(s_{n_0-1}^2 - s_{n_0-1}^4\sigma^{-2}\right) + g_{n_0}\left(s_{n_0-1}^4\sigma^{-2} - 2s_{n_0-1}^2 + \sigma^2\right)\right\}\right] \\
&= E_{\sigma^2}\Big[\Big\{d_{n_0}\left(s_{n_0-1}^2 - s_{n_0-1}^4\sigma^{-2}\right) \qquad (2.21)\\
&\qquad + g_{n_0}\left(s_{n_0-1}^4\sigma^{-2} - 2s_{n_0-1}^2 + \sigma^2\right) + \sigma^2(c_{n_0} - g_{n_0})\Big\}I_{[N \geq n_0]}\Big] \\
&\geq d_{n_0}E_{\sigma^2}\left[\left(s_{n_0-1}^2 - s_{n_0-1}^4\sigma^{-2}\right)I_{[N \geq n_0]}\right].
\end{aligned}
$$

If the extreme right of (2.21) ≥ 0, noting again that $s_n^2 < \sigma^2$ on the set $[N = n]$ for all $n \leq n_0 - 1$, one proves (2.18) from (2.19) and (2.21). Otherwise, noting that $(n_0 - 1)d_{n_0} < 1$, one gets from (2.18), (2.19) and (2.21),

lhs of (2.18)

$$
\begin{aligned}
&\geq \sum_{n=m}^{n_0-2} E_{\sigma^2}\Big[n^{-1}s_n^2\left(1 - s_n^2\sigma^{-2}\right)I_{[N=n]} \qquad (2.22)\\
&\qquad + (n_0 - 1)^{-1}s_{n_0-1}^2\left(1 - s_{n_0-1}^2\sigma^{-2}\right)I_{[N \geq n_0-1]}\Big].
\end{aligned}
$$

Proceed inductively to get either lhs of (2.18) ≥ 0, or finally end with

$$\text{lhs of (2.18)} \geq E_{\sigma^2}\left[m^{-1}s_m^2\left(1 - s_m^2\sigma^{-2}\right)I_{[N \geq m]}\right] \geq 0, \qquad (2.23)$$

as calculated earlier. The proof of Theorem 1 is now complete. □

3. Proof of Theorem 3. From (2.5)–(2.7), one gets

$$
\begin{aligned}
c^{-1}&\left[R\left(\theta,\sigma^2,\delta_N^b\right)-R\left(\theta,\sigma^2,\overline{\mathbf{X}}_N\right)\right]\\
&=c^{-1}E_{\sigma^2}\Bigg[\left(-2b(p-2)s_N^2+b^2\sigma^{-2}s_N^4\right)N^{-1}\\
&\qquad\times\int_0^\infty\prod_{i=1}^p\left(1+2ta_i^{-1}\right)^{-1/2}\exp\left(-Nt\sigma^{-2}\sum_{i=1}^p\zeta_i^2a_i^{-1}\left(a_i+2t\right)^{-1}\right)dt\Bigg]\\
&=E_{\sigma^2}\Bigg[\left(-2b(p-2)s_N^2+b^2\sigma^{-2}s_N^4\right)\left(Nc^{1/2}\right)^{-1}\int_0^\infty\prod_{i=1}^p\left(1+2tc^{1/2}a_i^{-1}\right)^{-1/2}\\
&\qquad\times\exp\left(-Ntc^{1/2}\sigma^{-2}\sum_{i=1}^p\zeta_i^2a_i^{-1}\left(a_i+2tc^{1/2}\right)^{-1}\right)dt\Bigg].
\end{aligned}\tag{3.1}
$$

As $c\to 0$, $s_N^2\to p(p+2)^{-1}\sigma^2$ a.s. and $Nc^{1/2}\to(\operatorname{tr}\mathbf{QV})^{1/2}\sigma=(\operatorname{tr}\mathbf{A})^{1/2}\sigma$ a.s. Hence, for $\zeta\neq\mathbf{0}$ (i.e., $\theta\neq\lambda$) as $c\to 0$,

$$
\begin{aligned}
&\text{the bracketed term in the rhs of (3.1)}\\
&\quad\to -\sigma^2p^2(p+2)^{-2}b\left\{2\left(p^2-4\right)p^{-1}-b\right\}(\operatorname{tr}\mathbf{QV})^{-1/2}\sigma^{-1}\\
&\qquad\times\int_0^\infty\exp\left(-t\sigma^{-1}(\operatorname{tr}\mathbf{QV})^{1/2}\left(\sum_{i=1}^p\zeta_i^2/a_i^2\right)\right)dt\\
&\quad=-\sigma^2p^2(p+2)^{-2}b\left\{2\left(p^2-4\right)p^{-1}-b\right\}(\operatorname{tr}\mathbf{QV})^{-1}\\
&\qquad\times\left\{(\theta-\lambda)'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\theta-\lambda)\right\}^{-1}\quad\text{a.s.,}
\end{aligned}\tag{3.2}
$$

using

$$
\sum_{i=1}^p\zeta_i^2a_i^{-2}=\zeta'\mathbf{A}^{-2}\zeta=(\theta-\lambda)'\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}(\theta-\lambda).
$$

Next for $m\geq 3$, proving the uniform integrability (in $c\leq c_0$) of the bracketed term in the rhs of (3.1), and using (1.6), one gets (1.7). The details of the uniform integrability proof appear in Ghosh, Nickerson and Sen (1985), and are omitted.

For $\zeta=\mathbf{0}$, i.e., $\theta=\lambda$, it follows from (3.1) that

$$
\begin{aligned}
c^{-1/2}&\left[R\left(\theta,\sigma^2,\delta_N^b\right)-R\left(\theta,\sigma^2,\overline{\mathbf{X}}_N\right)\right]\\
&=c^{-1/2}E_{\sigma^2}\left[\left(-2b(p-2)s_N^2+b^2\sigma^{-2}s_N^4\right)N^{-1}\right]\\
&\qquad\times\int_0^\infty\prod_{i=1}^p\left(1+2ta_i^{-1}\right)^{-1/2}dt.
\end{aligned}\tag{3.3}
$$

Using the result that if W_i's are iid χ_1^2,

$$
\begin{aligned}
E\left(\sum_{i=1}^pW_ia_i^{-1}\right)^{-1}&=\int_0^\infty E\left[\exp\left(-t\sum_{i=1}^pW_ia_i^{-1}\right)\right]dt\\
&=\int_0^\infty\prod_{i=1}^p\left(1+2ta_i^{-1}\right)^{-1/2}dt,
\end{aligned}
$$

along with $s_N^2 \to p(p+2)^{-1}\sigma^2$ a.s. and $Nc^{1/2} \to (\operatorname{tr} \mathbf{QV})^{1/2}\sigma$ a.s. as $c \to 0$, it follows that

$$c^{-1/2} \text{ times the bracketed term in the rhs of (3.3)}$$

$$\to \left(-2b(p-2)p(p+2)^{-1} + b^2p^2(p+2)^{-2}\right) \tag{3.4}$$

$$\times \sigma^2(\operatorname{tr}\mathbf{QV})^{-1/2}\sigma^{-1}E\left(\sum_{i=1}^{p} W_i a_i^{-1}\right)^{-1} \quad \text{a.s.}$$

The uniform integrability property of s_n^4 (in $c \le c_0$) when $m \ge 2$ along with (1.6) now lead to (1.8). The details appear in Ghosh, Nickerson and Sen (1985) and are omitted. □

4. A Monte Carlo study. For simplicity, consider in this section the case when $\boldsymbol{\lambda} = \mathbf{0}$, $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$. In this special case, Takada (1984) has shown that if one defines

$$\eta_n = (p+2)^{-1}\mathbf{Y}_n'\mathbf{Y}_n, \qquad n \ge 2, \tag{4.1}$$

then the estimator

$$\mathbf{T}_n^b = \overline{\mathbf{X}}_n - \frac{b\eta_n}{n\overline{\mathbf{X}}_n'\overline{\mathbf{X}}_n}\overline{\mathbf{X}}_n \tag{4.2}$$

dominates $\overline{\mathbf{X}}_n$ for all $0 < b < 2(p-2)$. Note that $\eta_n \sim (p+2)^{-1}\chi_p^2$, and does not utilize all the available $(n-1)p$ degrees of freedom for estimating σ^2. In this section, our objective is to compare the risk performance of $\overline{\mathbf{X}}_N$, $\boldsymbol{\delta}_N^b$ and $\mathbf{T}_N^b$, with $b = (p^2-4)/p$ when $p = 6$ and $p = 9$. For notational simplicity, we shall call $\boldsymbol{\delta}_N^{(p^2-4)/p}$ and $\mathbf{T}_N^{(p^2-4)/p}$, $\boldsymbol{\delta}_N$ and $\mathbf{T}_N$, respectively. Also consider $\boldsymbol{\delta}_N^+$ and $\mathbf{T}_N^+$, which are plus rule versions of James–Stein estimators and Takada's estimators, i.e.,

$$\boldsymbol{\delta}_n^+ = \left(1 - (p^2-4)p^{-1}s_n^2\left(n\overline{\mathbf{X}}_n'\overline{\mathbf{X}}_n\right)^{-1}\right)^+\overline{\mathbf{X}}_n$$

and

$$\mathbf{T}_n^+ = \left(1 - (p^2-4)p^{-1}\eta_n\left(n\overline{\mathbf{X}}_n'\overline{\mathbf{X}}_n\right)^{-1}\right)^+\overline{\mathbf{X}}_n,$$

for $n \ge 2$, where $a^+ = \max(a, 0)$. Although, we have no analytical results available showing the dominance of $\boldsymbol{\delta}_N^+$ or $\mathbf{T}_N^+$ over $\overline{\mathbf{X}}_N$, it is anticipated that they will perform better than $\boldsymbol{\delta}_N$ and $\mathbf{T}_N$ by preventing overshrinking [see, e.g., Lehmann (1983), page 302]. Indeed, for fixed samples, it is known that $\boldsymbol{\delta}_n^+$ dominates $\boldsymbol{\delta}_n$ for every $n \ge 2$.

Before describing the simulation procedure, note that $R(\boldsymbol{\theta}, \sigma^2, \overline{\mathbf{X}}_N)$ does not depend on $\boldsymbol{\theta}$, while the risks of the other estimators depend on $\boldsymbol{\theta}$ only through $\|\boldsymbol{\theta}\|$.

To simulate the sequential sampling procedure, and evaluate the estimators under consideration, a large pool of $N(\boldsymbol{\theta}, \mathbf{I}_p)$ variables were generated for $p = 6$ and 9 with $\|\boldsymbol{\theta}\| = 0$, 1 and 2. $\boldsymbol{\theta}$ was taken as $(0,0,0,0,\sqrt{0.5},\sqrt{0.5})$ and $(0,0,0,0,0,0,0,\sqrt{0.5},\sqrt{0.5})$, respectively, for $p = 6$ and 9 when $\|\boldsymbol{\theta}\| = 1$, and

$(0,0,0,0,\sqrt{2},\sqrt{2})$ and $(0,0,0,0,0,0,0,\sqrt{2},\sqrt{2})$, respectively, for $p = 6$ and 9 when $\|\boldsymbol{\theta}\| = 2$. We have run other simulations as well and have noted that the actual coordinates of $\boldsymbol{\theta}$ do not matter as long as $\|\boldsymbol{\theta}\|$ remains the same. Also, c, the cost per unit sample is taken as $c = 0.50, 0.25, 0.10, 0.05, 0.025$ and 0.01.

A single experiment would be sequential samples from the pool until the stopping criterion was met. At this point, the sampling would stop, the number of samples would be recorded and the estimators $\overline{\mathbf{X}}_N$, $\boldsymbol{\delta}_N$, $\mathbf{T}_n$, $\boldsymbol{\delta}_N^+$ and $\mathbf{T}_N^+$ and their associated losses are computed.

On the completion of 1000 experiments, we compute the average losses for all these estimators, and these are the simulated versions of the corresponding risks. Also, at this point we compute the percentage risk improvements $100(R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N) - R(\boldsymbol{\theta}, \boldsymbol{\delta}_N))/R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N)$, $100(R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N) - R(\boldsymbol{\theta}, \mathbf{T}_N))/R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N)$, $100(R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N) - R(\boldsymbol{\theta}, \boldsymbol{\delta}_N^+))/R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N)$ and $100(R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N) - R(\boldsymbol{\theta}, \mathbf{T}_N^+))/R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N)$, which are denoted by $\%\boldsymbol{\delta}_N$, $\%\mathbf{T}_N$, $\%\boldsymbol{\delta}_N^+$ and $\%\mathbf{T}_N^+$, respectively.

Our simulation findings are summarized in Table 1 for $p = 6$ and Table 2 for $p = 9$. It is clear from the tables that as in the fixed sample case when $\boldsymbol{\lambda} = \mathbf{0}$, the risk improvement of all the estimators is most substantial when $\|\boldsymbol{\theta}\| = 0$, and the improvement keeps diminishing as $\|\boldsymbol{\theta}\|$ moves further and further away from zero. Also, the risk improvements are more substantial when $p = 9$ than when $p = 6$. Also, for a fixed $\|\boldsymbol{\theta}\| \neq 0$ as c decreases, i.e., the average sample size gets larger, the percentage risk improvement decreases as in the fixed sample case. The opposite is the case when $\|\boldsymbol{\theta}\| = 0$. The main reason is that when $\boldsymbol{\lambda} = \mathbf{0}$, $N\|\overline{\mathbf{X}}_N\|^2$ behaves as a multiple of $c^{-1/2}\|\boldsymbol{\theta}\|^2$ when $\boldsymbol{\theta} \neq \mathbf{O}$, while $N\|\overline{\mathbf{X}}_N\|^2 \to_L \chi_p^2$ when $\boldsymbol{\theta} = \mathbf{0}$.

TABLE 1

The risks and the percentage risk improvements over $\overline{\mathbf{X}}_N$ for $p = 6$

$\|\boldsymbol{\theta}\|$	Cost	$\overline{N}$	$R(\boldsymbol{\theta}, \overline{\mathbf{X}}_N)$	$\%\boldsymbol{\delta}_N$	$\%\mathbf{T}_N$	$\%\boldsymbol{\delta}_N^+$	$\%\mathbf{T}_N^+$
0	0.50	3.81	3.6217	26.15	22.80	30.73	26.64
	0.25	5.23	2.5185	27.74	23.34	32.98	27.85
	0.10	8.12	1.5456	30.58	22.97	35.75	30.29
	0.05	11.36	1.1074	31.35	25.39	36.56	30.80
	0.025	15.84	0.7871	32.04	25.05	37.33	31.00
	0.01	24.90	0.4956	32.41	25.24	37.83	31.42
1	0.50	3.78	3.5405	14.38	12.87	16.51	14.66
	0.25	5.23	2.4995	12.30	10.60	13.84	12.34
	0.10	8.13	1.5530	10.97	8.80	11.46	9.51
	0.05	11.34	1.0927	9.14	7.51	9.22	7.66
	0.025	15.91	0.7791	6.44	5.11	6.44	5.19
	0.01	24.94	0.4958	4.42	3.37	4.42	3.37
2	0.50	3.77	3.5919	6.35	5.71	6.38	5.77
	0.25	5.24	2.4911	4.93	4.19	4.93	4.19
	0.10	8.06	1.5486	3.51	2.73	3.51	2.73
	0.05	11.37	1.0909	2.83	2.35	2.83	2.35
	0.025	15.84	0.7769	2.00	1.62	2.00	1.62
	0.01	24.94	0.4861	1.13	0.88	1.13	0.88

TABLE 2
The risks and the percentage risk improvements over $\overline{\mathbf{X}}_N$ for $p = 9$

$\|\theta\|$	Cost	$\bar{N}$	$R(\theta, \overline{\mathbf{X}}_N)$	$\%\delta_N$	$\%T_N$	$\%\delta_N^+$	$\%T_N^+$
0	0.50	4.63	4.2456	31.95	27.68	36.78	33.03
	0.25	6.42	3.0509	35.34	29.05	38.95	34.61
	0.10	9.94	1.8992	35.88	30.52	40.03	36.18
	0.05	13.81	1.3479	36.62	31.72	41.00	36.76
	0.025	19.42	0.9552	37.42	31.02	41.46	37.11
	0.01	30.50	0.6132	38.96	32.40	42.37	37.51
1	0.50	4.61	4.3622	21.66	19.18	23.35	21.30
	0.25	6.43	3.0317	19.30	16.45	20.66	18.36
	0.10	9.96	1.8811	15.82	13.87	16.39	14.50
	0.05	13.91	1.3488	12.97	10.79	13.00	11.07
	0.025	19.44	0.9488	11.13	9.13	11.20	9.39
	0.01	30.51	0.5995	7.74	6.46	7.74	6.47
2	0.50	4.64	4.3589	10.54	9.28	10.56	9.31
	0.25	6.43	3.0208	8.28	7.10	8.28	7.10
	0.10	9.95	1.9132	5.85	5.12	5.85	5.12
	0.05	13.84	1.3439	4.07	3.29	4.07	3.29
	0.025	19.36	0.9527	3.26	2.77	3.26	2.77
	0.01	30.45	0.6042	2.12	1.80	2.12	1.80

5. Concluding remarks. There are many possibilities to generalize the results of this paper. An immediate question to ask is whether the results of this paper can be generalized for an arbitrary p.d. unknown variance–covariance matrix Σ using the stopping rules of Ghosh, Sinha and Mukhopadhyay (1976) or Woodroofe (1977). For fixed samples, for an unknown p.d. Σ dominance of James–Stein type estimators over the sample mean are discussed in Berger, Bock, Brown, Casella and Gleser (1977) and in Berger and Haff (1983). Also, one can develop empirical Bayes stopping rules instead of the empirical minimax stopping rules as given in (1.3). Other types of robust Bayes stopping rules are also possible. We are currently exploring some of these ideas.

Acknowledgments. Thanks are due the Associate Editor and the referees for their helpful comments.

REFERENCES

BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645.

BERGER, J. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226.

BERGER, J., BOCK, M. E., BROWN, L. D., CASELLA, G. and GLESER, L. J. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771.

BERGER, J. and HAFF, L. R. (1983). A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Statist. Decisions* **1** 105–129.

CRESSIE, N., DAVIS, A. S., FOLKS, J. L. and POLICELLO, G. E. (1981). The moment-generating function and negative integer moments. *Amer. Statist.* **35** 148–150.

EFRON, B. and MORRIS, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11–21.

FINSTER, M. (1983). A frequentistic approach to sequential estimation in the general linear model. *J. Amer. Statist. Assoc.* **78** 403–407.

GHOSH, M. and MUKHOPADHYAY, N. (1980). Sequential point estimation of the difference of two normal means. *Ann. Statist.* **8** 221–225.

GHOSH, M., NICKERSON, D. M. and SEN, P. K. (1985). Sequential shrinkage estimation. Technical Report 240, Dept. of Statistics, Univ. of Florida.

GHOSH, M. and SEN, P. K. (1983). On two-stage James–Stein estimators. *Sequential Anal.* **2** 359–367.

GHOSH, M., SINHA, B. and MUKHOPADHYAY, N. (1976). Multivariate sequential point estimation. *J. Multivariate Anal.* **6** 281–294.

HUDSON, H. M. (1974). Empirical Bayes estimation. Technical Report 58, Dept. of Statistics, Stanford Univ.

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

KIEFER, J. (1957). Invariance, minimax, sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics* (U. Grenander, ed.) 235–245. Almqvist and Wiksell, Stockholm.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

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

TAKADA, Y. (1984). Inadmissibility of a sequential estimation rule of the mean of a multivariate normal distribution. *Sequential Anal.* **3** 267–271.

WOODROOFE, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.* **5** 984–995.

MALAY GHOSH
DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

DAVID M. NICKERSON
DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602

PRANAB K. SEN
DEPARTMENT OF BIOSTATISTICS 201H
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514