Pretest Estimation of Eigenvalues of a Wishart Matrix

S. E. Ahmed, A. I. Volodin
Department of Mathematics and Statistics
University of Regina
Regina, Saskatchewan
S4S 0A2 Canada
Internet: ahmed@math.uregina.ca
and
A. H. Joarder
School of Mathematics and Statistics
University of Sydney, NSW, Australia 2006

Abstract

In this paper, we discuss various large sample estimation techniques for eigen values of Wishart matrix. Pret est estimators are proposed and these are compared with the sample eigen values using asymptotic quadratic distributional risk. The relative dominance picture of the proposed estimators is presented. A criteria for the selection of the size of pretest is discussed. It is shown that the significance level for the proposed pretest estimator often coincides with the commonly used level of significance.

Key Words and Phrases: Wishart distribution, eigen values, covariance matrix, pretest estimators, asymptotic quadratic bias and risk.

1. INTRODUCTION

The main purpose of this investigation is to consider the problem of estimating the eigen values of the scale matrix $\mathbf{\Sigma}$ of a Wishrat distribution. Let the sample covariance matrix denoted by $\mathbf{S}$ has a non-singular Wishart distribution with unknown covariance matrix $\mathbf{\Sigma}$ and n degrees of freedom, i.e.,

$$n\mathbf{S} \sim \mathbf{W}_p(n, \mathbf{\Sigma}).$$

Let $\xi_1, \c{d}ots, \xi_p (> 0)$ denote the distinct roots, in the population, of the determinantal equation $|\mathbf{\Sigma} - \boldsymbol{\xi}\mathbf{I}| = \mathbf{0}$. Here we are primarily interested in the estimation of parameter vector $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_p)'$ when the *nonsample information (NSI)* given by

$$\boldsymbol{\xi} = \boldsymbol{\xi}_o, \quad \boldsymbol{\xi}_o = (\xi_{o_1}, \cdots, \xi_{o_p})', \tag{1.1}$$

may hold. Here $\boldsymbol{\xi}_o$ is a prior guessed valued vector of eigen values which may be obtained from the past experience of the experimenter. Thus, we are primarily interested in the estimation of $\boldsymbol{\xi}$ when we may have *nonsample* or *uncertain prior information (UPI)* in (1.1).

Various authors including James and Stein (1961), Olkin and Selliah (1977) and Dey and Srinivasan (1986) considered the problem of estimating $\mathbf{\Sigma}$ directly by perturbing the eigen values of $\mathbf{S}$. However for small samples, Dey (1988) estimated the eigen values directly by shrinking or expanding the sample eigen values towards their geometric means. Jin (1993) also considered the problem of simultaneous estimation of eigen values of multivariate normal covariance matrix and proposed a new class of estimators which is a generalization of Dey's result (1988). Leung (1992) considered the estimation of eigen values of the scale matrix of the multivariate F distribution. More recently, Joarder and Ahmed (1996) extended these results to a multivariate t distribution.

In the present investigation, emphasis is on a situation where sample size is taken large while the parameter vector is taken close to $\boldsymbol{\xi}_o$. In this context, the notion of *asymptotic distributional quadratic risk (ADQR)* will be used. This will enable us to study the large sample properties of the proposed pretest estimator to be defined in the next section.

Let the sample eigen values be $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \cdots, \tilde{\xi}_p)'$. The asymptotic distribution of $\tilde{\boldsymbol{\xi}}$ in the Guassian case is described by the following theorem.

Theorem: L et $\boldsymbol{\Upsilon}$ be the diagonal matrix of eigen values $\xi_1, \cdots, \xi_p$ of covariance matrix $\boldsymbol{\Sigma}$, then for large n, the quantity $n^{\frac{1}{2}}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi})$ is approximately $N_p(\mathbf{0}, 2\boldsymbol{\Upsilon}^2)$.

For proof we refer to Anderson (1963) and Girshick (1939). The above theorem implies that, for large n, the $\tilde{\xi}_i$, $i = 1, \cdots, p$ are independently distributed. Furthermore, $\tilde{\xi}_i$, has an Gaussian distribution with mean ξ_i and variance $2\xi_i^2/n$ respectively.

Alternatively, one can use the variance stabilizing transformation i.e., $log\tilde{\boldsymbol{\xi}} = (log\tilde{\xi}_1, \cdots, log\tilde{\xi}_p)'$, then the limiting distribution of $n^{\frac{1}{2}}(log\tilde{\boldsymbol{\xi}} - log\boldsymbol{\xi})$ is approximately $N_p(\mathbf{0}, 2\mathbf{I})$, where $\mathbf{I}$ is an identity matrix of order $p \times p$. The rest of the paper is organized as follows. Section 2 proposes pretest estimators of $\boldsymbol{\xi}$. The expressions for the quadratic bias and ADQR are provided in section 3.

The properties of the propo sed estimators and their comparison with the sample eigen values are given in section 4. Section 5 deals with the size of the test.

2. PROPOSED PRETEST ESTIMATION

The statistical objective is to estimate parameter vector $\boldsymbol{\xi}$ simultaneously when UPI is available. Note that the $\tilde{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ is based on sample data only, and does not incorporate the nonsample information in estimating $\boldsymbol{\xi}$. However, it may be advantageous to use the available nonsample information to obtain improved estimates. In the following subsections, we introduce two improved estimation methodologies.

2.1. Shrinkage Estimator

It is reasonable to shrink $\tilde{\boldsymbol{\xi}}$ towards $\boldsymbol{\xi}_o$ (Thompson, 1968). Thus, a *shrinkage estimator (SE)* of $\boldsymbol{\xi}$ is defined by

$$\hat{\boldsymbol{\xi}} = \tilde{\boldsymbol{\xi}} - (1 - \pi)(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o),$$

where $\pi \in (0, 1)$ is a coefficient reflecting degree of distrust in the prior information. Note that $\hat{\boldsymbol{\xi}}$ is a convex combination of ue and $\boldsymbol{\xi}_o$ via fixed value of π.

2.2. Standard Preliminary Test Estimator

For the pretest on $H_o : \boldsymbol{\xi} = \boldsymbol{\xi}_o$, we consider the test statistic: $\mathcal{Q}_n = \mathbf{y}'\tilde{\boldsymbol{\Omega}}^{-1}\mathbf{y}$, where $\mathbf{y} = \sqrt{n}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)$ and $\tilde{\boldsymbol{\Omega}} = 2\tilde{\boldsymbol{\Upsilon}}^2$ with $\tilde{\boldsymbol{\Upsilon}}$ be the diagonal matrix of sample eigen values $\tilde{\xi}_1, \cdots, \tilde{\xi}_p$ of sample covariance matrix $\mathbf{S}$. The usual *pretest estimator (PTE)* of $\boldsymbol{\xi}$ denoted by $\hat{\boldsymbol{\xi}}^{(P)} = (\xi_1^P, \cdots \xi_k^P)$ is obtained by replacing π by $I(\mathcal{Q}_n \leq q_{n,\alpha})$ in $\hat{\boldsymbol{\xi}}$ to have a random weight. Thus,

$$\hat{\boldsymbol{\xi}}^{(P)} = \tilde{\boldsymbol{\xi}} - (\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)I(\mathcal{Q}_n \leq q_{n,\alpha}),$$

where $I(A)$ is the indicator function of the set A and $q_{n,\alpha}$ be the upper $100\alpha\%$ $(0 < \alpha < 1)$ point of the test statistic.

2.3. Improved Pretest Estimator

The *improved preliminary test estimator (IPTE)* of $\boldsymbol{\xi}$ denoted by $\hat{\boldsymbol{\xi}}^{(I)}$is

$$\hat{\boldsymbol{\xi}}^{(I)} = \tilde{\boldsymbol{\xi}} - (1 - \pi)(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)I(\mathcal{Q}_n \leq q_{n,\alpha}).$$

The value of π may be completely determined by the experimenter, depending upon the degree of disbelief in the NSI. A value near 1 causes $\hat{\boldsymbol{\xi}}^{(I)}$ to be based essentially on the sample data alone. More specifically, for $\pi = 0$, $\hat{\boldsymbol{\xi}}^{(I)} = \hat{\boldsymbol{\xi}}^{(P)}$ and for $\pi = 1$, $\hat{\boldsymbol{\xi}}^{(I)} = \tilde{\boldsymbol{\xi}}$.

3. USEFUL ASYMPTOTIC RESULTS

In this section, we obtain the expressions for the quadratic bias and risk of the estimators. Let $\boldsymbol{\xi}^o$ be any estimator of $\boldsymbol{\xi}$, we use a quadratic loss function:

$$\mathcal{L}(\boldsymbol{\xi}^o) = n(\boldsymbol{\xi}^o - \boldsymbol{\xi})'\boldsymbol{\Gamma}(\boldsymbol{\xi}^o - \boldsymbol{\xi}),$$

where $\boldsymbol{\Gamma}$ is a positive semidefinite weighting matrix. Dey (1988) and Leung (1992) advocated this loss function with $\boldsymbol{\Gamma} = \mathbf{I}$. The *quadratic risk* for $\boldsymbol{\xi}^o$ is given by $\mathcal{R}(\boldsymbol{\xi}^o) = nE\{(\boldsymbol{\xi}^o - \boldsymbol{\xi})'\boldsymbol{\Gamma}(\boldsymbol{\xi}^o - \boldsymbol{\xi})\}$. Now we define a sequence $\{K_{(n)}\}$ of local alternatives as

$$K_{(n)} : \boldsymbol{\xi} = \boldsymbol{\xi}_{(n)}, \quad \text{where} \quad \boldsymbol{\xi}_{(n)} = \boldsymbol{\xi}_o + \frac{\boldsymbol{\lambda}}{\sqrt{n}}. \tag{3.1}$$

We compute the *asymptotic distributional quadratic risk (ADQR)* defined below. First, the *asymptotic distribution function* of $\{\sqrt{n}(\boldsymbol{\xi}^o - \boldsymbol{\xi}_{(n)})\}$ is given by

$$G(\mathbf{z}) = \lim_{n\to\infty} Pr\{\sqrt{n}(\boldsymbol{\xi}^o - \boldsymbol{\xi}_{(n)}) \leq \mathbf{z}\},$$

for which the limit in the above relation exists. Further, the dispersion matrix of the distribution G is $\mathbf{V}$. Finally, the ADQR is defined by $R(\boldsymbol{\xi}^o) = tr(\boldsymbol{\Gamma}\mathbf{V})$, where $tr(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$.

In the case of fixed alternatives,

$$
\begin{aligned}
n(\hat{\boldsymbol{\xi}}^{(P)} - \tilde{\boldsymbol{\xi}})'\boldsymbol{\Gamma}(\hat{\boldsymbol{\xi}}^{(P)} - \tilde{\boldsymbol{\xi}}) &= n(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)'\boldsymbol{\Gamma}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)I(\mathcal{Q}_n < q_{n,\alpha}) \\
&= \mathcal{Q}_n I(\mathcal{Q}_n < q_{n,\alpha})\frac{n(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)'\boldsymbol{\Gamma}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)}{n(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)'\tilde{\boldsymbol{\Omega}}^{-1}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_o)} \\
&\leq \{\mathcal{Q}_n I(\mathcal{Q}_n < q_{n,\alpha})\} ch_{\max}(\boldsymbol{\Gamma}\tilde{\boldsymbol{\Omega}}^{-1}) \\
&\leq \{\mathcal{Q}_n I(\mathcal{Q}_n < q_{n,\alpha})\} trace(\boldsymbol{\Gamma}\tilde{\boldsymbol{\Omega}}^{-1}) \qquad 3.2
\end{aligned}
$$

where $ch_{\max}(\mathbf{A})$ is the largest characteristic root of the matrix $\mathbf{A}$. Also, for $\boldsymbol{\xi} \notin H_o$, $E\{\mathcal{Q}_n I(\mathcal{Q}_n < q_{n,\alpha})\} \leq q_{n,\alpha}\{P(\mathcal{Q}_n < q_{n,\alpha})\}$. But the test statistic $\mathcal{Q}_n$ is consistent, hence $E\{\mathcal{Q}_n I(\mathcal{Q}_n < q_{n,\alpha})\} \to 0$ as $n \to \infty$. Thus, for fixed alternative $\tilde{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^{(P)}$ have asymptotically the same bounded risk. Finally, for any $\boldsymbol{\xi} \notin H_o$, $(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \xrightarrow{a.s.} \zeta(\neq 0)$, and

$$n(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\boldsymbol{\Gamma}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \xrightarrow{p} +\infty, \quad \text{as} \quad n \to \infty.$$

The asymptotic risk of $\hat{\boldsymbol{\xi}}$, for any $\boldsymbol{\xi} \notin H_o$, approaches $+\infty$. However, the asymptotic risk of $\tilde{\boldsymbol{\xi}}$ is bounded for every $\boldsymbol{\xi} \in \Omega$. The followin g theorem summarizes the results.

Theorem 3.1: When $\boldsymbol{\xi} \notin H_o, \hat{\boldsymbol{\xi}}$ has asymptotic risk of $+\infty$, while $\hat{\boldsymbol{\xi}}^{(P)}$ and $\tilde{\boldsymbol{\xi}}$ have the same finite asymptotic risk.

For this reason we consider a sequence $\{K_{(n)}\}$ of local alternatives in (3.1). First, we have the following lemma.

Lemma 3.1: Under local alternatives as $n \to \infty$, $\mathbf{y}$ follows approximately a multivariate Guassian distribution with mean $\boldsymbol{\lambda}$ and covariance matrix $2\boldsymbol{\Upsilon}^2$.

As a consequence of Lemma 1, under local alternatives, the test statistic $\mathcal{Q}_n$ follows asymptotically a non-central chi-square distribution with p degrees of freedom and non-centrality parameter $\Lambda = \frac{1}{2}\boldsymbol{\lambda}'\boldsymbol{\Upsilon}^{-2}\boldsymbol{\lambda}$. Thus, under the null hypothesis, $\boldsymbol{\xi} = \boldsymbol{\xi}_o$, $\mathcal{Q}_n$ will have a central chi-square distribution with p degrees of freedom.

Now, we present the expressions for the *asymptotic distributional biases (ADB)* of the estimators as fo llows. The ADB of an estimator $\boldsymbol{\xi}^o$ is defined as

$$ADB(\boldsymbol{\xi}^o) = \lim_{n\to\infty} E\{n^{\frac{1}{2}}(\boldsymbol{\xi}^o - \boldsymbol{\xi})\}.$$

Using the above definition of the ADB, under $\{K_n\}$ in (3.1), as $n \to \infty$,

$$
\begin{aligned}
ADB(\tilde{\boldsymbol{\xi}}) &= \mathbf{0}, \\
ADB(\hat{\boldsymbol{\xi}}) &= -(1-\pi)\boldsymbol{\lambda}, \\
ADB(\hat{\boldsymbol{\xi}}^{(P)}) &= -\boldsymbol{\lambda}\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda), \\
ADB(\hat{\boldsymbol{\xi}}^{(I)}) &= -(1-\pi)\boldsymbol{\lambda}\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda),
\end{aligned}
$$

where the notation $\Phi_p(x\ ;\Lambda)$ stands for the noncentral chi-square distribution function with noncentrality parameter Λ and p degrees of freedom.

The $ADB(\hat{\boldsymbol{\xi}}^{(I)})$ is obtained by direct computation and using the same argument as in Section 4.3 of Judge and Bock (1978). Consequently, for $\pi = 0$ we obtain ADB expression for $\hat{\boldsymbol{\xi}}^{(P)}$.

Further, we transform these functions in a scalar (quadratic) form by defining

$$B(.) = [ADB(\boldsymbol{\xi}^o)]'\frac{1}{2}\boldsymbol{\Upsilon}^{-2}[ADB(\boldsymbol{\xi}^o)]$$

as quadratic bias of an estimator $\boldsymbol{\xi}^o$ of parameter vector $\boldsymbol{\xi}$. Thus, by the above definition, we have the following

$$\begin{aligned} B(\tilde{\boldsymbol{\xi}}) &= 0, \\ B(\hat{\boldsymbol{\xi}}) &= (1-\pi)^2\Lambda, \\ B(\hat{\boldsymbol{\xi}}^{(P)}) &= \Lambda\{\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\}^2, \\ B(\hat{\boldsymbol{\xi}}^{(I)}) &= (1-\pi)^2\Lambda\{\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\}^2. \end{aligned}$$

We notice that only $\tilde{\boldsymbol{\xi}}$ is an asymptotically unbiased estimator of ξ. However, if the nonsample information is correct, then $\Lambda = 0$ and all the remaining estimators are also unbiased. For $\Lambda > 0$, the quadratic bias of $\hat{\boldsymbol{\xi}}$ is a function of Λ and is unbounded in Λ which goes to ∞ as Λ tends to ∞. On the other hand, quadratic bias function of all the remaining estimators are bounded in Λ. Noting that as $\Lambda \to \infty$

$$\frac{\Lambda e^{-\frac{\Lambda}{2}}(\frac{\Lambda}{2})^r}{r!} \to 0.$$

As Λ increases, $B(\hat{\boldsymbol{\xi}}^{(I)})$ increases monotonically at first, reaches a maximum and then monotonically decreases towards zero. Hence it is a bounded function of Λ. It is clear that $B(\hat{\boldsymbol{\xi}}^{(I)}) = (1-\pi)^2[B(\hat{\boldsymbol{\xi}}^{(P)})]$. Since $0 < \pi < 1$, $B(\hat{\boldsymbol{\xi}}^{(I)}) < B(\hat{\boldsymbol{\xi}}^{(P)})$. Thi s indicates that $\hat{\boldsymbol{\xi}}^{(I)}$ has an edge over $\hat{\boldsymbol{\xi}}^{(P)}$ from the quadratic bias point of view. Thus, $\hat{\boldsymbol{\xi}}^{(I)}$ can be viewed as a quadratic bias reduction technique over the usual pretest estimation.

Under local alternatives, we obtain the $ADQR$ functions of the estimators in the following theorem.

Theorem 3.2:

$$ADQR(\tilde{\boldsymbol{\xi}}) = 2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2), \tag{3.3a}$$

$$ADQR(\hat{\boldsymbol{\xi}}) = 2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2) - (1-\pi^2)2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2) + (1-\pi)^2\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \tag{3.3b}$$

$$ADQR(\hat{\boldsymbol{\xi}}^{(P)}) = 2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2) - 2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) + \boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda}\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) - \Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}, \tag{3.3c}$$

$$ADQR(\hat{\boldsymbol{\xi}}^{(I)}) = 2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2) - (1-\pi^2)2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) +$$
$$\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda}(1-\pi)\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) - (1+\pi)\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}, \qquad 3.3d$$

Proof. The computation of (3.3a) and (3.3b) are straightforward. After some tedious algebra and by the use of Lemma 3.1, the relations (3.3c) and (3.3d) are obtained with th e same arguments as in Section 4.3 of Judge and Bock (1978)

4. RISK COMPARISON OF THE ESTIMATORS

We now investigate the statistical properties of the various estimators using $ADQR$ functions and determine their dominance characteristics.

Comparison of $\tilde{\boldsymbol{\xi}}$ *and* $\hat{\boldsymbol{\xi}}$ First, note that $\tilde{\boldsymbol{\xi}}$ has a constant risk since it is unrelated to the nonsample information. The $R(\hat{\boldsymbol{\xi}})$ is an unbounded function of $\boldsymbol{\lambda}$. However, it is superior to $\tilde{\boldsymbol{\xi}}$ near the null hypothesis. It is seen that

$$R(\hat{\boldsymbol{\xi}}) \le R(\tilde{\boldsymbol{\xi}}) \quad \text{if} \quad \boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \le 2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)(1+\pi)(1-\pi)^{-1}.$$

Specifically, if $\boldsymbol{\lambda}$ is a null vector, that is under the null hypothesis, $\hat{\boldsymbol{\xi}}$ is superior to $\tilde{\boldsymbol{\xi}}$. Using the Mahalanobis (squared) distance asthe loss function, that is putting $\boldsymbol{\Gamma} = (2\boldsymbol{\Upsilon}^2)^{-1}$, $2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2) = p$ and $\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} = \Lambda$. For the above chice of $\boldsymbol{\Gamma}$ we will have

$$R(\hat{\boldsymbol{\xi}}) \le R(\tilde{\boldsymbol{\xi}}) \Longleftrightarrow \Lambda \in \left[0,\ \frac{p(1+\pi)}{(1-\pi)}\right]$$

and

$$R(\tilde{\boldsymbol{\xi}}) \le R(\hat{\boldsymbol{\xi}}) \Longleftrightarrow \Lambda \in \left(\frac{p(1+\pi)}{(1-\pi)},\ \infty\right).$$

Clearly, when Λ moves away from H_o beyond the value $\frac{p(1+\pi)}{(1-\pi)}$, the risk of $\hat{\boldsymbol{\xi}}$ increases and becomes unbounded. This clearly indicates that the performance of $\hat{\boldsymbol{\xi}}$ will strongly depend on the reliability of the nonsample information. The performance of $\tilde{\boldsymbol{\xi}}$ is always steady throughout $\Lambda \in [0, \infty)$.

Remark 1. In the light of above discussions, we may conclude that none of the two estimators $\tilde{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}$ dominate the other asymptotically.

However, under H_o, $\hat{\boldsymbol{\xi}} \prec \tilde{\boldsymbol{\xi}}$, where the notation $\prec$ stands for dominance. However, in reality one do not know whether H_o holds or not and generally the value of Λ is unknown. The above remark and conclusions drawn in sequel are rather of theoretical nature, which serve the purpose of the this investigation.

Comparison of $\tilde{\boldsymbol{\xi}}$ *with* $\hat{\boldsymbol{\xi}}^{(I)}$ *and* $\hat{\boldsymbol{\xi}}^{(P)}$

First, note that

$$\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda) \le \Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) \le \Phi_{p+2}(\chi^2_{p,\alpha};0) = 1-\alpha,$$

for $\alpha \in (0,1)$ and $\Lambda > 0$. The left hand side of the above relation converges to 0 as $\Lambda \to \infty$. Also, as $||\boldsymbol{\lambda}|| \to \infty \;\Rightarrow\; \Lambda \to \infty$, then $\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)$, $\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda}\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)$ and $\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda}\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)$ approach 0, and the risk of $\hat{\boldsymbol{\xi}}^{(I)}$ approaches $2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)$, i.e., the risk of $\tilde{\boldsymbol{\xi}}$. The risk of $\hat{\boldsymbol{\xi}}^{(I)}$ is smaller than the risk of $\tilde{\boldsymbol{\xi}}$ near the null hypothesis which keeps on increasing, crosses the line $2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)$, reaches to maximum then decreases monotonically to the risk of $\tilde{\boldsymbol{\xi}}$. Hence a pretest approach controls the magnitude of the risk. In fact, $\hat{\boldsymbol{\xi}}^{(I)}$ dominates $\tilde{\boldsymbol{\xi}}$ if $\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \in [0, u_1]$ where,

$$u_1 = \frac{2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)(1+\pi)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)}{\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) - (1+\pi)\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}}.$$

There are points in the parameter space for which $\hat{\boldsymbol{\xi}}^{(I)}$ is inferior to $\tilde{\boldsymbol{\xi}}$ and a sufficient condition is $\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \in (u_1, \infty)$. Moreover, as α (the le vel of significance of pretest) tends to 1, $R(\hat{\boldsymbol{\xi}}^{(I)})$ tends to $R(\tilde{\boldsymbol{\xi}})$. At $\Lambda = 0$, the $R(\hat{\boldsymbol{\xi}}^{(I)})$ assumes value $2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)[1-(1-\pi^2)\Phi_{p+2}(\chi^2_{p,\alpha};0)]$, then keeps on increasing crossing the line $2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)$, reaches to maximum then decreases monotonically to the $R(\tilde{\boldsymbol{\xi}})$.

For $\pi = 0$, we obtain the comparison of $\tilde{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^{(P)}$. Thus, $\hat{\boldsymbol{\xi}}^{(P)}$ performs better than $\tilde{\boldsymbol{\xi}}$ whenever $\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \in [0, u_2]$ where,

$$u_2 = \frac{2tr(\boldsymbol{\Gamma}\boldsymbol{\Upsilon}^2)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)}{\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) - \Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}},$$

and for $\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \in (u_2, \infty)$ opposite conclusion holds. Further, by comparing the u_1 and u_2 we find that $\hat{\boldsymbol{\xi}}^{(I)}$ provides a wider range than $\hat{\boldsymbol{\xi}}^{(P)}$ in which it has smaller risk than $\tilde{\boldsymbol{\xi}}$. This indicates the superiority of $\hat{\boldsymbol{\xi}}^{(I)}$ over $\hat{\boldsymbol{\xi}}^{(P)}$ in sense of dominance range. We will also demonstrate later in this paper that $\hat{\boldsymbol{\xi}}^{(I)}$ has an edge over $\hat{\boldsymbol{\xi}}^{(P)}$ with respect to the size of the pretest. This important fact was first noticed by Ahmed (1992a, 1992b).

We observe that performance of the pretest estimators, which combine sample informat ion with NSI, heavily depend on the correctness of the NSI. The gain in the risk is substantial over classical procedure when information is correct or nearly correct. However, $\hat{\boldsymbol{\xi}}^{(I)}$ and $\hat{\boldsymbol{\xi}}^{(P)}$ combine the information in a superior way than that of $\hat{\boldsymbol{\xi}}$ in the sense that their risk is a bounded function of the NSI.

Remark 2 None of the three estimators is inadmissible with respect to others. However, when the null hypothesis is true then the risks of the estimators may be ordered according to the magnitude of their risk as follows:

$$\hat{\boldsymbol{\xi}}^{(P)} \prec \hat{\boldsymbol{\xi}}^{(I)} \prec \tilde{\boldsymbol{\xi}}.$$

Comparison of $\hat{\boldsymbol{\xi}}$ *and* $\hat{\boldsymbol{\xi}}^{(I)}$

When the nonsample information is correct, then the risk difference

$$R(\hat{\boldsymbol{\xi}}^{(I)}) - R(\hat{\boldsymbol{\xi}}) = (1-\pi^2)2tr(\boldsymbol{\Gamma\Upsilon}^2)\{1-\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\} \geq 0.$$

This clearly indicates superiority of $\hat{\boldsymbol{\xi}}$ over $\hat{\boldsymbol{\xi}}^{(I)}$ at the null hypothesis. However, under local alternative, the risk difference indicates that $\hat{\boldsymbol{\xi}}$ will be superior to $\hat{\boldsymbol{\xi}}^{(I)}$ if

$$\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \leq \frac{(1+\pi)\{2tr(\boldsymbol{\Gamma\Upsilon}^2) - 2tr(\boldsymbol{\Gamma\Upsilon}^2)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\}}{\{(1-\pi) - 2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) + (1+\pi)\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}}.$$

Let us consider $\boldsymbol{\Gamma} = (2\boldsymbol{\Upsilon}^2)^{-1}$, then in term of Λ, $\hat{\boldsymbol{\xi}}$ is superior to $\hat{\boldsymbol{\xi}}^{(I)}$ if

$$0 \leq \Lambda \leq \frac{(1+\pi)p\{1-\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\}}{\{(1-\pi) - 2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) + (1+\pi)\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}},$$

while opposite holds if

$$\frac{(1+\pi)p\{1-\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\}}{\{(1-\pi) - 2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) + (1+\pi)\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}} < \Lambda < \infty.$$

The proposed estimators $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^{(P)}$ both use the data and NSI, however, neither $\hat{\boldsymbol{\xi}}^{(I)}$ nor $\hat{\boldsymbol{\xi}}$ is superior with respect to each other.

Remark 3 Under the null hypothesis $\hat{\boldsymbol{\xi}} \prec \hat{\boldsymbol{\xi}}^{(I)}$.

Comparison of $\hat{\boldsymbol{\xi}}^{(I)}$ *and* $\hat{\boldsymbol{\xi}}^{(P)}$

We now compare the risk performance of the $\hat{\boldsymbol{\xi}}^{(I)}$ and $\hat{\boldsymbol{\xi}}^{(P)}$ and determine the conditions under which $\hat{\boldsymbol{\xi}}^{(I)}$ performs better than $\hat{\boldsymbol{\xi}}^{(P)}$. First, under the null hypothesis $R(\hat{\boldsymbol{\xi}}^{(I)}) - R(\hat{\boldsymbol{\xi}}^{(P)}) = \pi^2 2tr(\boldsymbol{\Gamma\Upsilon}^2)\Phi_{p+2}(\chi^2_{p,\alpha};0) > 0$. Thus, under the null hypothesis $\hat{\boldsymbol{\xi}}^{(P)}$ is superior to $\hat{\boldsymbol{\xi}}^{(I)}$. However, the risk difference may not be noticeable for the smaller values of π. Alternatively, when the hypothesis error grows then $\hat{\boldsymbol{\xi}}^{(I)}$ will be superior to $\hat{\boldsymbol{\xi}}^{(P)}$ in the rest of the parameter space. More specifically,

$$R(\hat{\boldsymbol{\xi}}^{(I)}) < R(\hat{\boldsymbol{\xi}}^{(P)}) \Longleftrightarrow \boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} \geq \pi 2tr(\boldsymbol{\Gamma\Upsilon}^2)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) - \pi\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}^{-1}.$$

Let us consider the situation $\boldsymbol{\Gamma} = (2\boldsymbol{\Upsilon}^2)^{-1}$, then $\hat{\boldsymbol{\xi}}^{(P)}$ is superior to $\hat{\boldsymbol{\xi}}^{(I)}$ if

$$0 \leq \Lambda \leq \pi p\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) - \pi\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}^{-1},$$

while opposite result holds if

$$\pi p\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda) - \pi\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}^{-1} < \Lambda < \infty.$$

Let Λ_π be a point in the parameter space at which the risk of $\hat{\boldsymbol{\xi}}^{(I)}$ and $\hat{\boldsymbol{\xi}}^{(P)}$ intersect for a given π. Then, for $\Lambda \in (0, \Lambda_\pi]$, $\hat{\boldsymbol{\xi}}^{(P)}$ performs better than $\hat{\boldsymbol{\xi}}^{(I)}$, while for $\Lambda \in (\Lambda_\pi, \infty)$, $\hat{\boldsymbol{\xi}}^{(I)}$ is superior to $\hat{\boldsymbol{\xi}}^{(P)}$. Further, for large values of π (close to 1), the interval $(0, \Lambda_\pi]$ may be negligible. Nevertheless, $\hat{\boldsymbol{\xi}}^{(I)}$ and $\hat{\boldsymbol{\xi}}^{(P)}$ share a common asymptotic property that, as $\Lambda \to \infty$, their risk converge to a common limit, i.e., to the risk of $\tilde{\boldsymbol{\xi}}$ from the above.

Remark 4 None of $\hat{\boldsymbol{\xi}}^{(I)}$ and $\hat{\boldsymbol{\xi}}^{(P)}$ is inadmissible with respect to other. At $\Lambda = 0$, $\hat{\boldsymbol{\xi}}^{(P)} \prec \hat{\boldsymbol{\xi}}^{(I)}$.

Finally, by combining all the remarks we have made so far, we arrive at the following conclusion.

Conclusion None of the four estimators is inadmissible with respect to any of others. However, at $\Lambda = 0$, $\hat{\boldsymbol{\xi}} \prec \hat{\boldsymbol{\xi}}^{(P)} \prec \hat{\boldsymbol{\xi}}^{(I)} \prec \tilde{\boldsymbol{\xi}}$.

The statistical properties of the $\hat{\boldsymbol{\xi}}^{(I)}$ and hence that of $\hat{\boldsymbol{\xi}}^{(P)}$ depend, among other factors, on the size of the test chosen for the pretest which has not been given serious consideration. The size of the test plays an important role in selecting the estimator. Since the level of significance α is in the control of statistician, thus, we have a statistical decision problem for choosing α. In the following section, a method for the choice of α using efficiency criterion is discussed.

We demonstrate below that the optimal significance level is smaller for the proposed $\hat{\boldsymbol{\xi}}^{(I)}$ than that of $\hat{\boldsymbol{\xi}}^{(P)}$ and the smaller significance level often coincides with the traditional level of significance.

5. Size of the Pretest

One method to determine the value of α is to compute the minimum guaranteed asymptotic efficiency, a rule was first given in Han and Bancroft (1968) and extended by Ahmed (1992a, 1992b).

First, we introduce the notion of the asymptotic relative efficiency. The *asymptotic relative efficiency (ARE)* of an estimator $\boldsymbol{\xi}^\star$ to another estimator $\boldsymbol{\xi}^\diamond$ is defined by

$$ARE(\boldsymbol{\xi}^\star : \boldsymbol{\xi}^\diamond) = R(\boldsymbol{\xi}^\diamond, \boldsymbol{\Gamma})/R(\boldsymbol{\xi}^\star, \boldsymbol{\Gamma}).$$

Bear in mind that an ARE greater than 1 indicates the degree of asymptotic superiority of $\boldsymbol{\xi}^\star$ over $\boldsymbol{\xi}^\diamond$.

In order to facilitate numerical computation of risk functions of the various estimators, we consider the case $\boldsymbol{\Gamma} = (2\boldsymbol{\Upsilon}^2)^{-1}$ and then obtain the values of ARE on a digital computer .

Thus, the ARE of $\hat{\boldsymbol{\xi}}^{(I)}$ with respect to $\tilde{\boldsymbol{\xi}}$ is given by

$$ARE(\hat{\boldsymbol{\xi}}^{(I)}:\tilde{\boldsymbol{\xi}}) = [1-(1-\pi^2)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)+\frac{\Lambda}{p}\{2(1-\pi)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)-(1-\pi^2)\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}]^{-1}. \tag{5.1}$$

Note that for $\alpha = 0$ relation (5.1) defines the ARE of $\hat{\boldsymbol{\xi}}$ relative to $\tilde{\boldsymbol{\xi}}$ while for $\pi = 0$, we get the ARE of $\hat{\boldsymbol{\xi}}^{(P)}$ relative to $\tilde{\boldsymbol{\xi}}$. Further, when the null hypothesis is true then

$$ARE(\hat{\boldsymbol{\xi}}^{(I)}:\tilde{\boldsymbol{\xi}}) = ARE_{\max} = \left\{1-(1-\pi^2)\Phi_{p+2}(\chi^2_{p,\alpha};0)\right\}^{-1} > 1.$$

Note that for fixed α, π and p, $\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda) \leq \Phi_{p+2}(\chi^2_{p,\alpha};0)$. Thus, the maximum value of $ARE(\hat{\boldsymbol{\xi}}^{(I)}:\tilde{\boldsymbol{\xi}})$ occurs at $\Lambda = 0$. This maximum efficiency is decreasing function of α for fixed π and of π for fixed α. This picture is different for the non-null case. Noting that, for fixed values of α and π, $ARE(\hat{\boldsymbol{\xi}}^{(I)}:\tilde{\boldsymbol{\xi}})$ is a monotone decreasing function of Λ, where α and π are held constant. It crosses the line $ARE(\hat{\boldsymbol{\xi}}^{(I)}:\tilde{\boldsymbol{\xi}}) = 1$ at $\frac{p(1+\pi)\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)}{\{2\Phi_{p+2}(\chi^2_{p,\alpha};\Lambda)-(1+\pi)\Phi_{p+4}(\chi^2_{p,\alpha};\Lambda)\}}$, then decreases and attains a minimum value $ARE_{\min}$ at a point Λ_o and then increases asymptotically to 1. The minimum efficiency is an increasing function of α.

On the other hand, for any fixed value of α, the maximum value of the $ARE(\hat{\boldsymbol{\xi}}^{(I)}:\tilde{\boldsymbol{\xi}})$ is a decreasing function of π, while the minimum efficiency $ARE_{\min}$ is an increasing function of π. In order to determine the critical value for the pretest with minimum guaranteed efficiency $ARE_{\min}$, the researcher is willing to accept for fixed π, one needs to solve the equation

$$\sup_{\alpha}\left\{\inf_{\Lambda} ARE(\alpha,\pi,\Lambda)\right\} \geq ARE_{\min},$$

or

$$\sup_{\alpha_1}\left\{\inf_{\Lambda} ARE(\alpha,\pi,\Lambda)\right\} \geq ARE_{\min},$$

where α_1 is the required value. In the same way, by selecting $\pi = 0$, we obtain α_2 by using

$$\sup_{\alpha_2}\left\{\inf_{\Lambda} ARE(\alpha,0,\Lambda)\right\} \geq ARE_{\min},$$

where α_2 is the desired value. Finally, by the property of $\inf_{\Lambda} ARE_{\max}(\alpha,\pi,\Lambda)$, we observe that $\alpha_1 < \alpha_2$. Hence, $\hat{\boldsymbol{\xi}}^{(I)}$ has a remarkable edge over $\hat{\boldsymbol{\xi}}^{(P)}$ with respect to the size of the pretest. The use of the usual pretest estimation may be limited by the larger value of α, the level of significance. However, by employing a shrinkage technique one may control the value of size of the test.

5.1. Illustrative Example

In order to provide a numerical example we compute the table of maximum ($ARE_{\max}$) and minimum ($ARE_{\min}$) efficiencies along with corresponding value of the $\Lambda = \Lambda_{\min}$ at which the minimum efficiency occurred. The purpose of preparing the table is bi-fold; one is to appraise the behavior of the estimators and the other is to determine the optimum value of α for the pretest which provides the minimum guarantee d efficiency.

For an example, if $p = 2$ and the experimenter is looking for an estimator with a minimum ARE of at least 0.90, with $\pi = 0.6$ then from table 1 the value of α_1 is found to be 0.05 . Such a choice of α_1 would yield an estimator with a maximum efficiency of 2.05 at $\Lambda = 0$, with a minimum guaranteed efficiency of 0.91. If the experimenter selects $\pi = 0$, then from table 1 the size of the pretest, i.e., the value of α_2 will be approximately 0.30. Hence, $\hat{\boldsymbol{\xi}}^{(I)}$ outperforms $\hat{\boldsymbol{\xi}}^{(P)}$ with respect to the size of the preliminary test. Not only that, the maximum efficiency drops from 2.05 to 1.51. We conclude this section with the following remark.

Remark 5: The proposed $\hat{\boldsymbol{\xi}}^{(I)}$ has an edge over $\hat{\boldsymbol{\xi}}^{(P)}$ with respect to smaller level of significance as well as wider range of dominance over $\tilde{\boldsymbol{\xi}}$. Hence, $\hat{\boldsymbol{\xi}}^{(I)}$ is superior to $\hat{\boldsymbol{\xi}}^{(P)}$.

The computations for the table are carried out with a FORTRAN program.

ACKNOWLEDGMENTS

This work of first was supported by grant from the Natural Sciences and Engineering Research Council of Canada. The authors wish to thank Mr. Sohail M. Khan for preparing the table.

REFERENCES

Ahmed, S. E.(1992a). Shrinkage preliminary test estimation in multivariate normal distributions. Journal of Statistical Computation and Simulation **43**, 177-195.

Ahmed, S. E. (1992b). Large-sample pooling procedure for correlation. The Statistician **41**, 425-438.

Ahmed, S.E. and A.K.M.E. Saleh (1993). Improved estimation for the component mean-vector. Japan Journal of Statistics, **43**, 177-195.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. Annals of Mathematical Statistics, **34**, 122- 148.

Dey, D. K. (1988). Simultaneous estimation of eigen values. Ann Inst Statist Math, **40**, 137- 147.

Dey, D. K. and C. Srinivasan (1986). Trimmed minimax estimator of a covariance matrix. Ann Inst Statist Math, **38**, 47- 54.

Grischick, M. A. (1939). On the sampling theory of roots of determinantal equations. Annals of Mathematical Statistics, **10**, 203- 224.

Han, C.P. and T.A. Bancroft, (1968). On pooling means when variance is unknown. Journal of American Statistical Association, **63**, 1333-1342.

Joarder, A. H, and S. E. Ahmed (1996). Estimation of the characteristic roots of the scale matrix. Metrika, **44**, 259-267.

James W. and Stein C. (1961). Estimation with quadratic loss. Proceeding of the fourth Berkeley symposium on Mathematical statistics and Probability. University of California Press. 361-379.

Jin, C. (1993). A note on simultaneous estimation of eigen values of a multivariate normal covariance matrix. Statistics & Probaility Letters, **16**, 197-203.

Judge, G. G. and M. E. Bock (1978). The statistical implication of pre-test and Stein-rule estimators in econometrics. Amsterdam: North-Holland.

Lueng., P. L. (1992). Estimation of eigen values of the scale matrix of the multivariate F distribution. Communications in Statistics- Theory and methods, **21**, 1845-1856.

Olkin, I. and J. B. Selliah (1977). Estimating covariance matrix in a multivariate normal distribution. Statistical decision theory and related topics, II (eds. SS. Guppy and D. More), 313-326, Academic Press, New York.

Thompson, J.C. (1968). Some shrinkage techniques for estimating the mean. Journal of American Statistical Association **63**, 113-122.