ON THE ASYMPTOTIC DISTRIBUTIONAL RISKS OF SHRINKAGE AND PRELIMINARY TEST VERSIONS OF MAXIMUM LIKELIHOOD ESTIMATORS

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1478

December 1984

ON THE ASYMPTOTIC DISTRIBUTIONAL RISKS OF SHRINKAGE AND PRELIMINARY TEST VERSIONS OF MAXIMUM LIKELIHOOD ESTIMATORS *

By PRANAB KUMAR SEN

University of North Carolina, Chapel Hill, N.C., USA .

SUMMARY. Both the preliminary test and shrinkage versions of the maximum likelihood estimators aim to reduce , in a meaningful setup, the risk of the estimator. Asymptotic distributional risks for these versions of the maximum likelihood estimators are studied , and it is shown that none of these estimators dominate over the other in the light of the asymptotic distributional risk criterion.

1. INTRODUCTION

The *maximum likelihood estimators (MLE)* play a fundamental role in the classical parametric *estimation theory*. Under quite general regularity conditions, the MLE are (asymptotically) *optimal*, where the optimality has been interpreted in various ways; we may refer to Ibragimov and Has'minskii (1981) for an excellent account of this theory. Typically, in a *multi-parameter model*, a proper subspace (ω) of the parameter space (Ω) may be identified from extraneous considerations, and a *restricted MLE* may be defined by limiting considerations to ω , and this may fare better than the *unrestricted MLE* when the assumed parametric restraints actually hold. On the other hand, negation of such assumed restraints may render the restricted MLE as biased, inefficient, and even , possibly, inconsistent. Thus, the restricted MLE may have super-efficiency within a restricted part of the parameter space, but lacks the validity-robustness . This conflicting picture on the performance

* Work partially supported by the National Heart, Lung and Blood Institute, Contract : NIH-NHLBI-71-2243-L from the National Institutes of Health.

AMS Subject Classification Numbers: 62C15, 62E20, 62F12.

Key Words and Phrases : Asymptotic distribution; asymptotic distributional risk; James-Stein rule; maximum likelihood estimator; preliminary test estimation; shrinkage estimation.

characteristics of the restricted MLE's may be of genuine concern in a class of problems where the prior information on ω is rather uncertain (though quite likely to be true) : The use of the restricted MLE's may not be advocated without reservation , although it is tempting to take into consideration this prior information to achieve higher efficiency if possible. In such a situation, both the *preliminary test (PT)* and *shrinkage* versions of MLE's can be used with some distinct advantages. In a preliminary test estimation problem, first, a *preliminary test* (on the validity of ω) is made, and depending on the outcome of this test, a restricted or an unrestricted MLE is used. Thus, a PTMLE is a convex combination of the restricted and unrestricted MLE's , where the mixing coefficients are the indicator variables derived from the preliminary test. On the other hand, in a shrinkage estimation problem, a test statistic (for testing the validity of ω) is directly incorporated in the estimation rule in providing a smooth estimator of the parameter of interest. In either case, confined to the parameter space ω , the estimator performs better than the unrestricted MLE, though, the restricted MLE may still be better than either of them. On the other hand, unlike the case of the restricted MLE , when θ lies outside ω , either of these versions has a more robust performance characteristic, i.e., the negation of the assumed restraints affects the performance of either of these versions to a much lesser extent than the restricted MLE.

For (multivariate) normal and other specific parametric models, the theory of PTE and shrinkage estimation has been neatly developed by a host of workers [viz., Judge and Bock (1978) where other references are also cited]. The exactness of this theory stumbles into immensurable difficulties for a general parametric (or nonparametric) model where the underlying distribution may not necessarily belong to the *exponential family*. Nevertheless, in such a case, *asymptotic theory* has emerged with remarkable simplifications ; we may refer to Sen(1979,1984), Saleh and Sen (1978,1984) and Sen and Saleh (1979, 1985), among others. It is clear from these results that in an asymptotic setup, we need to confine ourselves to a *shrinking neighbourhood* of ω, for which the unrestricted, restricted, PT and shrinkage versions of the MLE would have distinct (and meaningful) performance characteristics , ameanable to comparative studies. For the PTMLE, the asymptotic theory has already been studied in detail in Sen(1979). The object of the current investigation is to focus mainly on the shrinkage MLE and to compare the asymptotic performances for the two versions (i.e., PTE and shrinkage) of the MLE's.

The PTMLE and shrinkage MLE are both introduced in Section 2 , where the preliminary notions are also presented. Asymptotic distribution theory of the shrinkage estimator is considered in Section 3 , and the asymptotic distributional risk results are also presented there. The last section deals with the asymptotic risk efficiency results for the PTMLE and shrinkage MLE. In this context, it is shown that none of these versions dominate over the other in the light of this asymptotic distributional risk criterion.

2. PTMLE, SHRINKAGE MLE AND PRELIMINARY NOTIONS

We closely follow the basic setup in Sen(1979) where the PTMLE have already been discussed. Consider $k(\geq 1)$ independent samples of sizes $n_1, \ldots, n_k$,respectively. Let $X_{i1}, \ldots, X_{in_i}$ be n_i independent and identically distributed (i.i.d.) random vectors (r.v.) with a distribution function (d.f.) $F_i(x, \underset{\sim}{\theta})$, $x \in E^m$ and $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_r)' \in \Omega \subset E^r$, for some $m \geq 1$, $r \geq 1$, for $i=1,\ldots,k$. For each i $(=1,\ldots,k)$ and $\theta \in \Omega$, we assume that $F_i(x, \underset{\sim}{\theta})$ admits a density function $f_i(x, \underset{\sim}{\theta})$ (with respect to some sigma-finite measure μ). Then, the log-likelihood function is defined by

$$\log L_n(\underset{\sim}{X}_n , \underset{\sim}{\theta}) = \Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} \log f_i(X_{ij}, \underset{\sim}{\theta}) , \quad \underset{\sim}{\theta} \in \Omega , \tag{2.1}$$

where $\underset{\sim}{X}_n = (X_{11}, \ldots, X_{kn_k})$ is the sample point ($\in E^{mn}$) ,and $n = n_1 + \ldots + n_k$.

We define the restricted parameter space ω as

$$\omega = \{\underset{\sim}{\theta}: \underset{\sim}{h}(\underset{\sim}{\theta}) = (h_1(\theta),\ldots,h_p(\theta)) = \underset{\sim}{0}\}\ , \text{ for some } p \leq r . \tag{2.2}$$

Then, an unrestricted and a restricted MLE of the true parameter $\underset{\sim}{\theta}$, denoted by $\tilde{\underset{\sim}{\theta}}_n$ and $\hat{\underset{\sim}{\theta}}_n$, respectively, are defined as

$$\log L_n(\underset{\sim}{X}_n , \tilde{\underset{\sim}{\theta}}_n) = \sup_{\theta\varepsilon\Omega} \log L_n(\underset{\sim}{X}_n , \underset{\sim}{\theta}) \quad ; \tag{2.3}$$

$$\log L_n(\underset{\sim}{X}_n , \hat{\underset{\sim}{\theta}}_n) = \sup_{\theta\varepsilon\omega} \log L_n(\underset{\sim}{X}_n , \underset{\sim}{\theta}) \quad . \tag{2.4}$$

For testing the null hypothesis $H_0: \underset{\sim}{\theta}\ \varepsilon\ \omega$, the classical (log-) likelihood ratio test statistic is

$$\mathcal{L}_n = -2 \log\{L_n(\underset{\sim}{X}_n, \hat{\underset{\sim}{\theta}}_n)/L_n(\underset{\sim}{X}_n , \tilde{\underset{\sim}{\theta}}_n)\} . \tag{2.5}$$

The null hypothesis H_0 is then accepted or rejected according as $\mathcal{L}_n$ is $\leq$ or $>$ $\ell_{n,\alpha}$, the level α $(0 < \alpha < 1)$ critical value for $\mathcal{L}_n$. The PTMLE $\hat{\underset{\sim}{\theta}}{}^{PT}_n$ is then defined as

$$\hat{\underset{\sim}{\theta}}{}^{PT}_n = \begin{cases} \hat{\underset{\sim}{\theta}}_n , & \text{if } \mathcal{L}_n \leq \ell_{n,\alpha} , \\ \tilde{\underset{\sim}{\theta}}_n, & \text{if } \mathcal{L}_n > \ell_{n,\alpha} . \end{cases} \tag{2.6}$$

To introduce the shrinkage MLE, we consider the usual James-Stein(1961) rule and define

$$\hat{\underset{\sim}{\theta}}{}^{S}_n = \tilde{\underset{\sim}{\theta}}_n + (\hat{\underset{\sim}{\theta}}_n - \tilde{\underset{\sim}{\theta}}_n)(p-2)/\mathcal{L}_n . \tag{2.7}$$

It may be noted that in (2.7), instead of (p-2), we may consider a sequence $\{c_n\}$ of positive numbers, such that as $n \to \infty$, c_n converges to a positive constant c: $0 < c < 2(p-2)$. However, the asymptotic picture with the risk of such estimators would remain the same, and, in this context, $c=(p-2)$ appears to be the most desirable solution. Also, instead of the scalar multiplication of (p-2), it is possible to use some matrix multiplication (as is usually done in a general shrinkage estimation of the multinormal mean). However, such a matrix has to be specifically chosen from the sample estimates of the unknown *information matrix*, and would therefore complicate the procedure. The end picture ,however, remains essentially the same. Hence, for the sake of simplicity, we shall specifically consider the particular shrinkage estimator

$\hat{\theta}_n^S$ in (2.7) and append a small discussion on the other possibilities. Our primary concern is to study the asymptotic distributional risk properties of $\hat{\theta}_n^S$ and to compare these with those of the PTMLE and the other two versions of the MLE.

For this study, we need to introduce the regularity conditions under which the proposed estimators have the desirable asymptotic properties. For the classical MLE, these regularity conditions have been considered in increasing generality in Cramér (1946), Huber (1967), Hájek (1970), Inagaki (1973) and others. In Sen (1979), these regularity conditions have been stated along with the possible generalizations (to cover more general cases). Hence, for the sake of brevity, we omit these details. However, we introduce the following notations which will be needed in the sequel.

Ω is assumed to be a convex, compact subspace of E^r, and, for each i (=1,...,k), $\log f_i(x,\theta)$ is almost everywhere (a.e.) differentiable with respect to θ (at least twice), these derivatives are dominated by some integrable functions, and the second derivative matrix has the continuity property in the first mean. Further, each of these densities has a finite Kullback-Leibler information. Let us define for each $i(=1,\ldots,k)$,

$$B_{\theta}^{(i)} = ((\textstyle\int\!\!\int_{E^m} (\partial/\partial\theta_j)\log f_i(x,\theta)(\partial/\partial\theta_\ell)\log f_i(x,\theta)dF_i(x,\theta)\ ))_{j,\ell=1,\ldots,r} \quad (2.8)$$

and assume that $B_{\theta}^{(1)}, \ldots, B_{\theta}^{(k)}$ are all continuous in θ in some neighbourhood of the true parameter point θ_o, and that

$$B_{\theta_o}^{*} = \Sigma_{i=1}^{k} (n_i/n)\, B_{\theta_o}^{(i)} \text{ is positive definite (p.d.)} \quad (2.9)$$

We also assume that $h(\theta)$ possesses continuous first and second order derivatives with respect to θ, for every $\theta \in \Omega$. Let then

$$H_{\theta} = ((\ (\partial/\partial\theta)h(\theta)\ ))\ \text{(of order rxp)}, \quad (2.10)$$

and, we assume that

$$H_{\theta_o} \text{ is of rank } p\ (<r)\ . \quad (2.11)$$

Regarding the individual sample sizes $n_1,\ldots,n_k$, we make the conventional assumption that there exist positive numbers $\rho_1, \ldots, \rho_k$, such that

$$\lim_{n\to\infty} (n_i/n) = \rho_i \text{ exists for } i=1,\ldots,k;\ \Sigma_{i=1}^{k}\rho_i = 1\ . \quad (2.12)$$

Finally, we assume that the following matrix of order (p+r)x(p+r)

$$\begin{pmatrix} \underset{\sim}{B}^{*}_{\underset{\sim}{\theta}_o} & -\underset{\sim}{H}_{\underset{\sim}{\theta}_o} \\ -\underset{\sim}{H}'_{\underset{\sim}{\theta}_o} & \underset{\sim}{O} \end{pmatrix} \text{ is p.d.} \tag{2.13}$$

Note that under (2.12), $\underset{\sim}{B}^{*}_{\underset{\sim}{\theta}_o}$ in (2.9) converges to

$$\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} = \Sigma_{i=1}^{k} \rho_i \underset{\sim}{B}^{(i)}_{\underset{\sim}{\theta}_o} . \tag{2.14}$$

Also, note that if we denote the reciprocal matrix of (2.13) by

$$\begin{pmatrix} \underset{\sim}{P}^{*}_{\underset{\sim}{\theta}_o} & \underset{\sim}{Q}^{*}_{\underset{\sim}{\theta}_o} \\ \underset{\sim}{Q}^{*\prime}_{\underset{\sim}{\theta}_o} & \underset{\sim}{R}^{*}_{\underset{\sim}{\theta}_o} \end{pmatrix} \tag{2.15}$$

then, under (2.12), (2.15) converges to

$$\begin{pmatrix} \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} & \bar{\underset{\sim}{Q}}_{\underset{\sim}{\theta}_o} \\ \bar{\underset{\sim}{Q}}'_{\underset{\sim}{\theta}_o} & \bar{\underset{\sim}{R}}_{\underset{\sim}{\theta}_o} \end{pmatrix} \text{ where } \begin{pmatrix} \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} & \bar{\underset{\sim}{Q}}_{\underset{\sim}{\theta}_o} \\ \bar{\underset{\sim}{Q}}'_{\underset{\sim}{\theta}_o} & \bar{\underset{\sim}{R}}_{\underset{\sim}{\theta}_o} \end{pmatrix} \begin{pmatrix} \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} & -\underset{\sim}{H}_{\underset{\sim}{\theta}_o} \\ -\underset{\sim}{H}'_{\underset{\sim}{\theta}_o} & \underset{\sim}{O} \end{pmatrix} = \underset{\sim}{I}_{p+r} . \tag{2.16}$$

Note that $\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}$, $\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$ and $\bar{\underset{\sim}{R}}_{\underset{\sim}{\theta}_o}$ are all symmetric matrices. Also, we need the following definition of the <u>likelihood score function</u> :

$$\underset{\sim}{\Lambda}_n(\underset{\sim}{\theta}) = n^{-\frac{1}{2}}(\partial/\partial\underset{\sim}{\theta})\log L_n(\underset{\sim}{X}_n ,\underset{\sim}{\theta}) , \quad \underset{\sim}{\Lambda}^o_n = \underset{\sim}{\Lambda}_n(\underset{\sim}{\theta}_o) , \tag{2.17}$$

where $\underset{\sim}{\theta}_o$ is the ture parameter point.

3. ASYMPTOTIC DISTRIBUTION OF SHRINKAGE MLE

For simplicity of presentation, we consider first the case where $\underset{\sim}{\theta}_o \in \omega$. It follows from Sen(1979) that under the assumed regularity conditions, as $n \to \infty$,

$$n^{\frac{1}{2}}(\tilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) = \bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o} \underset{\sim}{\Lambda}^o_n + \underset{\sim}{o}_p(1) \overset{\mathcal{D}}{\sim} \mathcal{N}(\underset{\sim}{0}, \bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o}) , \tag{3.1}$$

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) = \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} \underset{\sim}{\Lambda}^o_n + \underset{\sim}{o}_p(1) \overset{\mathcal{D}}{\sim} \mathcal{N}(\underset{\sim}{0}, \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}) , \tag{3.2}$$

$$\mathcal{L}_n = \underset{\sim}{\Lambda}^{o\prime}_n (\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} - \bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o})\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} (\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} - \bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o})\underset{\sim}{\Lambda}^o_n + o_p(1) \overset{\mathcal{D}}{\sim} \chi^2_p , \tag{3.3}$$

where χ^2_p is a r.v. having the central chi square d.f. with p degrees of freedom (DF); we denote the upper 100α% point of this d.f. by $\chi^2_{p,\alpha}$. We may also write

$$\mathcal{L}_n = n(\hat{\underset{\sim}{\theta}}_n - \tilde{\underset{\sim}{\theta}}_n)'\bar{B}_{\underset{\sim}{\theta}_o}(\hat{\underset{\sim}{\theta}}_n - \tilde{\underset{\sim}{\theta}}_n) + o_p(1) . \tag{3.4}$$

By virtue of (3.1) through (3.4), under H_0: $\underset{\sim}{\theta}_o \in \omega$, we have

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}^S_n - \underset{\sim}{\theta}_o) = n^{\frac{1}{2}}(\tilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) + (p-2)(\underset{\sim}{\Lambda}^{o'}_n \underset{\sim}{A}\underset{\sim}{\Lambda}_n)^{-1}.n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n - \tilde{\underset{\sim}{\theta}}_n) + \underset{\sim}{o}_p(1)$$

$$= [\underset{\sim}{I} - (p-2)(\underset{\sim}{\Lambda}^{o'}_n\underset{\sim}{A}\underset{\sim}{\Lambda}^o_n)^{-1}(\underset{\sim}{I}-\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o})]\bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o}\underset{\sim}{\Lambda}^o_n + \underset{\sim}{o}_p(1) , \tag{3.5}$$

where

$$\underset{\sim}{A} = (\underset{\sim}{I} - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o})\bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o}(\underset{\sim}{I} - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}) . \tag{3.6}$$

Consider next, the case of $\underset{\sim}{\theta}_o \notin \omega$. It follows from the results in Sen (1979) that for every $\underset{\sim}{\theta}_o \notin \omega$, $n^{-1}\mathcal{L}_n$ converges , in probability, to a positive constant (depending on $\underset{\sim}{h}(\underset{\sim}{\theta}_o) \neq \underset{\sim}{0}$), so that by (2.7),

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}^S_n - \underset{\sim}{\theta}_o) = n^{\frac{1}{2}}(\tilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) + \underset{\sim}{o}_p(1) . \tag{3.7}$$

The situation is quite different when $\underset{\sim}{\theta}_o$ lies near the boundary of ω . For this case, we consider the following sequence $\{K_n\}$ of local alternatives :

$$K_n : \underset{\sim}{h}(\underset{\sim}{\theta}_o) = n^{-\frac{1}{2}}\underset{\sim}{\gamma} , \quad \underset{\sim}{\gamma} \text{ a real vector in } E^p . \tag{3.8}$$

Note that for the nonparametric estimation problems, for shrinkage estimators, it has been shown in Sen(1984) and Sen and Saleh(1985) that in such a shrinking neighbourhood of the null hypothesis parameter space, the shrinkage estimators have nice asymptotic properties and dominate the usual ones. In the current study, we like to present the same picture for the shrinkage MLE.

First, we note that by virtue of (2.10)-(2.11) and (3.8), we can conceive of a sequence $\{\underset{\sim}{\theta}^{(n)}_o\}$ of parametric points , such that $\underset{\sim}{h}(\underset{\sim}{\theta}^{(n)}_o) = \underset{\sim}{0}$, and

$$\lim_{n\to\infty} n^{\frac{1}{2}}(\underset{\sim}{\theta}^{(n)}_o - \underset{\sim}{\theta}_o) = \underset{\sim}{\gamma}^* \text{ exists , where } \underset{\sim}{\gamma} = \underset{\sim}{H}'_{\underset{\sim}{\theta}_o}\underset{\sim}{\gamma}^* . \tag{3.9}$$

As such, proceeding as in Section 4 of Sen (1979), it follows that under $\{K_n\}$,

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) = \underset{\sim}{\gamma}^* + \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}\underset{\sim}{\Lambda}^o_n + \underset{\sim}{o}_p(1) , \tag{3.10}$$

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n - \tilde{\underset{\sim}{\theta}}_n) = \underset{\sim}{\gamma}^* + (\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} - \bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o})\underset{\sim}{\Lambda}^o_n + \underset{\sim}{o}_p(1) , \tag{3.11}$$

and

$$\mathcal{L}_n = [(\underset{\sim}{I} - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}) \underset{\sim}{\Lambda}_n^o - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \underset{\sim}{\gamma}^*]' \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1} [(\underset{\sim}{I} - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}) \underset{\sim}{\Lambda}_n^o - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \underset{\sim}{\gamma}^*] + o_p(1) \tag{3.12}$$

where we note that by letting $\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} = \underset{\sim}{D}\underset{\sim}{D}'$ ($\underset{\sim}{D}$ nonsingular) ,

$$\underset{\sim}{D}^* = \underset{\sim}{I} - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} \quad \text{and} \quad \underset{\sim}{D}_o = \underset{\sim}{I} - \underset{\sim}{D}' \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} \underset{\sim}{D} \quad \text{are idempotent matrices,} \tag{3.13}$$

(of rank p) . Further, (3.1) holds under $\{K_n\}$ as well. Thus, from (2.7), (3.1), (3.10), (3.11), (3.12) and (3.13), we have under $\{K_n\}$ and the assumed regularity conditions,

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n^S - \underset{\sim}{\theta}_o) = \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1} \left\{ \underset{\sim}{\Lambda}_n^o - \frac{(p-2)[\underset{\sim}{D}^* \underset{\sim}{\Lambda}_n^o - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \gamma^*]}{[\underset{\sim}{D}^* \underset{\sim}{\Lambda}_n^o - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \underset{\sim}{\gamma}^*]' \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1} [\underset{\sim}{D}^* \underset{\sim}{\Lambda}_n^o - \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \underset{\sim}{\gamma}^*]} \right\} + \underset{\sim}{o}_p(1)$$

$$= (\underset{\sim}{D}')^{-1} \left\{ \underset{\sim}{U}_n - \frac{(p-2)[\underset{\sim}{D}_o \underset{\sim}{U}_n - \underset{\sim}{D}' \underset{\sim}{\gamma}^*]}{[\underset{\sim}{D}_o \underset{\sim}{U}_n - \underset{\sim}{D}' \underset{\sim}{\gamma}^*]' [\underset{\sim}{D}_o \underset{\sim}{U}_n - \underset{\sim}{D}' \underset{\sim}{\gamma}^*]} \right\} + o_p(1) , \tag{3.14}$$

where

$$\underset{\sim}{U}_n = \underset{\sim}{D}^{-1} \underset{\sim}{\Lambda}_n^o \xrightarrow{\mathfrak{D}} \underset{\sim}{U} \overset{\mathfrak{D}}{=} \mathcal{N}(\underset{\sim}{0}, \underset{\sim}{I}_r) \tag{3.15}$$

and

$$[\underset{\sim}{D}_o \underset{\sim}{U}_n - \underset{\sim}{D}' \underset{\sim}{\gamma}^*]' [\underset{\sim}{D}_o \underset{\sim}{U}_n - \underset{\sim}{D}' \underset{\sim}{\gamma}^*] \xrightarrow{\mathfrak{D}} [\underset{\sim}{D}_o \underset{\sim}{U} - \underset{\sim}{D}' \underset{\sim}{\gamma}^*]' [\underset{\sim}{D}_o \underset{\sim}{U} - \underset{\sim}{D}' \underset{\sim}{\gamma}^*]$$

$$\overset{\mathfrak{D}}{=} \chi_p^2(\Delta) \quad ; \quad \Delta = \underset{\sim}{\gamma}^{*\prime} \underset{\sim}{D}\underset{\sim}{D}' \underset{\sim}{\gamma}^* = \underset{\sim}{\gamma}^{*\prime} \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o} \underset{\sim}{\gamma}^* , \tag{3.16}$$

where $\chi_p^2(\Delta)$ stands for a r.v. having the non-central chi square d.f. with p DF and noncentrality parameter Δ .

We are now in a position to define the asymptotic distributional risk (ADR) of the shrinkage and other estimators. For a suitable estimator $\underset{\sim}{\theta}_n^*$ of $\underset{\sim}{\theta}_o$, we denote the asymptotic d.f. by

$$G^*(\underset{\sim}{x}) = \lim_{n\to\infty} P\{ n^{\frac{1}{2}}(\underset{\sim}{\theta}_n^* - \underset{\sim}{\theta}_o) \le \underset{\sim}{x} \mid K_n \} , \tag{3.17}$$

where, we assume that this asymptotic d.f. is nondegenerate . Also, we consider a suitable positive semidefinite (p.s.d.) matrix $\underset{\sim}{W}$. Then, the ADR of $\underset{\sim}{\theta}_n^*$,at the point $\underset{\sim}{\theta}_o$,is defined by

$$\rho(\underset{\sim}{\theta}^* ; \underset{\sim}{W}) = \text{Trace}(\underset{\sim}{W}[\int \cdots \int \underset{\sim}{x}\underset{\sim}{x}' dG^*(\underset{\sim}{x})]) . \tag{3.18}$$

Note that by (3.1) and (3.18),

$$\rho(\tilde{\theta}\,;\,W) = \mathrm{Tr}(W\bar{B}_{\theta_0}^{-1}) \text{ , for every } \gamma^* \in E^r , \tag{3.19}$$

while, by (3.10) and (3.18),

$$\rho(\hat{\theta}\,;\,W) = \mathrm{Tr}(W\bar{P}_{\theta_0}) + (\gamma^{*\prime}W\gamma^*) , \tag{3.20}$$

where, we have made use of the identity that $\bar{P}_{\theta_0}\bar{B}_{\theta_0}\bar{P}_{\theta_0} = \bar{P}_{\theta_0}$. Let us denote by $W_0 = (D)^{-1}W(D')^{-1}$, where D is defined before (3.12). Then, by (3.14) through (3.18), we obtain that

$$\rho(\hat{\theta}^S\,;\,W) = \mathrm{Tr}(W_0) + (p-2)^2 E\left\{\frac{(D_0U-D'\gamma^*)'W_0(D_0U-D'\gamma^*)}{([D_0U-D'\gamma^*]'[D_0U-D'\gamma^*])^2}\right\}$$

$$- 2(p-2)E\left\{\frac{(D_0U-D'\gamma^*)'W_0U}{([D_0U-D'\gamma^*]'[D_0U-D'\gamma^*])}\right\} , \tag{3.21}$$

where

$$\mathrm{Tr}(W_0) = \mathrm{Tr}((D')^{-1}W(D)^{-1}) = \mathrm{Tr}(W\bar{B}_{\theta_0}^{-1}) . \tag{3.22}$$

Note that by (3.13), $(I - D_0)D_0 = D_0 - D_0 = 0$, so that $(I - D_0)U$ and D_0U are independent r.v.'s with normal distributions with null mean vectors and dispersion matrix $I-D_0$ and D_0, respectively. For the last term on the right hand side of (3.21), we write $(D_0U - D'\gamma^*)'W_0U = (D_0U - D'\gamma^*)'W_0D_0U + (D_0U - D'\gamma^*)'W_0(I- D_0)U$, and make use of the independence of D_0U and $(I - D_0)U$. With this manipulation, the last term on the right hand side of (3.21) is equal to

$$-2(p-2)E\{([D_0U- D'\gamma^*]'[D_0U- D'\gamma^*])^{-1}(D_0U- D'\gamma^*)'W_0D_0U\} + 0$$

$$= -2(p-2)\{E(\chi_{p+2}^{-2}(\Delta))\mathrm{Tr}(W_0) + (\gamma^{*\prime}W\gamma^*)[E(\chi_{p+2}^{-2}(\Delta))+ E(\chi_{p+4}^{-2}(\Delta))]\} , \tag{3.23}$$

where Δ is defined by (3.16), and in the last step, we have made use of the Stein identity [viz., Appendix B of Judge and Bock (1978)]. In this context note that $\mathrm{Tr}((DD')W_0(\gamma^*\gamma^{*\prime})) = \mathrm{Tr}(W(\gamma^*\gamma^{*\prime})) = (\gamma^{*\prime}W\gamma^*)$, as it appeared in (3.20). Similarly, by using the Stein identity, the second term on the right hand side of (3.21) reduces to

$$(p-2)^2\{E(\chi_{p+2}^{-4}(\Delta))\mathrm{Tr}(W_0) + (\gamma^{*\prime}W\gamma^*)E(\chi_{p+4}^{-4}(\Delta))\} . \tag{3.24}$$

Therefore, from (3.21), (3.22), (3.23) and (3.24), we conclude that the ADR of of the shrinkage MLE is given by

$$\rho(\hat{\underset{\sim}{\theta}}^{S};\underset{\sim}{W}) = \mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1}) + (p-2)^2\{\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1})E(\chi_{p+2}^{-4}(\Delta)) + (\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{W}\underset{\sim}{\gamma}^{*})E(\chi_{p+4}^{-4}(\Delta))\}$$
$$- 2(p-2)\{\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1})E(\chi_{p+2}^{-2}(\Delta)) + (\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{W}\underset{\sim}{\gamma}^{*})[E(\chi_{p+2}^{-2}(\Delta))+E(\chi_{p+4}^{-2}(\Delta))]\}, \tag{3.25}$$

where we note that by the Courant theorem,

$$\Delta^{*} = \underset{\sim}{\gamma}^{*\prime}\underset{\sim}{W}\underset{\sim}{\gamma}^{*} = \Delta(\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{W}\underset{\sim}{\gamma}^{*})/(\underset{\sim}{\gamma}^{*\prime}\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{\gamma}^{*}) \le \Delta ch_1(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1})$$
$$\le \Delta\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1}), \text{ for every } \underset{\sim}{\gamma}^{*} \in E^{r}. \tag{3.26}$$

We conclude this section with the following result on the ADR of the PTMLE, which follows directly from the results in Section 5 of Sen (1979) :

$$\rho(\hat{\underset{\sim}{\theta}}^{PT};\underset{\sim}{W}) = \mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1}) - \mathrm{Tr}(\underset{\sim}{W}(\bar{\underset{\sim}{B}}_{\theta_o}^{-1} - \bar{\underset{\sim}{P}}_{\theta_o}))\Pi_{p+2}(\chi_{p,\alpha}^{2};\Delta) +$$
$$\Delta^{*}[\Pi_{p+2}(\chi_{p,\alpha}^{2};\Delta) - \Pi_{p+4}(\chi_{p,\alpha}^{2};\Delta)]$$
$$= [1 - \Pi_{p+2}(\chi_{p,\alpha}^{2};\Delta)]\rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W}) + \Pi_{p+2}(\chi_{p,\alpha}^{2};\Delta)\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W}) +$$
$$\Delta^{*}[\Pi_{p+2}(\chi_{p,\alpha}^{2};\Delta) - \Pi_{p+4}(\chi_{p,\alpha}^{2};\Delta)], \tag{3.27}$$

where $\Pi_q(x;\delta)$ is the non-central chi square d.f. with q DF and noncentrality parameter δ.

4. ASYMPTOTIC RISK-EFFICIENCY RESULTS

First, we note that by (3.13), $\bar{\underset{\sim}{B}}_{\theta_o}^{-1} - \bar{\underset{\sim}{P}}_{\theta_o} = \bar{\underset{\sim}{B}}_{\theta_o}^{-1}\underset{\sim}{D}^{*}$ is of rank p, and we assume that $p > 2$ (otherwise, the shrinkage estimation would not lead to any reduction in the risk). To avoid trivial consequences, we therefore assume that for the chosen $\underset{\sim}{W}$,

$$\mathrm{Tr}(\underset{\sim}{W}(\bar{\underset{\sim}{B}}_{\theta_o}^{-1} - \bar{\underset{\sim}{P}}_{\theta_o})) = c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\theta_o},\bar{\underset{\sim}{P}}_{\theta_o}) > 0. \tag{4.1}$$

From (3.19) and (3.20), we have then

$$\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W}) = \rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W}) - c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\theta_o},\bar{\underset{\sim}{P}}_{\theta_o}) + \Delta^{*}, \tag{4.2}$$

so that

$$\{\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W})/\rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W})\} \gtreqless 1 \text{ according as } (\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{W}\underset{\sim}{\gamma}^{*}) \gtreqless c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\theta_o},\bar{\underset{\sim}{P}}_{\theta_o}). \tag{4.3}$$

In particular, under H_0, $\underset{\sim}{\gamma}^{*} = \underset{\sim}{0}$, so that $\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W}) < \rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W})$. On the other hand, outside the ellipsoid : $\{\underset{\sim}{\gamma}^*: \underset{\sim}{\gamma}^{*\prime}\underset{\sim}{W}\underset{\sim}{\gamma}^* \le c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\theta_o},\bar{\underset{\sim}{P}}_{\theta_o})\}$, $\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W}) > \rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W})$, and moreover, as $\underset{\sim}{\gamma}^*$ moves away from the origin (i.e., Δ^* or Δ increases), $\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W})$ converges to $+\infty$, while $\rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W})$ does not depend on $\underset{\sim}{\gamma}^*$ (or Δ^* or Δ). Thus, in the light of the A.D.R., none of $\tilde{\underset{\sim}{\theta}}_n$ and $\hat{\underset{\sim}{\theta}}_n$ dominates over the other. Further, this also reflects the lack of <u>(risk-) efficiency robustness</u> of the restricted MLE against any departure from the assumed restraints for which Δ^* is not small.

From (3.19) and (3.25), we have

$$\rho(\hat{\underset{\sim}{\theta}}^S;\underset{\sim}{W})/\rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W}) = 1 + (p-2)^2\{ E(\chi_{p+2}^{-4}(\Delta)) + [\Delta^*/\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1})]E(\chi_{p+4}^{-4}(\Delta))\} - 2(p-2)\{ E(\chi_{p+2}^{-2}(\Delta)) + [\Delta^*/\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1})]E(\chi_{p+2}^{-2}(\Delta)+ \chi_{p+4}^{-2}(\Delta))\}, \quad (4.4)$$

where, by virtue of (3.26), $\Delta^*/\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1}) \le \Delta$, for every $\underset{\sim}{\gamma}^* \varepsilon E^r$. This familiar expression arises in connection with the James-Stein rule estimator of the multivariate normal mean vector, and hence, we may conclude that the following holds:

(i) Under H_0: $\theta_o \ \varepsilon\ \omega$, i.e., $\underset{\sim}{\gamma}^* = \underset{\sim}{0}$, $\Delta = \Delta^* = 0$, so that (4.4) reduces to $2/p$ (≤ 1) , for every $p \ge 2$. The larger is the value of p, the smaller is the ratio in (4.4), so that greater is the reduction of the ADR , under H_0.

(ii) The right hand side of (4.4) is never greater than 1; in fact, the upper asymptote 1 is attained as Δ or Δ^* goes to $+\infty$. Thus, in the light of the ADR, the shrinkage estimator $\hat{\underset{\sim}{\theta}}_n^S$ dominates over the unrestrained MLE $\hat{\underset{\sim}{\theta}}_n$.

Consider next the case of $\{\hat{\underset{\sim}{\theta}}_n^S\}$ vs. $\{\hat{\underset{\sim}{\theta}}_n\}$. By (3.20), (3.25) , (4.2) and (4.4), we have then

$$\rho(\hat{\underset{\sim}{\theta}}^S;\underset{\sim}{W})/\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W}) = \{\rho(\hat{\underset{\sim}{\theta}}^S;\underset{\sim}{W})/\rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W})\}\{ 1 - [c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\theta_o},\bar{\underset{\sim}{P}}_{\theta_o})-\Delta^*]/\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1})\}^{-1}. \quad (4.5)$$

Note that for every $\underset{\sim}{\gamma}^* \notin \mathcal{E} = \{\underset{\sim}{\gamma}^* : \underset{\sim}{\gamma}^{*\prime}\underset{\sim}{W}\underset{\sim}{\gamma}^* \le c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\theta_o},\bar{\underset{\sim}{P}}_{\theta_o})\}$, $\Delta^* > c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\theta_o},\bar{\underset{\sim}{P}}_{\theta_o})$, so that

$$\sup_{\gamma^*\notin\mathcal{E}} \frac{\rho(\hat{\underset{\sim}{\theta}}^S;\underset{\sim}{W})}{\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W})} < \sup_{\gamma^*\notin\mathcal{E}} \frac{\rho(\hat{\underset{\sim}{\theta}}^S;\underset{\sim}{W})}{\rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W})} \le 1 . \quad (4.6)$$

On the other hand, the picture may be quite different for $\underset{\sim}{\gamma}^* \varepsilon\ \mathcal{E}$ (and hence, under H_0 as well). Note that

$$c_p(\underset{\sim}{W},\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o},\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o})/\text{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1}) = 1 - \text{Tr}((\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1})\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o})/\text{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1}) , \tag{4.7}$$

where $\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$ is an idempotent matrix of rank r-p. Thus, whenever $\underset{\sim}{W}$ is of full rank (r), the right hand side of (4.7) reduces to (r-p)/r. On the other hand, $\underset{\sim}{W}$ need not be of full rank (as will be explained later on). In that case, if $\underset{\sim}{W}$ has the rank $r'(\leq r)$, $\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1}$ will have the same rank r', while the rank of $\underset{\sim}{W}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$ is bounded from below by $r' + (r-p) - r$ $(= r'-p)$ and above by min(r', r-p). We denote the rank of $\underset{\sim}{W}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$ by r^*, so that $r'-p \leq r^* \leq \min(r',r-p)$. In this case, some standard matrix manipulations lead us to conclude that (4.7) reduces to $1 - r^*/r'$, so that under H_0: $\underset{\sim}{\gamma}^* = \underset{\sim}{0}$, (4.5) reduces to

$$(2/p)(r'/r^*) = 2r'/(pr^*) \gtreqless 1 \text{ according as } r^* \lesseqgtr 2r'/p . \tag{4.8}$$

Thus, in one extreme case, when $r'= r$ and $r^* = r-p$, in order that (under H_0) (4.5) is less than one, we need that $r > (p-2)^{-1}p^2$, $p > 2$; the opposite inequality holds (in (4.5)) when $r < (p-2)^{-1}p^2$. In the other extreme case, when $r' = p$ and $r^* \leq 1$, in order that (4.5) is greater than one, we need $p \geq 1$, i.e., in this case, under H_0, $\hat{\underset{\sim}{\theta}}_n$ dominates the shrinkage MLE $\hat{\underset{\sim}{\theta}}_n^S$. This is ,of course, not surprising, as under r'=p, for the restricted MLE $\hat{\underset{\sim}{\theta}}_n$, tacitly, $\underset{\sim}{W}$ attaches full weight only to the restaints , so that it has no contribution from the remaining (r-p) components , and $\text{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}) \leq p^{-1}.\text{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1})$. To sumup, we may therefore conclude that unlike the case of the unrestrained MLE, here, the shrinkage MLE may or may not dominate over the restricted MLE (under H_0), depending on the ranks r' and r^*. To illustrate this point, we consider the following classical example.

Let $\{\underset{\sim}{X}_i, i \geq 1\}$ be i.i.d.r.v.'s having the m-variate normal distribution with mean vector $\underset{\sim}{\mu}$ and dispersion matrix $\underset{\sim}{\Sigma}$. First, consider the case of known $\underset{\sim}{\Sigma}$, and let $\underset{\sim}{\theta} = \underset{\sim}{\mu}$ and $\underset{\sim}{h}(\underset{\sim}{\theta}) = \underset{\sim}{\theta}$. Thus, H_0 refers to $\underset{\sim}{\theta} = \underset{\sim}{0}$. In this case, p=m while $\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$ is a null matrix, so that $r^* = 0$ and $r' = p$. Thus, (4.7) is equal to 1, and hence, (4.5) is $+\infty$. This is not surprising as $\hat{\underset{\sim}{\theta}}_n = \underset{\sim}{0}$,with probability 1, and hence, under H_0, the ADR of the restricted MLE is equal to $\underset{\sim}{0}$, while the ADR of the shrinkage MLE is a positive number. Consider next the case of an unknown $\underset{\sim}{\Sigma}$, and denote by $\underset{\sim}{\sigma}$ the $\binom{m+1}{2}$-vector consisting of the elements in the upper triangle of $\underset{\sim}{\Sigma}$. Let then $\underset{\sim}{\theta} = (\underset{\sim}{\mu}', \underset{\sim}{\sigma}')' = (\underset{\sim}{\theta}_1', \underset{\sim}{\theta}_2')'$, and let $\underset{\sim}{h}(\underset{\sim}{\theta}) = \underset{\sim}{\theta}_1$. Then, $p = m$ and $r = m(m+3)/2$. If our sole interest lies in the estimation of $\underset{\sim}{\mu}$ (treating $\underset{\sim}{\sigma}$ as a nuisance parameter), we may choose $\underset{\sim}{W}$ such that $\underset{\sim}{W} = \begin{pmatrix} \underset{\sim}{W}^o & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}$ where $\underset{\sim}{W}^o$ is a p.d. matrix of order $p \times p$. In that case, we have again $r' = p$ and $r^* = 0$, so that (4.5) is equal to $+\infty$ (under H_0). On the otherhand, if $\underset{\sim}{W}$ is chosen to be a p.d. matrix , then, we have $r' = r = p(p+3)/2$ and $r^* = r - p = \binom{p+1}{2}$, so that $p^{-1}2r' = p+3$ and r^* is $>$ (or $\geq$) $p^{-1}2r'$, for every $p >$ (or $\geq$) 3. Thus, for $m \geq 4$, $\hat{\underset{\sim}{\theta}}{}_n^S$ dominates over $\hat{\underset{\sim}{\theta}}_n$ (under H_0 as well as elsewhere) , while for $m = 3$, under H_0, they have the same ADR, though, elsewhere $\hat{\underset{\sim}{\theta}}{}_n^S$ would dominate over $\hat{\underset{\sim}{\theta}}_n$. Consider next the case, where $\underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{I}_m$, σ unknown. Here, we have $r = m+1 = p+1$ and $r^* = r-p = 1$. Thus, for $\underset{\sim}{W}$ of rank $p+1$, we have $r' = r = p+1$, and hence, $2r'/(pr^*) = 2r'/p = 2 + 2/p > 1$, so that under H_0 , (4.5) is equal to $2(1+p^{-1}) > 2$. Similarly, if $\underset{\sim}{\Sigma} = \sigma^2[\, (1-\rho)\underset{\sim}{I} + \rho \underset{\sim}{1}\underset{\sim}{1}' \,]$,where σ and ρ are unknown, we have $r = m+2 = p+2$ and $r^* = 2$. Thus, for $\underset{\sim}{W}$ of rank r $(= p+2)$, we have $2r'/(pr^*) = (p+2)/p = 1 + 2/p > 1$, and hence, under H_0 , (4.5) exceeds 1 . In either of the last two cases, if we take W of rank p (with null elements for the last one or two rows and columns), as before, we have under H_0 , (4.5) equal to $+\infty$. This explains clearly the dependence of the ADR-efficiency on the choice of $\underset{\sim}{W}$, particularly through the rank of $\underset{\sim}{W}\bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$.

Let us finally compare the ADR of the shrinkage and the PT MLE's. From (3.25), (3.27), (4.4) and (4.5), we obtain that

$$\frac{\rho(\hat{\underset{\sim}{\theta}}{}^S;\, \underset{\sim}{W})}{\rho(\hat{\underset{\sim}{\theta}}{}^{PT};\, \underset{\sim}{W})} = \left\{\frac{\rho(\hat{\underset{\sim}{\theta}}{}^S;\, \underset{\sim}{W})}{\rho(\tilde{\underset{\sim}{\theta}}\, ;\, \underset{\sim}{W})}\right\}\left\{ 1 - \Pi_{p+2}(\chi^2_{p,\alpha};\Delta) + \Pi_{p+2}(\chi^2_{p,\alpha};\Delta) \left\{\frac{\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W})}{\rho(\tilde{\underset{\sim}{\theta}};\underset{\sim}{W})}\right\} + (\Delta^*/\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}{}^{-1}_{\underset{\sim}{\theta}_o}))\,[\, \Pi_{p+2}(\chi^2_{p,\alpha};\Delta) - \Pi_{p+4}(\chi^2_{p,\alpha};\Delta) \,]\right\}^{-1}. \tag{4.9}$$

Note that for given p, $\alpha: 0 < \alpha < 1$ and Δ (≥ 0), $\Pi_q(\chi^2_{p,\alpha};\Delta)$ is nonincreasing in q (≥ 1), and for given p, q and α, it is nonincreasing in Δ. Thus, for every $\alpha \in (0,1)$ and Δ (≥ 0), $\Pi_{p+2}(\chi^2_{p,\alpha};\Delta) - \Pi_{p+4}(\chi^2_{p,\alpha};\Delta)$ is nonnegative ; by using (3.26), it also follows that $(\Delta^*/\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}^{-1}_{\underset{\sim}{\theta}_o}))[\Pi_{p+2}(\chi^2_{p,\alpha};\Delta) - \Pi_{p+4}(\chi^2_{p,\alpha};\Delta)]$ is equal to 0 at $\Delta = 0$ and it converges to 0 as $\Delta \to +\infty$. In fact, this is a bounded nonnegative function, where the (finite) upper bound depends on p, α, $\underset{\sim}{W}$ and $\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}$. Hence, defining the ellipsoid $\mathcal{E}$ as in before (4.6), we obtain from (4.4), (4.7) and (4.9) that

$$\rho(\hat{\underset{\sim}{\theta}}^S; \underset{\sim}{W}) / \rho(\hat{\underset{\sim}{\theta}}^{PT}; \underset{\sim}{W}) \leq 1 \text{, for every } \underset{\sim}{\gamma}^* \notin \mathcal{E} \quad ; \tag{4.10}$$

actually, the left hand side is usually close to 1 and it converges to 1 as $\Delta \to +\infty$. Thus, the dominance of the shrinkage MLE over the PTMLE (in the light of the ADR) in the domain $\{\underset{\sim}{\gamma}^* \notin \mathcal{E}\}$ follows from (4.10). Using the results in (4.7) and (4.8), we also claim that under H_0, (4.9) reduces to

$$\frac{2}{p}\{1 - \Pi_{p+2}(\chi^2_{p,\alpha};0) + \Pi_{p+2}(\chi^2_{p,\alpha};0)(r^*/r')\}^{-1} , \tag{4.11}$$

where r' and r* are defined as in before (4.8). Consequently, under H_0 (and for $p \geq 2$),

$$(4.9) \text{ is } \gtreqless 1 \text{ according as } \frac{r^*}{r'} \text{ is } \lesseqgtr 1 - (p-2)/\{p\Pi_{p+2}(\chi^2_{p,\alpha};0)\} , \tag{4.12}$$

where $\Pi_{p+2}(\chi^2_{p,\alpha};0) < \Pi_p(\chi^2_{p,\alpha};0) = 1 - \alpha$. Since $r^* \leq r'$, for $p = 2$, (4.9) is always greater than or rqual to 1. On the other hand, for $p \geq 3$, the picture depends on the rank of $\underset{\sim}{W}$ and other factors. When $\underset{\sim}{W}$ is of full rank (r), we have,as in before, $r' = r$ and $r^* = r - p$, so that (4.12) reduces to

$$(4.9) \text{ is } \gtreqless 1 \text{ according as } r \lesseqgtr (p-2)^{-1}p^2\Pi_{p+2}(\chi^2_{p,\alpha};0) . \tag{4.13}$$

In particular, if $r = p$ (as was the case discussed for the normal mean vector with known $\underset{\sim}{\Sigma}$), (4.9) exceeds one whenever $\Pi_{p+2}(\chi^2_{p,\alpha};0) > (p-2)/p$. This is the case for $\alpha = 0.10$ when $p \leq 11$ and for $\alpha = 0.05$ when $p \leq 21$. Thus, unless p is very large (and/ or α is very small), under H_0, for the case of $r=p$, the PTMLE dominates over the shrinkage MLE. Let us next go back to (4.13) when

$r > p$; there, under H_0, (4.9) will be less than 1 when $\Pi_{p+2}(\chi^2_{p,\alpha};0)$ is < $rp^{-2}(p-2)$. Since $\Pi_{p+2}(\chi^2_{p,\alpha};0) < \Pi_p(\chi^2_{p,\alpha};0) = 1-\alpha$, a sufficient condition for (4.9) to be less than 1 (under H_0) is that $1-\alpha \leq r(p-2)/p^2$, and this condition is easy to verify when r is large compared to p. It also shows that choosing a large α (the significance level) for the preliminary test may lead to an increased ADR for the PTMLE. Let us consider next the case where $\underset{\sim}{W}$ is not of full rank (as was discussed in the normal d.f. example earlier). In this case, we may have an extreme situation where $r' = p$ and r^* is either 0 or 1. For $r^* = 0$, by virtue of (4.12), in order that under H_0, (4.9) is greater than one, we need to have $\Pi_{p+2}(\chi^2_{p,\alpha};0) > (p-2)/p$, and this agrees with (4.13) when $r = p$. Thus, for $\alpha = 0.10$, for $p \leq 11$ and for $\alpha = 0.05$, for $p \leq 21$, (4.9) is greater than 1, indicating the dominance of the PTMLE (over the shrinkage MLE) under H_0. For $r'=p$ and $r^* = 1$, in order that under H_0, (4.9) exceeds one, we need to have $\Pi_{p+2}(\chi^2_{p,\alpha};0) > (p-2)/(p-1) = 1-(p-1)^{-1}$. Again, by reference to the chi square distributional tables, we conclude that for $\alpha = 0.05$, for all $p \leq 9$ (and for all $p \leq 5$ for $\alpha = 0.10$), (4.9) exceeds 1 (under H_0). In such a case, depending on p and α , the shrinkage MLE may dominate over the PTMLE (under H_0 and hence, elsewhere too) , though this feature is not generally true.

The general conclusion from the above discussion is that the shrinkage MLE may not dominate over the PTMLE , though it does so outside a small neighbourhood of the parameter space (ω) under H_0 . Moreover, over the parameter space where the shrinkage estimator dominates, the PTMLE and the shrinkage MLE both have very close ADR , both converging to a common limit [$\mathrm{Tr}(\underset{\sim}{W}\underset{\sim}{\overline{B}}^{-1}_{\underset{\sim}{\theta}_0})$] as $\underset{\sim}{\gamma}^*$ moves away from the origin (i.e., Δ or Δ^* converges to $+\infty$); however, in this case, the ADR of the shrinkage MLE always lies below the asymptote, while that of the PTMLE may be slightly above the asymptote for some intermediate values of $\underset{\sim}{\gamma}^*$. Further, the PTMLE does not need that $p \geq 3$ (as needed by the shrinkage MLE), and hence, for smaller values of p (say, less than 6), the PTMLE may have an edge over the shrinkage MLE. On the other hand, as p becomes larger, the shrinkage MLE seems to have a more dominating picture.

We conclude this section with some remarks on a general form of the shrinkage MLE. We may rewrite (2.7) as

$$\hat{\underset{\sim}{\theta}}_n^S = \hat{\underset{\sim}{\theta}}_n + [1 - (p-2)/\mathcal{L}_n](\tilde{\underset{\sim}{\theta}}_n - \hat{\underset{\sim}{\theta}}_n). \tag{4.14}$$

In line with the general shrinkage estimator of the normal mean vector with unknown covariance matrix, we may replace the scalar factor in the second term on the right hand side of (4.14) by an appropriate matrix, and consider a general shrinkage estimator as

$$\underset{\sim}{\theta}_n^* = \hat{\underset{\sim}{\theta}}_n + [\underset{\sim}{I} - c_n d_n \mathcal{L}_n^{-1} \underset{\sim}{W}^{-1} \underset{\sim}{S}_n^{-1})(\tilde{\underset{\sim}{\theta}}_n - \hat{\underset{\sim}{\theta}}_n), \tag{4.15}$$

where $\{c_n\}$ is a sequence of positive numbers converging to a positive c (as $n \to \infty$) with $c \in (0, 2(p-2))$,

$$d_n = \text{smallest characteristic root of } \underset{\sim}{W}\underset{\sim}{S}_n, \tag{4.16}$$

and $\underset{\sim}{S}_n$ is a BAN estimator of $\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1} - \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$. Ideally, c may be taken as p-2. In (4.2), we have tacitly assumed that $\underset{\sim}{S}_n\underset{\sim}{W}$ is of full rank, in probability. If $\underset{\sim}{W}\underset{\sim}{S}_n$ is not of full rank, we may work with a generalized inverse, and in that case, for d_n in (4.16), we take the smallest positive root. Replacing $\underset{\sim}{S}_n$ by $\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1} - \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o}$ in (4.16), the corresponding quantity is denoted by δ. Then,

$$\underset{\sim}{S}_n \overset{p}{\to} \bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1} - \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o} \quad \text{and} \quad d_n \overset{p}{\to} \delta, \text{ as } n \to \infty, \tag{4.17}$$

so that if we define

$$\underset{\sim}{\theta}_n^{**} = \hat{\underset{\sim}{\theta}}_n + [\underset{\sim}{I} - c\delta \mathcal{L}_n^{-1} \underset{\sim}{W}^{-1} (\bar{\underset{\sim}{B}}_{\underset{\sim}{\theta}_o}^{-1} - \bar{\underset{\sim}{P}}_{\underset{\sim}{\theta}_o})^{-1}](\tilde{\underset{\sim}{\theta}}_n - \hat{\underset{\sim}{\theta}}_n), \tag{4.18}$$

then, it follows by a direct use of the Slutzky theorem that under $\{K_n\}$,

$$n^{1/2} \| \underset{\sim}{\theta}_n^* - \underset{\sim}{\theta}_n^{**} \| \overset{p}{\to} 0, \text{ as } n \to \infty, \tag{4.19}$$

so that both the estimators in (4.15) and (4.18) would have the same A.D.R. On the other hand, for (4.18), we may use (3.10)-(3.13) and derive that under $\{K_n\}$ and the assumed regularity conditions ,

$$n^{\frac{1}{2}}(\underset{\sim}{\theta}_n^* - \underset{\sim}{\theta}_o) = (\bar{\underset{\sim}{P}}_{\theta_o}\underset{\sim}{\Lambda}_n^o + \underset{\sim}{\gamma}^*) + [\underset{\sim}{I} - c\delta\mathcal{L}_n^{-1}\underset{\sim}{W}^{-1}(\underset{\sim}{D}^*)^{-1}\bar{\underset{\sim}{B}}_{\theta_o}](\bar{\underset{\sim}{B}}_{\theta_o}^{-1}\underset{\sim}{D}^*\underset{\sim}{\Lambda}_n^o - \underset{\sim}{\gamma}^*) + \underset{\sim}{o}_p(1) \tag{4.20}$$

$$= (\underset{\sim}{D}')^{-1}\underset{\sim}{U}_n - c\delta(\underset{\sim}{W}^{-1}\underset{\sim}{D}\underset{\sim}{D}_o^{-1})(\underset{\sim}{D}_o\underset{\sim}{U}_n - \underset{\sim}{D}'\underset{\sim}{\gamma}^*)/(\underset{\sim}{D}_o\underset{\sim}{U}_n - \underset{\sim}{D}'\underset{\sim}{\gamma}^*)'(\underset{\sim}{D}_o\underset{\sim}{U}_n - \underset{\sim}{D}'\underset{\sim}{\gamma}^*) + \underset{\sim}{o}_p(1),$$

where we adopt the same notationa as in (3.13)-(3.15), and assume that $\underset{\sim}{W}$ is of full rank (note that $\underset{\sim}{D}_o$ and $\underset{\sim}{D}^*$ are idempotent matrices, so that their (generalized) inverses are the same matrices , and this adjustment is necessary when $\underset{\sim}{D}^*$ or $\underset{\sim}{D}$ is not of full rank). We define $\underset{\sim}{W}_o$ as in after (3.20), and proceed as in (3.21) through (3.24), and obtain that

$$\rho(\underset{\sim}{\theta}^* ; \underset{\sim}{W}) = \mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1}) - 2c\delta\{ E(\chi_{p+2}^{-2}(\Delta)\mathrm{Tr}(\underset{\sim}{D}_o) + (\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{D}^*\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{\gamma}^*)[E(\chi_{p+2}^{-2}(\Delta)) +$$
$$+ E(\chi_{p+4}^{-2}(\Delta))]\} + c^2\delta^2\{ E(\chi_{p+2}^{-4}(\Delta))\mathrm{Tr}(\underset{\sim}{W}_o^{-1}\underset{\sim}{D}_o) +$$
$$E(\chi_{p+4}^{-4}(\Delta))(\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{D}^*\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{W}^{-1}\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{D}^*\underset{\sim}{\gamma}^*)\} . \tag{4.21}$$

This expression is quite similar to that of (3.25), and parallel to (4.4), it can be shown that $\underset{\sim}{\theta}_n^*$ dominates over the unrestricted MLE $\tilde{\underset{\sim}{\theta}}_n$ in the light of the ADR. The discussions following (4.7) (through (4.9)) also pertain to the case of $\rho(\underset{\sim}{\theta}^*;\underset{\sim}{W})/\rho(\hat{\underset{\sim}{\theta}};\underset{\sim}{W})$ (under H_0), and this shows that $\underset{\sim}{\theta}_n^*$ may not dominate over the retricted MLE when H_0 holds. The same conclusion holds for the PTMLE. However, the relative picture of (3.25) and (4.21) is a lot more complicated and depends heavily on all the matrices appearing in the two expressions. In particular, if we let $\underset{\sim}{W} = \bar{\underset{\sim}{B}}_{\theta_o}$, then we have $\mathrm{Tr}(\underset{\sim}{W}\bar{\underset{\sim}{B}}_{\theta_o}^{-1}) = p$, δ = smallest positive characteristic root of $\underset{\sim}{D}^*$ (= 1), $\mathrm{Tr}(\underset{\sim}{D}_o)$ = rank of $\underset{\sim}{D}^*$, $\mathrm{Tr}(\underset{\sim}{W}_o^{-1}\underset{\sim}{D}_o)$ = $\mathrm{Tr}(\underset{\sim}{D}_o)$ = rank of $\underset{\sim}{D}_o$ and $\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{D}^*\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{\gamma}^* = \underset{\sim}{\gamma}^{*\prime}\underset{\sim}{D}^*\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{W}^{-1}\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{D}^*\underset{\sim}{\gamma}^* = \underset{\sim}{\gamma}^{*\prime}\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{\gamma}^* - \underset{\sim}{\gamma}^{*\prime}\bar{\underset{\sim}{B}}_{\theta_o}\bar{\underset{\sim}{P}}_{\theta_o}\bar{\underset{\sim}{B}}_{\theta_o}\underset{\sim}{\gamma}^* \leq \Delta$, so that the two expressions become very much comparable.

R E F E R E N C E S

CRAMÉR, H. (1946). Mathematical Methods of Statistics. Princeton Univ. Press, New Jersey.

HUBER, P.J. (1967). The behavior of maximum likelihood estimators under non-standard conditions. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1, 221-234.

IBRAGIMOV, I.A. AND HAS'MINSKII, R.Z. (1981). Statistical Estimation: Asymptotic Theory. Springer-Verlag, New York.

INAGAKI, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. Ann. Inst. Statist. Math. 25, 1-26.

JUDGE, G.G. AND BOCK, M.E. (1978). The Statistical Implications of Pretest and Stein-Rule Estimators in Econometrics. North - Holland, Amsterdam.

SALEH, A.K.Md.E. AND SEN, P.K. (1978). Nonparametric estimation of location parameter after a preliminary test on regression. Ann. Statist. 6, 154-168.

SALEH, A.K.Md.E. AND SEN, P.K. (1984). Least squares and rank order preliminary test estimation in general multivariate linear models. Proc. Indian Statist. Inst., Golden Jub. Confer. Statist. Appl. New Direc. (ed. J.K.Ghosh),237-253.

SEN, P.K. (1979). Asymptotic properties of maximum likelihood estimators based on conditional specifications. Ann. Statist. 7, 1019-1033.

SEN, P.K. (1984). A James-Stein detour of U-statistics. Commun. Statist. Theor. Meth. A13 , 2725-2747.

SEN, P.K. AND SALEH, A.K.Md.E. (1979). Nonparametric estimation of location parameter after a preliminary test on regression in the multivariate case. Jour. Multivar. Anal. 9 , 322-331.

SEN, P.K. AND SALEH, A.K.Md.E. (1985). On some shrinkage estimators of multi-variate location. Ann. Statist. 13, in press.