Res. Lett. Inf. Math. Sci., 2005, Vol.**8**, pp77-96
Available online at http://iims.massey.ac.nz/research/letters/

Appoximation-assisted estimation of eigenvectors under quadratic loss

S.E. Ahmed[†]

Department of Mathematics and Statistics
University of Windsor
Ontario, Canada
[†]seahmed@uwindsor.ca

Improved estimation of eigen vector of covariance matrix is considered under uncertain prior information (*UPI*) regarding the parameter vector. Like statistical models underlying the statistical inferences to be made, the prior information will be susceptible to uncertainty and the practitioners may be reluctant to impose the additional information regarding parameters in the estimation process. A very large gain in precision may be achieved by judiciously exploiting the information about the parameters which in practice will be available in any realistic problem.

Several estimators based on preliminary test and the Stein-type shrinkage rules are constructed. The expressions for the bias and risk of the proposed estimators are derived and compared with the usual estimators. We demonstrate that how the classical large sample theory of the conventional estimator can be extended to shrinkage and preliminary test estimators for the eigenvector of a covariance matrix. It is established that shrinkage estimators are asymptotically superior to the usual sample estimators. For illustration purposes, the method is applied to three datasets.

1 Introduction

Principal component analysis (PCA) is a classical multivariate technique and it has become increasingly popular in multivariate statistical theory and applications. PCA takes its place in many applications such as detection, recognition, image processing and compression. Using PCA, the red, green and blue color channels of an image can be transformed into three different, de-correlated channels. PCA is widely used in signal processing and neural computing. In some application areas, this is also called the (discrete) Karhunen-Love transform, or the Hotelling transform. In biometric applications, principal components are frequently interpreted as independent factors determining the growth size and shape of an organism. Tipping and Bishop (1999) discuss the probabilistic approach to PCA. A comparative study of principal component analysis techniques with applications to neural networks is given in Calvo, Partridge and Jabri (1998). PCA involves a mathematical procedure that transform a set of correlated response variable into a smaller set of uncorrelated variables called *principal components*. We refer to book by Jolliffe (2002) for comprehensive treatment on the subject. More importantly PCA is a function of eigen vectors and hence the motivation of this paper. In the present investigation the problem of estimating the eigen vectors is considered. Noting that the eigen vectors is the essence of a principal component procedure. The eigen vectors determine the direction of the maximum variability. In order to estimate the population principal components we need to estimate the corresponding eigen vectors.

Let $\mathbf{X}$ be an $m \times 1$ random vector with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\mathbf{\Sigma}$. Let $\lambda_1, \lambda_2, \cdots, \lambda_m (> 0)$ be the characteristic roots interchangeably with eigen values of $\mathbf{\Sigma}$ and let $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m)$ be an $m \times m$ orthogonal matrix such that

$$\mathbf{E}'\mathbf{\Sigma}\mathbf{E} = \mathbf{\Lambda} \equiv \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m),$$

so that $\mathbf{e}_i$ is the ith eigen vector of $\mathbf{\Sigma}$ corresponding to eigen value λ_i. Let $Y_1 = \mathbf{e}_1'\mathbf{X},\ Y_2 = \mathbf{e}_2'\mathbf{X}, \cdots Y_m = \mathbf{e}_m'\mathbf{X}$, then $Cov(\mathbf{Y}) = \mathbf{\Lambda}$, so that $Y_1, \cdots, Y_m$ are all uncorrelated and $Var(Y_i) = \lambda_i,\ i = 1, \cdots, m$. The components $Y_1, \cdots .Y_m$ of $\mathbf{Y}$ are called *principal components* of $\mathbf{X}$. Hotelling (1933) developed the principal component procedure. In multivariate data analysis, an experimenter is often encountered with a large set of correlated variables. Principal components usually serve as intermediate steps in much larger studies. PCA is frequently used by the practitioner to extract the main relations in data of high dimensionality.

Here we are primarily interested in the estimation of coefficient $\mathbf{e}_j$ of the jth principal component that may be equal to a specified value $\mathbf{e}_j^o$. This information may be regarded as *uncertain prior information (UPI)* regarding parameter vector $\mathbf{e}_j$.

The statistical objective is how to incorporate this information in the estimation process. Generally speaking, consequences of incorporating UPI depend on the *quality or reliability* of information introduced in the estimation process. This uncertain prior information, in the form of the null hypothesis, can be used in two different ways in the estimation procedure. In the first place, it is natural to perform a preliminary test on the validity of the *UPI* in the form of parametric restrictions, and then choose between the restricted and unrestricted estimation procedure depending upon the outcome of the preliminary test. This idea was initially conceived by T. A. Bancroft in 1944. However, this may be partly motivated by the remarks made by Berkson (1942). In the later case, the James-Stein estimation procedure is adopted. For the past three decades, the researchers have paid considerable attention to the James-Stein type estimation, in the small sample as well as in the large sample set-up. Readers may find the sufficient amount of developments in the current literature. James and Stein (1961) presented an explicit form of an estimator, which dominates the classical estimator for the mean vector of a multivariate normal distribution. Their estimator shrinks the usual estimator towards a null vector and such an estimator is generally called a *shrinkage estimator (SE)*. However, there is no reason why the usual sample estimator must shrink towards the null vector. In general, we may shrink the classical estimator towards any arbitrary fixed vector. That is what we intend to do in this communication. We plan to shrink the classical estimator of $\mathbf{e}_j$ towards an arbitrary vector $\mathbf{e}_j^o$.

Thus, there are other estimation strategies available which are theoretically superior and often lead to efficient methods of estimation. To this end, the preliminary test and Stein-type estimators are a family of superior estimation methods that use historical data or uncertain prior information. The basic principle behind these estimators is simple yet very powerful. If there is some prior information (or prior estimates) available from the past, then it can be used to improve the estimate for the current estimation problem.

1.1 Motivating examples

Let us consider the following examples to motivate the problem at hand.

Stocks Data: Consider the monthly rates of rate return for four or more stocks at a given stock exchange. Generally speaking, the observations for successive weeks appear to be independently distributed. However, the rate of return across stock are correlated since as one expects stocks tend to move together in response to general economic conditions. A financial analyst wishes to compute a few sample principal components of monthly rates of return. For example, the first component may be an equally weighted sum, or *index* of the stocks. Perhaps this component may be called a *general stock-market component*, or simply a market component. The second component may represent a contrast between two major stocks. It might be called an *industry component*. In most cases most of the variation in stock returns is due to market activity and uncorrelated industry activity. The analyst suspects that the eigen vectors of this month's return may not differ from the past few months average and hence wishes to calculate the sample principal components of this month's rates for a return that incorporates the prior information. This make sense since it is unusual that users rely on only current sample data for the estimation of returns, volatility, and correlation for the variable of interest in the study.

Application in Genetics: Geneticists are often concerned with the inheritance of characteristics that can be measured several times during an animal's lifetime. For the purpose of illustration consider body weights for 100 female mice that were obtained immediately after birth of their first 4 litters. We consider the situation that correlations may be close enough to provide uncertain prior information regarding first eigenvector. In some cases, the first principal component may account for most of the variation. The average post-birth weights is fairly well explained by the first principal component. Jolicoeur and Mosimann (1960) studied the relationships of size and shape for painted turtles measured carapace length, width and height. For the male turtles the PCA indicates that the first principal component explains approximately 96% of the total variance, which has interesting subject-matter interpretations. Since the first principal component may be viewed as logarithm(volume) of a box with adjusted dimension. In order to use the proposed methodology one needs to have the prior knowledge about the first eigenvector only which may be taken nearly as $\frac{1}{\sqrt{m}}$ for each entry, but may not be equal.

1.2 Statement of the problem

Let $\mathbf{X}_1,\ \mathbf{X}_2, \cdots \mathbf{X}_n$be the random sample from a multivariate normal population with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. Further, we assume that the (unknown) eigen values $(\lambda_1, \cdots, \lambda_m)$ of the covariance matrix $\boldsymbol{\Sigma}$ are distinct and positive. Let $\mathbf{e}_1, \cdots, \mathbf{e}_m$ be the corresponding normalized eigen vectors. We try to motivate this problem from two viewpoints. First, suppose that it were thought a priori likely, though not certain, that $\mathbf{e}_j = \mathbf{e}_j^o$, where $\mathbf{e}_j^o$ is a specified $m \times 1$ vector $\mathbf{e}_j^o = (\mathbf{e}_{j1}^o, \cdots \mathbf{e}_{jm}^o)'$. We are interested in the estimation of coefficient $\mathbf{e}_j$ of the jth principal component under the above information. The restriction stated in the form of the null hypothesis is

$$H_o:\ \mathbf{e}_j = \mathbf{e}_j^o. \tag{1.1}$$

This information can be used to construct preliminary test and shrinkage estimators. Second, one may apply the Bayesian methodology to obtain an empirical Bayes estimator.

Various authors including Olkin and Selliah (1977) considered the problem of estimating $\boldsymbol{\Sigma}$ directly by perturbing the eigenvalues of the sample covariance matrix. Ahmed (1998) and Dey (1988) estimated the eigenvalues using shrinkage methods. Leung (1992) considered the estimation of eigenvalues of the scale matrix of the multivariate F distribution. Joarder and Ahmed (1996) considered the estimation problem for a multivariate t distribution. Daniels and Kass (2001) proposed the shrinkage estimation for covariance matrices. Recently, Judge and Mittelhammer (2004) applied shrinkage technique in semiparametric regression model.

A plan of this paper is as follows. The estimators are formally introduced in Section 2. The properties of the estimators are investigated in Section 3. Some computational analysis is presented in Section 4. Examples based on real data are given Section in 5 and section 6 summarizes the findings.

2 Large-sample estimation strategies

Assume that the data $\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_n$ represent n independent drawings from a m-dimensional normal population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. These data yield the sample mean vector $\bar{\mathbf{x}}$, and the sample covariance matrix $\mathbf{S}$. Let $\hat{\mathbf{e}}_1, \cdots, \hat{\mathbf{e}}_m$ be the normalized eigen vectors of the sample covariance matrix $\mathbf{S}$ corresponding to eigen values $\hat{\lambda}_1, \cdots, \hat{\lambda}_m$. We call $\hat{\mathbf{e}_j}^U$ the *unrestricted estimator (UE)* of $\mathbf{e}_j$. Note that $\hat{\mathbf{e}_j}^U$ of $\mathbf{e}_j$ is based on sample data only and hence does not incorporate the UPI in estimating $\mathbf{e}_j$. However, it may be advantageous to use the available UPI to obtain improved estimates. In the following sub-sections, we introduced some improved estimation methodologies.

2.1 Simple shrinkage estimator

Our objective is to identify a weighted linear combination of $\hat{\mathbf{e}}_j{}^U$ and $\mathbf{e}_j^o$ with smaller expected quadratic risk than the sample estimator $\hat{\mathbf{e}}_j{}^U$. Toward, this end, define a *simple shrinkage (SS)* estimator of $\mathbf{e}_j$ as

$$\hat{\mathbf{e}}_j{}^{SS} = \varpi \mathbf{e}_j^o + (1-\varpi)\hat{\mathbf{e}}_j{}^U, \quad \varpi \in (0,1), \tag{2.1}$$

where ϖ is a constant, called the shrinkage coefficient. One might consider ϖ and $(1-\varpi)$ as two weighted constants of $\mathbf{e}_j^o$ and $\hat{\mathbf{e}}_j{}^U$ respectively. The value of ϖ reflects the degree of confidence in the UPI. The proposed estimator $\hat{\mathbf{e}}_j{}^{SS}$ is a linear and convex combination of $\hat{\mathbf{e}}_j{}^U$ and $\mathbf{e}_j^o$ via a fixed value of ϖ. This is an example of the simple shrinkage estimator: the sample estimator is 'shrunk' towards $\mathbf{e}_j^o$. The result is a biased estimator with a smaller risk near the hypothesized value.

2.2 Shrinkage preliminary test estimation

The *shrinkage preliminary test (SPT)* estimator of $\mathbf{e}_j$ is defined by

$$\hat{\mathbf{e}}_j{}^{SP} = \hat{\mathbf{e}}_j{}^{SS} I(\mathcal{T}_j < c_{(n,\alpha)}) + \hat{\mathbf{e}}_j{}^U I(\mathcal{T}_j \geq c_{(n,\alpha)}), \tag{2.2}$$

where $I(A)$ is an indicator function of a set A. $\mathcal{T}_j$ is a test-statistic for the null hypothesis H_o in (1.1) and will be given in the next section. For $\varpi = 1$, we obtain the simple *preliminary test estimator (PT)* given by

$$\hat{\mathbf{e}}_j{}^P = \mathbf{e}_j^o I(\mathcal{T}_j < c_{(n,\alpha)}) + \hat{\mathbf{e}}_j{}^U I(\mathcal{T}_j \geq c_{(n,\alpha)}). \tag{2.3}$$

Both preliminary test estimators are convex combinations of $\hat{\mathbf{e}}_j{}^U$, $\hat{\mathbf{e}}_j{}^{SS}$ and $\hat{\mathbf{e}}_j{}^U$ and non-sample information ($\mathbf{e}_j{}^o$) respectively, via a test-statistic for testing H_o. These estimators abound in a wide range of statistical applications, as evidenced by the bibliography by Giles and Giles (1993). In recent literature a discussion about preliminary testing can be found in Magnus (1999), Ohanti (1999), Reif and VIcek (2002), Khan and Ahmed (2003).

It is known that the preliminary test estimators are sensitive to departure from H_o and may not be perfect for all $\mathbf{e}_j$ (Cohen, 1968). Further, for the multivariate normal mean vector, Stein (1956) and James and Stein (1961) considered a shrinkage estimator which dominates the sample mean vector. Their basic theory has been extended in various directions. To avoid the imperfection in preliminary test estimation, we propose two variants of *shrinkage estimators* in the following sub-section.

2.3 Doubly shrinkage estimation

We adopt the Stein-rule estimation approach to obtain the estimators that will dominate $\hat{\mathbf{e}}_j{}^U$ over the entire parameter space notwithstanding of how correct the UPI is. First, we propose the *doubly shrinkage (DS)* estimator as follows:

$$\begin{aligned} \hat{\mathbf{e}}_j{}^{DS} &= \hat{\mathbf{e}}_j{}^{SS} + \{1-(m-3)\mathcal{T}_j^{-1}\}(\hat{\mathbf{e}}_j{}^U - \hat{\mathbf{e}}_j{}^{SS}) \\ &= \hat{\mathbf{e}}_j{}^U - \varpi(m-3)\mathcal{T}_j^{-1}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j{}^o), \quad m \geq 4. \end{aligned} \tag{2.4}$$

Note that for $\varpi = 1$, we get the *ordinary shrinkage estimator (OSE)*, based on Stein-rule

$$\hat{\mathbf{e}}_j{}^S = \hat{\mathbf{e}}_j{}^U - (m-3)\mathcal{T}_j^{-1}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j{}^o), \quad m \geq 4. \tag{2.5}$$

The proposed shrinkage estimators use the test-statistic $\mathcal{T}_j$ to combine the sample data with the hypothesis H_o. In this respect their course of action is like the preliminary test estimator. However, the estimator based on the preliminary test uses the statistic $\mathcal{T}_j$ to select between $\hat{\mathbf{e}}_j{}^{SS}$ and the unrestricted estimator whereas the proposed estimator $\hat{\mathbf{e}}_j{}^{DS}$ uses the test statistic $\mathcal{T}_j$ to shrink the estimator $\hat{\mathbf{e}}_j{}^U$ towards the $\hat{\mathbf{e}}_j{}^{SS}$.

Remark 1: Noting that $\hat{\mathbf{e}}_j{}^{DS}$ is not a convex combination of $\hat{\mathbf{e}}_j{}^{SS}$ and $\hat{\mathbf{e}}_j{}^U$ because $[(m-3)\mathcal{T}_j^{-1}]$ may be greater than 1.

Remark 2: Shrinkage estimator has a alarming peculiarity of over shrinking which may make the coordinates of shrinkage estimator have a different sign from the coordinates of the maximum likelihood estimator. One may agree on adjusting the magnitude of $\hat{\mathbf{e}}_j{}^U$, but the change of sign is somewhat a grave matter and it would make a practitioner rather uncomfortable if $\mathcal{T}_j$ near zero is observed. However, it is important to note that this behavior does not adversely affect the risk performance of the shrinkage estimators.

A small adjustment of the shrinkage estimators leads to a convex combination of $\hat{\mathbf{e}}_j{}^U$ and $\hat{\mathbf{e}}_j{}^{SS}$. We trim $\hat{\mathbf{e}}_j{}^{DS}$ with its positive-part and propose the *positive-part doubly shrinkage (PDS)* estimator which may be written as

$$\begin{aligned} \hat{\mathbf{e}}_j{}^{DS+} &= \hat{\mathbf{e}}_j{}^U - \varpi(m-3)\mathcal{T}_j^{-1}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j^o) \\ &-\varpi\{1-(m-3)\mathcal{T}_j^{-1}\}I(\mathcal{T}_j < m-3)(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j^o), \quad m \geq 4. \end{aligned} \tag{2.6}$$

Naturally, this estimator prevents changing the sign of $\hat{\mathbf{e}}_j{}^U$ because if $\mathcal{T}_j$ is observed small $\hat{\mathbf{e}}_j{}^{DS+}$ assumes the value of $\hat{\mathbf{e}}_j{}^{SS}$. We shall later see that the DS estimator is dominated by its truncated version. For $\varpi = 1$, we obtain *ordinary positive-part shrinkage* estimator is defined as

$$\begin{aligned} \hat{\mathbf{e}}_j{}^{S+} &= \hat{\mathbf{e}}_j{}^U - (m-3)\mathcal{T}_j^{-1}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j^o) \\ &-\{1-(m-3)\mathcal{T}_j^{-1}\}I(\mathcal{T}_j < m-3)(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j{}^o), \quad m \geq 4. \end{aligned} \tag{2.7}$$

For review of the shrinkage estimators, we may refer to Kubokowa (1998), Hoffmann (1992), Brandwein and Strawderman (1990), Stigler (1990), Ahmed and Saleh (1999), and Ahmed and Sen (2004) among others.

Remark 3: We have defined several possible estimators for the parameter vector of interest. Although the proposed estimators no longer preserve the unit norms, however, the risk performance of these estimators is superior to that of $\hat{\mathbf{e}}_j{}^U$.

In the present investigation we shall study the properties of the proposed estimators under an asymptotic set up in the light of the usual weighted quadratic loss function. Considering a sequence of local alternatives $\{K_{(n)}\}$

$$K_{(n)} : \mathbf{e}_j = \mathbf{e}_{j_n}, \quad \text{where} \quad \mathbf{e}_{j_n} = \mathbf{e}_j^o + \frac{\boldsymbol{\delta}}{\sqrt{n}}. \tag{2.8}$$

The *asymptotic distribution function (ADF)* of $\{\sqrt{n}(\mathbf{e}_j^* - \mathbf{e}_j)\}$ is given by

$$G(\mathbf{y}) = \lim_{n\to\infty} P\{\sqrt{n}(\mathbf{e}_j^* - \mathbf{e}_j) \leq \mathbf{y}\}, \tag{2.9}$$

where $\mathbf{e}_j{}^*$ is any estimator of $\mathbf{e}_j$ for which the limit in (2.9) exists. Further, $\mathbf{M} = \int\int\cdots\int \mathbf{y}\mathbf{y}' dG(\mathbf{y})$. Then, the *asymptotic distributional risk (ADR)* is defined by $ADR(\mathbf{e}_j^*; \mathbf{e}_j) = \text{trace}(\mathbf{VM})$, where $\mathbf{V}$ is a given positive semi-definite matrix.

3 Main results

Anderson (1963) established the following theorem which gives asymptotic distribution for the sample eigenvetors $\mathbf{e}_j$.

Theorem (Anderson, 1963): Let $\mathbf{X}$ be an $m \times 1$ random vector from a multivariate population with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. Suppose that the eigen values of $\boldsymbol{\Sigma}$ are $\lambda_1, \lambda_2, \cdots, \lambda_m (> 0)$ and let $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$ be corresponding normalized eigen vectors. Label $\hat{\mathbf{e}}_1, \cdots \hat{\mathbf{e}}_m$ be the normalized eigen values of the sample covariance matrix $\mathbf{S}$ corresponding to eigen roots $\hat{\lambda}_1, \cdots, \hat{\lambda}_m (> 0)$. If λ_j is a distinct root, then as $n \to \infty$, $n^{\frac{1}{2}}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j)$ has a limiting multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Gamma}$, $N_m(\mathbf{0}, \boldsymbol{\Gamma})$, where

$$\boldsymbol{\Gamma} = \lambda_j \sum_{i=1}^{m} \sum_{j\neq i}^{m} \left[\frac{\lambda_i}{(\lambda_j - \lambda_i)^2}\right] \mathbf{e}_i \mathbf{e}_i{}'.$$

It is important to note that covariance matrix $\boldsymbol{\Gamma}$ is singular. Further, we may represent $\boldsymbol{\Gamma}$ as follows

$$\boldsymbol{\Gamma} = \mathbf{E}_{(j)} \mathbf{D}_{(j)}^2 \mathbf{E}_{(j)}',$$

where

$$\mathbf{E}_{(j)} = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_{j-i}, \mathbf{e}_{j+1}, \cdots \mathbf{e}_m)$$

$$\mathbf{D}_{(j)}^2 = diag\left(\frac{\lambda_j\lambda_1}{(\lambda_j - \lambda_1)^2}, \cdots, \frac{\lambda_j\lambda_{j-1}}{(\lambda_j - \lambda_{j-1})^2}, \frac{\lambda_j\lambda_{j+1}}{(\lambda_j - \lambda_{j+1})^2}, \cdots, \frac{\lambda_j\lambda_m}{(\lambda_j - \lambda_m)^2}.\right)$$

Let us define the following transformation

$$\mathbf{z} = \mathbf{D}_{(j)}^{-1} \mathbf{E}_{(j)} [n^{\frac{1}{2}}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j)],$$

then the limiting distribution of $\mathbf{z}$ is $\aleph_{m-1}(\mathbf{0}, \mathbf{I})$, where $\mathbf{I}$ is an identity matrix. Thus,

$$\mathbf{z}'\mathbf{z} = [n^{\frac{1}{2}}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j)]' \mathbf{Q} [n^{\frac{1}{2}}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j)],$$

where $\mathbf{Q} = \mathbf{E}_{(j)} \mathbf{D}_{(j)}^{-2} \mathbf{E}_{(j)}'$. Hence the limiting distribution of $\mathbf{z}'\mathbf{z}$ is chi-square distribution with $m-1$ degrees of freedom. The matrix $\mathbf{Q}$ may be displayed in the following relation

$$\mathbf{Q} = \lambda_j \boldsymbol{\Sigma}^{-1} - 2\mathbf{I} + \frac{1}{\lambda_j} \boldsymbol{\Sigma}.$$

Hence, the limiting distribution of the following quadratic form

$$[n^{\frac{1}{2}}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j)]' \left(\hat{\lambda}_j \mathbf{S}^{-1} - 2\mathbf{I} + \frac{1}{\hat{\lambda}_j}\mathbf{S} \right) [n^{\frac{1}{2}}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j)]$$

is a central chi-square distribution (Anderson, 1963). Further, it can be shown that

$$\mathcal{T}_j = n \left(\hat{\lambda}_j \mathbf{e}_j^{o\prime} \mathbf{S}^{-1} \mathbf{e}_j^o + \frac{1}{\hat{\lambda}_j} \mathbf{e}_j^{o\prime} \mathbf{S} \mathbf{e}_j^o - 2 \right)$$

is distributed asymptotically as a central chi-square distribution when the null hypothesis in (1.1) is true with $(m-1)$ degrees of freedom. For given $\alpha(0 < \alpha < 1)$, the critical value of $\mathcal{T}_j$ may be approximated by $\chi^2_{m-1,\alpha}$, the upper $100\alpha\%$ point of the chi-square distribution with $(m-1)$ degrees of freedom under H_o.

Note that for fixed alternatives it is straightforward to show that $\mathcal{T}_j$ is a consistent test. For this obvious reason all the estimators will be ADR equivalent to $\hat{\mathbf{e}}_j{}^U$ while $\hat{\mathbf{e}}_j{}^{SS}$ will have a unbounded ADR. To avoid this technical difficulty we confine to a sequence of local alternatives $\{K_{(n)}\}$ defined in relation (2.8) and have the following useful lemmas.

Lemma 3.1: Under local alternatives and usual regularity conditions as n increases $\sqrt{n}(\hat{\mathbf{e}}_j{}^U - \mathbf{e}_j{}^o) \xrightarrow{L} \aleph_m(\boldsymbol{\delta}, \boldsymbol{\Gamma})$. Hence, the test statistic $\mathcal{T}_j$ is distributed asymptotically as a non-central chi-square distribution with $(m-1)$ degrees of freedom and non-centrality parameter $\Delta = \boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}$.

Let us define *asymptotic distributional bias vector (adbv)* of an estimator $\mathbf{e}_j^*$ of $\mathbf{e}_j$ as

$$\mathbf{adbv}(\mathbf{e}_j^*) = \lim_{n\to\infty} E\{\sqrt{n}(\mathbf{e}_j^* - \mathbf{e}_j\}.$$

Further, we transform various bias functions in scalar (quadratic) form by defining

$$qb(\mathbf{e}_j^*) = [\mathbf{abdv}(\mathbf{e}_j^*)]'\mathbf{Q}[\mathbf{abdv}(\mathbf{e}_j^*)],$$

Thus, $qb(\mathbf{e}_j^*)$ is called quadratic bias of an estimator $\mathbf{e}_j^*$ of parameter vector $\mathbf{e}_j$.

It is seen that only $\hat{\mathbf{e}}_j{}^U$ is an asymptotically unbiased estimator of $\mathbf{e}_j$. On the other hand, $adbv(\hat{\mathbf{e}}_j{}^{SS})$ is $-\varpi\boldsymbol{\delta}$, hence the quadratic bias of $\hat{\mathbf{e}}_j{}^{SS}$ is $\varpi^2\Delta$. Thus, quadratic bias of $\hat{\mathbf{e}}_j{}^{SS}$ increases with Δ without an upper bound. This estimator will achieve the property of unbiasedness if and only if $\Delta = 0$ which in return it requires $\boldsymbol{\delta}$ to be a null vector.

Theorem 3.1: The *adbv* of $\hat{\mathbf{e}}_j{}^{SP}$ is

$$\mathbf{adbv}(\hat{\mathbf{e}}_j{}^{SP}) = -\varpi\boldsymbol{\delta} H_{m+1}(\chi^2_{m-1,\alpha}; \Delta),$$

where $H_m(x\ ;\Delta)$ stands for the non-central chi-square distribution function with non-centrality parameter Δ and m degrees of freedom. This expression is established by using the results of Lemma 3.1.

The quadratic bias is $qb(\hat{\mathbf{e}_j}^{SP}) = \varpi^2\Delta[H_{m+1}(\chi^2_{m-1,\alpha};\Delta)]^2$. Obviously, the qb of $\hat{\mathbf{e}_j}^{SP}$ is a function of Δ, ϖ and α. As a function of Δ (for fixed ϖ and α), it starts from 0 increases to a point, then is decreased gradually to zero. As a function of α for fixed Δ and ϖ, it is a decreasing function of α with a maximum value $\varpi^2\Delta$ at $\alpha = 0$ and 0 at $\alpha = 1$. The statistical properties of $\hat{\mathbf{e}_j}^{P}$ may be obtained for $\varpi = 1$ in the $qb(\hat{\mathbf{e}_j}^{SP})$.

Theorem 3.2: Using Lemma 3.1 the bias expression for $\hat{\mathbf{e}_j}^{DS}$ is

$$\mathbf{adbv}(\hat{\mathbf{e}_j}^{DS}) = -\varpi(m-3)\boldsymbol{\delta}E(\chi^{-2}_{m+1}(\Delta)),$$

where $E\left(\chi_m^{-2k}(\Delta)\right) = \int_0^\infty x^{-2k}dH_m(x\,;\Delta)$. Hence $qb(\hat{\mathbf{e}_j}^{DS}) = \varpi^2(m-3)^2\Delta[E(\chi^{-2}_{m+1}(\Delta))]^2$. $E(\chi^{-2}_{m+1}(\Delta))$ is a decreasing log-convex function of Δ. Hence, the qb of $\hat{\mathbf{e}_j}^{DS}$ starts at $\Delta = 0$ then decreases to a point and then increases towards the origin.

Figure 1 validates the behavior of the quadratic bias of the proposed estimator.

Theorem 3.3: By virtue of Lemma 3.1 the expressions for **adbv** and qb $\hat{\mathbf{e}_j}^{DS+}$ are given in the following relations, respectively

$$\mathbf{adbv}(\hat{\mathbf{e}_j}^{DS+}) = -\varpi\boldsymbol{\delta}\left[H_{m+1}(m-3;\Delta) + E\{\chi^{-2}_{m+1}(\Delta)I(\chi^2_{m+1}(\Delta)) > (m-3))\}\right]$$

$$qb(\hat{\mathbf{e}_j}^{DS+}) = \varpi^2\Delta\left[H_{m+1}(m-3;\Delta) + E\{\chi^{-2}_{m+1}(\Delta)I(\chi^2_{m+1}(\Delta)) > (m-3))\}\right]^2.$$

The graph of $qb(\hat{\mathbf{e}_j}^{DS+})$ follows the same pattern as that of $\hat{\mathbf{e}_j}^{DS}$. However, $\hat{\mathbf{e}_j}^{DS+}$ is less biased than that of $\hat{\mathbf{e}_j}^{DS}$ for all $\varpi \in [0,1)$.

[Figure 1 about here]

Now, we provide asymptotic distributional risk analysis for the proposed estimators under local alternatives.

It is seen that ADR of $\hat{\mathbf{e}_j}^{U}$ has a constant value trace$(\mathbf{V\Gamma})$. The ADR of $\hat{\mathbf{e}_j}^{SS}$ is $(1-\varpi)^2$trace$(\mathbf{V\Gamma}) + \varpi^2\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta}$. It is an unbounded function of Δ and satisfies the inequality

$$\begin{aligned}(1-\varpi)^2tr(\mathbf{V\Gamma}) \quad &+\varpi^2\Delta Ch_{\min}(\mathbf{V\Gamma}) \le \text{ADR}(\hat{\mathbf{e}_j}^{SS}) \\ \le (1-\varpi)^2tr(\mathbf{V\Gamma}) \quad &+\varpi^2\Delta Ch_{\max}(\mathbf{V\Gamma}),\end{aligned} \tag{3.1}$$

where $Ch_{min}(\mathbf{A})$ and $Ch_{\max}(\mathbf{A})$ are the smallest and the largest eigen values of the matrix $\mathbf{A}$. Thus, the two bounds of ADR $(\hat{\mathbf{e}_j}^{SS})$ intersects with the ADR$(\hat{\mathbf{e}_j}^{U})$ at

$$\Delta_l = \frac{2-\varpi}{\varpi}\frac{tr(\mathbf{V\Gamma})}{Ch_{\max}(\mathbf{V\Gamma})} \quad \text{and} \quad \Delta_m = \frac{2-\varpi}{\varpi}\frac{tr(\mathbf{V\Gamma})}{Ch_{\min}(\mathbf{V\Gamma})}$$

respectively. Also, if

$$0 \le \Delta \le \Delta_l \tag{3.2}$$

then $\hat{\mathbf{e}_j}^{SS}$ has a smaller ADR than that of $\hat{\mathbf{e}_j}^{U}$. On the other hand, if $\Delta_m \le \Delta$ then $\hat{\mathbf{e}_j}^{U}$ has a smaller ADR than $\hat{\mathbf{e}_j}^{SS}$.

Theorem 3.4: Under local alternatives and as $n \to \infty$, the ADR of $\hat{\mathbf{e}_j}^{SP}$ is

$$\begin{aligned} ADR(\hat{\mathbf{e}_j}^{SP}, \mathbf{e}_j) &= \text{trace}(\mathbf{V\Gamma}) - \varpi(2-\varpi)\text{trace}(\mathbf{V\Gamma})H_{m+1}(\chi^2_{m-1,\alpha};\Delta) \\ &+\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta}\{2\varpi H_{m+1}(\chi^2_{m-1,\alpha};\Delta) - \varpi(2-\varpi)H_{m+3}(\chi^2_{m-1,\alpha};\Delta)\}. \end{aligned} \tag{3.3}$$

Proof: We use the same argument as in Ahmed (2001) to arrive at the above relation (and thereby avoiding the proof).

From relation 3.3 it is clear that if

$$0 \le \Delta \le \frac{(2-\varpi)tr(\mathbf{V\Gamma})H_{m+1}(\chi^2_{m-1}(\alpha);\Delta)}{Ch_{\max}(\mathbf{V\Gamma})[2H_{m+1}(\chi^2_{m-1}(\alpha);\Delta) - (2-\varpi)H_{m+3}(\chi^2_{\varpi-1}(\alpha);\Delta)]}, \tag{3.4}$$

then the ADR of $\hat{\mathbf{e}_j}^{SP}$ is smaller than that of $\hat{\mathbf{e}_j}^{U}$. As α (the level of significance) approaches to 1, ADR($\hat{\mathbf{e}_j}^{SP}$) tends to ADR($\hat{\mathbf{e}_j}^{U}$). This is also true when $\Delta \to \infty$. Undoubtedly, at $\Delta = 0$, the ADR($\hat{\mathbf{e}_j}^{SP}$) assumes the smallest possible value, i.e., $tr(\mathbf{V\Gamma})\{1 - \varpi(2-\varpi)H_{m+1}(\chi^2_{m-1}(\alpha);0\}$, which keeps on increasing crossing the line $tr(\mathbf{V\Gamma})$, reaches to maximum then decreases monotonically to the ADR($\hat{\mathbf{e}_j}^{U}$).

For $\varpi = 1$, we obtain the comparison of $\hat{\mathbf{e}_j}^{U}$ and $\hat{\mathbf{e}_j}^{P}$. Thus, $\hat{\mathbf{e}_j}^{P}$ performs better than $\hat{\mathbf{e}_j}^{U}$ whenever Δ is in the interval

$$0 \le \Delta \le \frac{tr(\mathbf{V\Gamma})H_{m+1}(\chi^2_{m-1}(\alpha);\Delta)}{Ch_{\max}(\mathbf{V\Gamma})\{2H_{m+1}(\chi_{m+1}(\chi^2_{m-1}(\alpha);\Delta) - H_{m+3}(\chi^2_{m-1}(\alpha);\Delta)\}}. \tag{3.5}$$

Interestingly, while comparing the right hand side of (3.4) to that of (3.5) we find that $\hat{\mathbf{e}_j}^{SP}$ provides a wider range than $\hat{\mathbf{e}_j}^{P}$ in which it has a smaller ADR than $\hat{\mathbf{e}_j}^{U}$. This indicates that the performance of $\hat{\mathbf{e}_j}^{SP}$ is relatively superior to $\hat{\mathbf{e}_j}^{P}$.

If $\Delta = 0$, then the ADR of $\hat{\mathbf{e}_j}^{SS}$ takes the value $(1-\varpi)^2 tr(\mathbf{V\Gamma})$ whereas ADR of $\hat{\mathbf{e}_j}^{SP}$ is $tr(\mathbf{V\Gamma})\{1 - \varpi(2-\varpi)H_{m+1}(\chi^2_{m-1}(\alpha);0\}$. The ADR difference at $\Delta = 0$ is

$$\text{ADR}(\hat{\mathbf{e}_j}^{SS}) - \text{ADR}(\hat{\mathbf{e}_j}^{SP}) = -\varpi(2-\varpi)\{1 - H_{m+1}(\chi^2_{m-1}(\alpha),0)\}.$$

It is concluded that $\hat{\mathbf{e}_j}^{SS}$ performs better than $\hat{\mathbf{e}_j}^{SP}$ under H_o. However, for $\Delta \neq 0$, the ADR difference shows that $\hat{\mathbf{e}_j}^{SS}$ performs better than $\hat{\mathbf{e}_j}^{SP}$ whenever

$$0 \le \Delta \le \frac{(2-\varpi)tr(\mathbf{V\Gamma})\{1 - H_{m+1}(\chi^2_{m-1}(\alpha);\Delta)\}}{Ch_{\max}(\mathbf{V\Gamma})\{\varpi - 2H_{m+1}(\chi^2_{m-1}(\alpha);\Delta) + (2-\varpi)H_{m+3}(\chi^2_{m-1}(\alpha);\Delta)\}}.$$

It can be safely said that neither $\hat{\mathbf{e}_j}^{SS}$ nor $\hat{\mathbf{e}_j}^{SP}$ is asymptotically admissible with respect to each other.

Theorem 3.5: For large n , the ADR function for $\hat{\mathbf{e}_j}^{DS+}$ under the local alternatives is

$$\begin{aligned} ADR(\hat{\mathbf{e}_j}^{DS+}, \mathbf{e}_j) &= ADR(\hat{\mathbf{e}_j}^{DS}, \mathbf{e}_j) - \varpi^2\text{trace}(\mathbf{V\Gamma}) \\ &E[\{1-(m-3)\chi^{-2}_{m+1}(\Delta)\}^2 I(\chi^2_{m+1}(\Delta) \le (m-3)] \\ &+\varpi\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta}\big[E[2\{1-(m-3)\chi^{-2}_{m+1}(\Delta)\}I(\chi^2_{m+1}(\Delta) \le (m-3)] \\ &-\varpi E[\{1-(m-3)\chi^{-2}_{m+3}(\Delta)\}^2 I(\chi^2_{m+3}(\Delta) \le (m-3)]\big], \end{aligned} \tag{3.6}$$

where

$$\begin{array}{ll} ADR(\hat{\mathbf{e}}_j^{DS}, \mathbf{e}_j) & = \mathrm{trace}(\mathbf{V\Gamma}) + \boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta}\varpi(m-3)[4+\varpi(m-3)]E(\chi_{m+3}^{-4}(\Delta)) \\ & -\varpi(m-3)\mathrm{trace}(\mathbf{V\Gamma})\{2E(\chi_{m+1}^{-2}(\Delta)) - c(m-3)E(\chi_{m+1}^{-4}(\Delta))\}. \end{array} \quad (3.7)$$

Proof: We derive the above expressions by using the argument as in Ahmed (2001) and after some tedious algebra and repeated application of Stein's identities.

We determine the value of ϖ at which minimum is attained for the ADR of $\hat{\mathbf{e}}_j^{DS}$. Using the Stein's identity

$$E[\chi_{p-2}^{-2}(\Delta)] - (p+2)E[\chi_{p+2}^{-4}(\Delta)] = \Delta E[\chi_{p+4}^{-4}(\Delta)],$$

taking $p = m-1$ in the above identity then

$$E[\chi_{m+1}^{-2}(\Delta)] - (m-3)E[\chi_{m+1}^{-4}(\Delta)] = \Delta E[\chi_{m+3}^{-4}(\Delta)].$$

Let $g_1 = E[\chi_{m+3}^{-4}(\Delta)],\ g_2 = E[\chi_{m+1}^{-2}(\Delta)]$ and $g_3 = E[\chi_{m+1}^{-4}(\Delta)]$. Then $\Delta g_1 = g_2 - (m-3)g_3$. Hence,

$$\begin{array}{ll} ADR(\hat{\mathbf{e}}_j^{DS}, \mathbf{e}_j) & = (m-1) + \Delta(m-3)g_1(4+\varpi)(m-3))\varpi \\ & -\varpi(m-3)(m-1)[2g_2 - \varpi(m-3)g_3] \\ & = \varpi^2[(m-3)^2\Delta g_1 + (m-3)^2(m-1)g_3] + \varpi[4\Delta g_1(m-3)- \\ & 2g_2(m-1)(m-3)] + (m-1). \end{array}$$

Noting that this is a quadratic function of ϖ, it has form $a\varpi^2 + b\varpi + c$ and $a > 0$. Then this function has minimum at the point $-\frac{b}{2a}$. In this case

$$\begin{array}{ll} \varpi_o & = -\frac{b}{2a} = \frac{2g_2(m-1)(m-3)-4\Delta g_1(m-3)}{2(m-3)^2[\Delta g_1+(m-1)g_3]} \\ & = \frac{g_2(m-1)-2\Delta g_1}{(m-3)[\Delta g_1+(m-1)g_3]}. \end{array}$$

Now by applying the Stein's identity for Δg_1 both in numerator and denominator, we get

$$\varpi_o = \frac{g_2(m-1) - 2g_2 + 2(m-3)g_3}{(m-3)[g_2 - (m-3)g_3 + (m-1)g_3]} = 1.$$

Thus, the minimum of ADR function of $\hat{\mathbf{e}}_j^{DS}$ is attained at $\varpi_o = 1$. Further, it can be seen that ADR of $\hat{\mathbf{e}}_j^{DS}$ decreases as ϖ increases.

In order to investigate the comparative properties of the proposed estimators, let us consider the class of positive semi-definite matrices $\mathcal{S}$ defined by

$$\mathcal{S} = \left\{\mathbf{V} : \frac{tr(\mathbf{V\Gamma})}{Ch_{\max}(\mathbf{V\Gamma})} \geq \frac{4+\varpi(m-3)}{2}\right\}. \quad (3.8)$$

We find that

$$\mathrm{ADR}(\hat{\mathbf{e}}_j^{DS+}, \mathbf{e}_j) \leq \mathrm{ADR}(\hat{\mathbf{e}}_j^{DS}, \mathbf{e}_j) \leq \mathrm{ADR}(\hat{\mathbf{e}}_j^{U}, \mathbf{e}_j),$$

for all Δ and $\mathbf{V} \in \mathcal{S}$. Thus $\hat{\mathbf{e}}_j^{DS+}$ is asymptotically superior to the other two estimators for $m \geq 4$.

Under the null hypothesis the ADR of $\hat{\mathbf{e}}_j^{SP}$ and of $\hat{\mathbf{e}}_j^{DS+}$ are

$$\mathrm{ADR}(\hat{\mathbf{e}}_j^{SP}, \mathbf{e}_j) = tr(\mathbf{V\Gamma})\{1 - \varpi(2-\varpi)H_{m+1}(\chi^2_{m-1}(\alpha); 0)\},$$

$$\mathrm{ADR}(\hat{\mathbf{e}}_j^{DS+}, \mathbf{e}_j) = tr(\mathbf{V\Gamma})\left\{1 - \varpi(2-\varpi)\frac{m-3}{m-1}\right\}.$$

The ADR difference is

$$\mathrm{ADR}(\hat{\mathbf{e}}_j^{DS+}, \mathbf{e}_j) - \mathrm{ADR}(\hat{\mathbf{e}}_j^{SP}, \mathbf{e}_j) = tr(\mathbf{V\Gamma})\varpi(2-\varpi)\left\{H_{m+1}(\chi^2_{m-1}(\alpha); 0) - \frac{m-3}{m-1}\right\},$$

which is positive if α belongs to the set

$$A_\alpha = \left\{\alpha : H_{m+1}(\chi^2_{m-1}(\alpha), 0) > \frac{m-3}{m-1}\right\}.$$

If $\alpha \in A_\alpha$, then $\mathrm{ADR}(\hat{\mathbf{e}}_j^{DS+}) > \mathrm{ADR}(\hat{\mathbf{e}}_j^{SP}, \mathbf{e}_j)$ for all $\varpi \in [0, 1)$. In this case clearly $\hat{\mathbf{e}}_j^{SP}$ performs better than $\hat{\mathbf{e}}_j^{DS+}$, otherwise $\hat{\mathbf{e}}_j^{DS+}$ is superior. If we are willing to ignore those values of α that fall in $A(\alpha)$ then under H_o

$$\mathrm{ADR}(\hat{\mathbf{e}}_j^{DS+}, \mathbf{e}_j) \leq \mathrm{ADR}(\hat{\mathbf{e}}_j^{DS}, \mathbf{e}_j) \leq \mathrm{ADR}(\hat{\mathbf{e}}_j^{SP}, \mathbf{e}_j),$$

with strict inequality holds for some Δ. However, it is evident that as Δ departs from 0, i.e., H_o is violated then the dominance picture changes. Nevertheless, as $\Delta \to \infty$, these three estimators are asymptotically equivalent. Otherwise, the curves of ADR of $\hat{\mathbf{e}}_j^{SP}, \hat{\mathbf{e}}_j^{DS}$ and $\hat{\mathbf{e}}_j^{DS+}$ intersect for the set A_α, and none of the estimators are asymptotically admissible relative to each other. However, for those values of $\alpha \notin A_\alpha$, the picture remains as in the above relation. We conclude this section with the following remark.

Remark 4: If $m \leq 3$, one would prefer the use of $\hat{\mathbf{e}}_j^{SP}$ against $\hat{\mathbf{e}}_j^{SS}$ as the size of Δ is usually unknown. However, if $m \geq 4$, then $\hat{\mathbf{e}}_j^{DS+}$ is the most desirable candidate as an estimator of $\mathbf{e}_j$, unless the size of Δ is known to be very small, in this case $\hat{\mathbf{e}}_j^{SS}$ will have the smallest ADR.

4 Some computed risk analysis

We realize that $\mathbf{\Gamma}$ is a singular matrix for computational sake we may choose $\mathbf{V} = \mathbf{Q}$. In this case, $\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta} = \Delta$ and

$$\begin{aligned} tr(\mathbf{V\Gamma}) = \;& trace(\mathbf{Q\Gamma}) = trace(\mathbf{E}_{(j)}\mathbf{E}'_{(j)}) = \\ & trace(\mathbf{I} - \mathbf{e}_j\mathbf{e}'_j) = trace(\mathbf{I}) - trace(\mathbf{e}_j\mathbf{e}'_j) = \\ & m - trace(\mathbf{e}'_j\mathbf{e}_j) = m - 1. \end{aligned}$$

Thus, in sequel we consider $\mathbf{V} = \mathbf{Q}$ to facilitate numerical computation of the various ADR functions and with this substitution, we have the following Lemma.

Lemma 4.1:
$$ADR(\hat{\mathbf{e}}_j^{\,U}, \mathbf{e}_j) = m - 1, \tag{4.1}$$
$$ADR(\hat{\mathbf{e}}_j^{\,SS}, \mathbf{e}_j) = (1-\varpi)^2(m-1) + \varpi^2\Delta \tag{4.2}$$
$$\begin{aligned} ADR(\hat{\mathbf{e}}_j^{\,SP}, \mathbf{e}_j) \;&= (m-1)[1 - \varpi(2-\varpi)H_{m+1}(\chi^2_{m-1,\alpha};\Delta)] \\ &+\Delta\{2\varpi H_{m+1}(\chi^2_{m-1,\alpha};\Delta) - \varpi(2-\varpi)H_{m+3}(\chi^2_{m-1,\alpha};\Delta)\}, \end{aligned}$$
$$\begin{aligned} ADR(\hat{\mathbf{e}}_j^{\,DS}, \mathbf{e}_j) \;&= m - 1 + \Delta\varpi(m-3)[4 + \varpi(m-3)]E(\chi^{-4}_{m+3}(\Delta)) \\ &-\varpi(m-3)(m-1)\{2E(\chi^{-2}_{m+1}(\Delta)) - \varpi(m-3)E(\chi^{-4}_{m+1}(\Delta))\}, \end{aligned} \tag{4.3}$$
$$\begin{aligned} ADR(\hat{\mathbf{e}}_j^{\,DS+}, \mathbf{e}_j) \;&= ADR(\hat{\mathbf{e}}_j^{\,DS}, \mathbf{e}_j) - \varpi^2(m-1) \\ &E[\{1-(m-3)\chi^{-2}_{m+1}(\Delta)\}^2 I(\chi^2_{m+1}(\Delta) \le (m-3)] \\ &+\varpi\Delta\big[E[2\{1-(m-3)\chi^{-2}_{m+1}(\Delta)\}I(\chi^2_{m+1}(\Delta) \le (m-3)] \\ &-\varpi E[\{1-(m-3)\chi^{-2}_{m+3}(\Delta)\}^2 I(\chi^2_{m+3}(\Delta) \le (m-3)]\big]. \end{aligned} \tag{4.4}$$

We have numerically calculated the values of ADR of the estimators versus Δ at selected values of m, α and ϖ. The results are presented graphically in Figures 2-3. These graphs validate our analytical findings as discussed in the previous section. Not surprisingly, proposed Stein-rule estimators outshine their competitors in the entire parameter space induced by non-centrality parameter Δ. However, it is important to note that for small values of ϖ the performance of $\hat{\mathbf{e}}_j^{\,SS}$ is worth considering. In these cases the $\hat{\mathbf{e}}_j^{\,SS}$ dominates the rest of the estimators in a very large parameter space. All the estimators have maximum risk gain over the $\hat{\mathbf{e}}_j^{\,U}$ at $\Delta = 0$ and the value of the improvement is a decreasing function of Δ. However, at $\Delta = 0$, the improvement of $\hat{\mathbf{e}}_j^{\,SS}$ over $\hat{\mathbf{e}}_j^{\,U}$ is the largest as compared to other proposed estimators, which decreases as Δ increases without a bound. The value of the improvement of $\hat{\mathbf{e}}_j^{\,SP}$ over $\hat{\mathbf{e}}_j^{\,U}$ follows a similar pattern, however, it is bounded in Δ. On the other hand, both $\hat{\mathbf{e}}_j^{\,DS}$ and $\hat{\mathbf{e}}_j^{\,DS+}$ improve upon $\hat{\mathbf{e}}_j^{\,U}$ in the entire parameter space.

[Figures 2-3 about here]

5 Examples

In this section we present three examples for practical purposes.

Example 1: We return to the second motivating example described in the Section 1, and consider the analysis of the data provided in Jolicoeur and Mosimann (1960). In a study of size and shape relationships for painted turtles, they measured carapace length, width, and height.

The natural logarithms of the dimensions of 24 male turtles have sample mean vector $\bar{\mathbf{x}} = [4.725, 4.478, 3.703]'$, and covariance matrix
$$\mathbf{S} = 10^{-3}\begin{pmatrix} 11.072 & 8.019 & 8.160 \\ 8.019 & 6.417 & 6.005 \\ 8.160 & 6.005 & 6.773 \end{pmatrix}.$$

A principal component analysis yields the first estimated principal component

$$\hat{\mathbf{e}}_1^U = [0.683, 0.510, 0.523]',$$

which explains 96% of the total variance. There is some evidence that the corresponding population correlation matrix may be of equal-correlation and positively correlated. This information is very well used to specify the value of first the eigenvector. Thus, the first eigenvector may be specified as $\frac{1}{\sqrt{3}}$ plus a perturbation to ensure that the eigenvalues of covariance matrix are distinct. With this remark we may specify $\mathbf{e}_1$ as follows

$$\mathbf{e}_1^o = [0.70, 0.50, 0.51]'.$$

The usual sample estimate is

$$\hat{\mathbf{e}}_1^U = [0.683, 0.510, 0.523]',$$

so the data does seem to be consonant with the prior belief. Further, we find $\mathcal{T}_1 = 0.9722$ for testing the prior belief null hypothesis. In this example, we have $m = 3$, hence with 2 degrees of freedom, the critical value for a 5% significance test is $\chi^2_{(2,.05)} = 5.99$, so the prior belief hypothesis is tenable. Using $\varpi = 0.5$, we have applied some of the estimators defined in this communication to this set of data with the following results.

$$\hat{\mathbf{e}}_1^{SS} = \hat{\mathbf{e}}_1^{SP} = [0.6900, 0.5050, 0.5165]'.$$

We recommend applying pretest estimation in this situation. Noting that in this example we cannot use the Stein-rule estimators due to a dimensional restriction, i.e. $m \geq 4$. The application of Stein-rule estimators is given in the following example.

Example 2: Returning to the first motivating example, we consider monthly rates of return for four stocks, Mobil, Texaco, IBM and DEC (Digital Equipment Company) for the period January 1978 through December 1986. That monthly data set is on pages 109-111, *Stock Market Analysis Using the SAS System: Portfolio Selection and Evaluation* published by the SAS institute, 1994. The observations in 108 successive months appear to be independently distributed, but the rates of return across stocks are correlated. This is expected since, stocks tend to perform together in response to general economic situations.

Let $x_1, \cdots, x_4$ denote observed monthly rates of return for Mobil, Texaco, IBM and DEC, respectively. Then a principal component analysis using SAS yields the following summary.

$$\bar{\mathbf{x}} = [0.0171, 0.0121, 0.0103, 0.0182]'$$

and

$$\mathbf{S} = \begin{pmatrix} 0.0060 & 0.0041 & 0.0009 & 0.0024 \\ 0.0041 & 0.0057 & 0.0007 & 0.0005 \\ 0.0009 & 0.0007 & 0.0030 & 0.0018 \\ 0.0024 & 0.0005 & 0.0018 & 0.0082 \end{pmatrix}.$$

The eigenvalues and corresponding normalized eigenvectors of $\mathbf{S}$ are given below.

$$\begin{array}{llll} \hat{\lambda}_1 = & 0.01241650, & \hat{\mathbf{e}}_1^U = [0.582086, 0.511612, 0.214029, 0.594660]' \\ \hat{\lambda}_2 = & 0.00641397, & \hat{\mathbf{e}}_2^U = [-.386420, -.506979, 0.184045, 0.748185]' \\ \hat{\lambda}_3 = & 0.00251218, & \hat{\mathbf{e}}_3^U = [-.054494, -.018083, 0.959331, -.276382]' \\ \hat{\lambda}_4 = & 0.00170143, & \hat{\mathbf{e}}_4^U = [-.713362, 0.693468, 0.001664, 0.101058]' \end{array}$$

Using the above information, we obtain the first two principal components

$$\begin{array}{ll} \hat{y}_1 = & (\hat{\mathbf{e}}_1^U)'\mathbf{x} = 0.582086x_1 + 0.511612x_2 + 0.214029x_3 + 0.594660x_4 \\ \hat{y}_2 = & (\hat{\mathbf{e}}_2^U)'\mathbf{x} = -0.386420x_1 - 0.506979x_2 + 0.184045x_3 + 0.748185x_4. \end{array}$$

These components, which account for 81.72% of the total sample variance. The first component can be viewed as weighted sum of returns, weights are roughly equal except for the IBM stock. The second component represents a contrast between oil stocks and Tech stocks. Further, from the past data, the initial estimates of $\mathbf{e}_1$ and $\mathbf{e}_2$ available as

$$\mathbf{e}_1^o = [0.60, 0.50, 0.17, 0.60]', \quad \mathbf{e}_2^o = [-0.40, -0.50, 0.30, 0.70]'$$

We calculate $\mathcal{T}_1 = 0.8113$ and $\mathcal{T}_2 = 2.0058$, respectively, for testing the prior belief null hypotheses and with 3 degrees of freedom, the asymptotic critical value for a 5% significance test is $\chi^2_{(3,.05)} = 7.81$. Hence, the prior belief hypotheses are tenable in both cases. Hence, with $\varpi = 0.5$, first principal component is estimated by $\hat{\mathbf{e}}_1^{SS} = \hat{\mathbf{e}}_1^{SP} = [0.5910, 0.5058, 0.1936, 0.5973]'$, and $\hat{\mathbf{e}}_1^{DS+} = \hat{\mathbf{e}}_1^{DS} = [0.5931, 0.5045, 0.1889, 0.5980]'$. Similarly, for $\varpi = 0.5$ the improved estimates of $\mathbf{e}_2$ are $\hat{\mathbf{e}}_2^{SS} = \hat{\mathbf{e}}_2^{SP} = [-0.3932, -0.5035, 0.2501, 0.7241]'$, and $\hat{\mathbf{e}}_2^{DS+} = \hat{\mathbf{e}}_2^{DS} = [-0.3898, -0.5052, 0.2170, 0.7362]'$.

Example 3: Di Vesta and Walls (1970) studied mean semantic differential ratings given by fifth graders for a large number of words. These rating were obtained on the following eight scales: friendly/unfriendly (1), good/bad (2), nice/awful (3), brave/not brave (4), big/little (5), strong/weak (6), moving/still (7), and fast/slow (8). Based on the 292 words and the raw data matrix given by Di Vesta and Walls we have the following: $\hat{\lambda}_1 = 5.77$ and

$$\hat{\mathbf{e}}_1^U = [0.484, 0.569, 0.551, 0.243, 0.100, 0.232.0.087, 0.096]'.$$

Further, from the past experiment

$$\mathbf{e}_1^o = [0.5, 0.6, 0.5, 0.2, 0.1, 0.2., 0.2, 0.1]'.$$

We find $\mathcal{T}_1 = 21.89$ for testing the prior belief null hypothesis and with 7 degrees of freedom, the critical value for a 5% significance test is $\chi^2_{(7,.05)} = 14.07$. For $\varpi = 0.5$,

$$\hat{\mathbf{e}}_1^{SS} = [0.4820, 0.5845, 0.5255, 0.2215, 0.1000, 0.2160, 0.1435.0.0980]',$$

and $\hat{\mathbf{e}}_1^{SP} = \hat{\mathbf{e}}_1^P = \hat{\mathbf{e}}_1^U$.

$$\hat{\mathbf{e}}_1^{DS+} = \hat{\mathbf{e}}_1^{DS} = [0.4842, 0.5496, 0.5504, 0.2425, 0.1000, 0.2316, 0.0883, 0.0960]'.$$

We note in this example the usual estimate is not very different from $\hat{\mathbf{e}}_1^{SS}$, the positive part is the same as the Stein-rule estimator and the pretest estimators reduces to the same value as that of $\hat{\mathbf{e}}_1^{U}$. This is not surprising due to a large value of test statistic. Both $\hat{\mathbf{e}}_1^{SS}$ and Stein-rule estimators adjust the magnitude of $\hat{\mathbf{e}}_1^{U}$. More importantly, both $\hat{\mathbf{e}}_1^{DS+}$ and $\hat{\mathbf{e}}_1^{DS}$ are superior than $\hat{\mathbf{e}}_1^{U}$.

Example 4: European Marketing data are the percentage employed in different industries in Europe countries during 1979. the job categories are agriculture, mining, manufacture, power supplies, construction, service industries, finance, social and personal services, and transport and communications. The sample mean and variance-covariance matrix are:

$$\bar{\mathbf{x}} = [19.1308, 1.2538, 27.0077, 0.9077, 8.1654, 12.9577, 4.0000, 20.0231, 6.5462],'$$

and

$$\mathbf{S} = \begin{pmatrix} 241.696 & 0.540 & -73.114 & -2.340 & -13.772 & -52.484 & -9.592 & -79.291 & -12.221 \\ 0.540 & 0.941 & 3.026 & 0.148 & -0.041 & -1.761 & -1.205 & -1.862 & 0.211 \\ -73.114 & 3.026 & 79.109 & 1.016 & 5.702 & 6.538 & -3.065 & 7.379 & 3.420 \\ -2.340 & 0.148 & 1.016 & 0.142 & 0.037 & 0.348 & 0.116 & 0.340 & 0.196 \\ -13.772 & -0.041 & 5.702 & 0.037 & 2.708 & 2.681 & 0.075 & 1.778 & 0.888 \\ -52.484 & -1.761 & 6.538 & 0.346 & 2.681 & 20.983 & 4.703 & 17.905 & 1.197 \\ -9.592 & -1.205 & -3.065 & 0.116 & 0.075 & 4.703 & 7.877 & 02.063 & -0.960 \\ -79.291 & -1.862 & 7.379 & 0.340 & 1.778 & 17.905 & 2.063 & 46.643 & 5.397 \\ -12.221 & 0.211 & 3.420 & 0.197 & 0.888 & 1.197 & -0.960 & 5.397 & 1.936 \end{pmatrix},$$

respectively. The eigenvalues vector of the variance-covariance matrix is given below

$$\hat{\lambda} = [303.4850, 43.7109, 15.2184, 5.6404, 2.4445, 1.0459, 0.4209, 0.0652, 0.0019],'$$

The principal components were constructed using the variance-covariance matrix. It can be seen from the eigenvalues vector that the first two covariance principal components explain about 93%of the variability of the European data set. The first two principal components can be written as linear functions of the original nine variables as follows:

$$\hat{y}_1 = -0.8917x_1 - 0.0019x_2 + 0.2712x_3 + 0.0084x_4 + 0.0496x_5 + 0.1920x_6 + 0.0311x_7 + 0.2980x_8 + 0.0454$$
$$\hat{y}_2 = 0.0070x_1 - 0.0923x_2 - 0.7703x_3 - 0.0120x_4 - 0.0690x_5 + 0.2350x_6 + 0.1301x_7 + 0.5665x_8 + 0.0098x$$

The first principal component has a high positive loading on the agriculture variable, x_1, and small loadings on all other job categories. This component may be interpreted as distinguishing between countries with agricultural and industrial economics. The second component can be viewed as a contrast between mining and social and personal services. Further, from the past data, the initial estimates of $\mathbf{e}_1$ and $\mathbf{e}_2$ available as

$$\mathbf{e}_1^o = [-.90, -0.012, 0.25, 0.01, 0.06, 0.15, 0.04, 0.30, 0.098]',$$

and

$$\mathbf{e}_2^o = [-0.007, -0.10, -0.80, -0.015, -0.07, 0.20, 0.15, 0.45, 0.283]'$$

Example 5:

Four measurements were made of male Egyptian skulls from five different periods ranging from 4000 B.C. to 150 A.D.

Let $x_1, \cdots, x_4$ denote observed monthly rates of return for Mobil, Texaco, IBM and DEC, respectively. Then a principal component analysis using SAS yields the following summary.

$$\bar{\mathbf{x}} = [0.0171, 0.0121, 0.0103, 0.0182]'$$

and

$$\mathbf{S} = \begin{pmatrix} 0.0060 & 0.0041 & 0.0009 & 0.0024 \\ 0.0041 & 0.0057 & 0.0007 & 0.0005 \\ 0.0009 & 0.0007 & 0.0030 & 0.0018 \\ 0.0024 & 0.0005 & 0.0018 & 0.0082 \end{pmatrix}.$$

The eigenvalues and corresponding normalized eigenvectors of $\mathbf{S}$ are given below.

$$\begin{array}{llll} \hat{\lambda}_1 = & 0.01241650, & \hat{\mathbf{e}}_1^U = [0.582086, 0.511612, 0.214029, 0.594660]' \\ \hat{\lambda}_2 = & 0.00641397, & \hat{\mathbf{e}}_2^U = [-.386420, -.506979, 0.184045, 0.748185]' \\ \hat{\lambda}_3 = & 0.00251218, & \hat{\mathbf{e}}_3^U = [-.054494, -.018083, 0.959331, -.276382]' \\ \hat{\lambda}_4 = & 0.00170143, & \hat{\mathbf{e}}_4^U = [-.713362, 0.693468, 0.001664, 0.101058]' \end{array}$$

Using the above information, we obtain the first two principal components

$$\begin{array}{ll} \hat{y}_1 = & (\hat{\mathbf{e}}_1^U)'\mathbf{x} = 0.582086x_1 + 0.511612x_2 + 0.214029x_3 + 0.594660x_4 \\ \hat{y}_2 = & (\hat{\mathbf{e}}_2^U)'\mathbf{x} = -0.386420x_1 - 0.506979x_2 + 0.184045x_3 + 0.748185x_4. \end{array}$$

These components, which account for 81.72% of the total sample variance. The first component can be viewed as weighted sum of returns, weights are roughly equal except for the IBM stock. The second component represents a contrast between oil stocks and Tech stocks. Further, from the past data, the initial estimates of $\mathbf{e}_1$ and $\mathbf{e}_2$ available as

$$\mathbf{e}_1^o = [0.60, 0.50, 0.17, 0.60]', \quad \mathbf{e}_2^o = [-0.40, -0.50, 0.30, 0.70]'$$

We calculate $\mathcal{T}_1 = 0.8113$ and $\mathcal{T}_2 = 2.0058$, respectively, for testing the prior belief null hypotheses and with 3 degrees of freedom, the asymptotic critical value for a 5% significance test is $\chi^2_{(3,.05)} = 7.81$. Hence, the prior belief hypotheses are tenable in both cases. Hence, with $\varpi = 0.5$, first principal component is estimated by $\hat{\mathbf{e}}_1^{SS} = \hat{\mathbf{e}}_1^{SP} = [0.5910, 0.5058, 0.1936, 0.5973]'$, and $\hat{\mathbf{e}}_1^{DS+} = \hat{\mathbf{e}}_1^{DS} = [0.5931, 0.5045, 0.1889, 0.5980]'$. Similarly, for $\varpi = 0.5$ the improved estimates of $\mathbf{e}_2$ are $\hat{\mathbf{e}}_2^{SS} = \hat{\mathbf{e}}_2^{SP} = [-0.3932, -0.5035, 0.2501, 0.7241]'$, and $\hat{\mathbf{e}}_2^{DS+} = \hat{\mathbf{e}}_2^{DS} = [-0.3898, -0.5052, 0.2170, 0.7362]'$.

6 Concluding remarks

In this paper we continue the search started a few decades ago to find ways on improving on conventional estimators. In the context of estimation of eigen vectors of covariance matrix, we consider methods for optimally combining, under quadratic loss, estimation problem involving estimators under full and reduce models that have different sampling characteristics. It is concluded that Stein-type estimators provide a superior (in the sense of quadratic risk) basis for combining estimators and thus possibility of combining estimation problems. The proposed estimation strategy can be extended in various directions. Research on the statistical implications of proposed and related estimators is on-going. It may be worth mentioning that this is one of the two areas Professor Efron predicted for the early 21st century (RSS News, January 1995). Shrinkage and likelihood-based methods continue to play extremely useful techniques for combining estimation problems.

Although the point estimation implications of proposed estimators are encouraging, there are some roadblocks for the confidence set estimation. In linear models, at least in an asymptotic setup, the estimator or the estimating function that yields the estimator, is assumed to be (multi-) normally distributed. This enables to apply well known distribution theory relating to quadratic forms that can be used to construct suitable confidence set. In the multiparameter case, as encountered here, the Scfeffe and Tukey methods of construction of confidence sets in MANCOVA models, as well as classical likelihood ratio statistic based confidence sets are most popular. However, if we look into the Stein-type and the preliminary test estimators, even in the classical parametric cases, it is seen that are not normally distributed even in an asymptotic setup. Although the distribution theory of Stein-type estimators have been extensively investigated in the literature, the distribution theory of such quadratic norms does not produce the fruitful result so as to facilitate the construction of confidence sets having desirable properties. We need to attack this problem from a somewhat different direction, and relegate this study to a future communication.

The asymptotic distribution theory of the proposed estimators and ADR of all the estimators rest on the asymptotic normality of $\hat{\mathbf{e}}_j{}^U$ and on the asymptotic non-central chi-square distribution of the test statistic. We find that Stein type estimation strategies are superior to $\hat{\mathbf{e}}_j{}^U$ in the whole parameter space (induced by the noncentrality parameter Δ), while the performance of simple shrinkage estimator and estimation techniques based on the preliminary test rule are purely Δ dependant. It is important to note that shrinkage estimators based on Stein rule can only be used for $m > 3$. Thus, in the present investigation proposed shrinkage estimators are superior to usual estimators for $m \geq 4$ while $\hat{\mathbf{e}}_j{}^U$ is admissible for $m = 1, 2$ and $m = 3$. Thus, the use of the shrinkage estimation may be limited due to a dimensional restriction. As such, in up to trivariate cases we will be unable to use the additional information in the estimation process. In this situation, we recommend using estimators based on the preliminary-test rule. The analytical results are well supported by the computational work presented in the graphs.

References

Ahmed, S.E. (2001). Shrinkage estimation of regression coefficients from censored data with multiple observations. Empirical Bayes and Likelihood Inference, Lecture Notes in Statistics, 148, 103-120. Editors: S.E. Ahmed and N. Reid, Springer-Verlag: New York.

Ahmed, S. E. (1998). Large-Sample Estimation Strategies for Eigenvalues of a Wishart Matrix. Metrika, 47, 35-45.

Ahmed, S. E and A. K. M. E. Saleh (1999). Improved nonparametric estimation of location vector in a multivariate regression Model. *Nonparametric Statistics*, 11,51-78.

Ahmed, S.E. and Sen, P. K. (2004). Positive-part shrinkage M-estimation in linear models. *Preprint*.

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

Brekson, J. (1942). Test of significance considered as evidence. Journal of American Statistical Association, **37**, 325-335.

Brandwein, A.C. and W.E. Strawderman (1990). Stein estimation: The spherically symmetric case. Statistical Science, **5**, 356-369.

Calvo, R. A., M. G. Partridge, and M. A. Jabri (1998). A comparative study of principal component analysis techniques. In Proc. Ninth Australian Conf. on Neural Networks, Brisbane, QLD, Feb. 1998.

Cohen, A. (1968). A note on the admissibility of pooling in the analysis of variance. *Ann. Math. Statist.* **39**, 1744-1746.

Daniels, M. J. and Kass, R. E. (2001). Shrinkage estimators for covarince matrices. *Biometrics* **57**, 1173-1184

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

Di Vista, F. J. and R. T. Walls (1970). Factor analysis of the semantic attributes of 487 words and some relationships to the conceptual behavior of fifth-grade children. Journal of Educational Psychology, **61**, 6, Pt. 2.

Giles, J. A. and D. E. A. Giles (1993). Pre-test estimation and testing in econometrics: Recent developments. *Journal of Economic Review Surveys* **7**, 145-197.

Hotelling, H. (1933). Analysis of a complex statistical variables into principal components. Journal of Educational Psychology, **24**, 417-441, 498-520.

Jolliffe, I. T. (2002). Principal component analysis, 2nd edition. New York: Springer.

Judge, G .G. and Mittelhammaer, R. C. (2004). A Semiparametric basis for combining estimation problem under quadratic loss. *Journal of American Statistical Association*, **99**, 479-487.

Hoffmann, K. (1992). Improved estimation of distribution parameters: Stein-Type estimators. B. G. Teubner Verlgsgesellschaft: Stuttgart.

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

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

Jolicoeur, P. and Mosimann (1960). Size and shape variation in the painted turtle: a principal component analysis. Growth, **24**, 339-354.

Jolliffe, I.T. (1986). Principal Component Analysis. Springer-Verlag, New York.

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

Magnus, J. R. (1999). The traditional pretest estimator. Theory of Probability and its Applications, 44, 293-308.

Ohanti, K.(1999). MSE performance of a heterogeneous pre-test estimator. Statistics & Probability Letters, 41, 65-71.

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

Reif, J. and VIcek, K. (2002). Optimal pre-test estimators in regression. Journal of Econometrics, 110, 91-102.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceeding of the third Berkeley symposium on Mathematical statistics and Probability. University of California Press. volume 1, 197-206.

Stigler, S. M. (1990). The 1988 Neyman Memorial Lecture: A Galtonian perspective on shrinkage estimators Statistical Science. **5**, 147-155.

Tipping, M. E. and C. M. Bishop (1999). Probabilistic principal component analysis. Journal of the Royal Statistical Society, Series B, 21, 611–622.