Journal of Multivariate Analysis 122 (2013) 190–201

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On extension of some identities for the bias and risk functions in elliptically contoured distributions

Sévérien Nkurunziza *, Fuqi Chen

Department of Mathematics and Statistics, University of Windsor, Ontario N9B 3P4, Canada

ARTICLE INFO

Article history:
Received 29 September 2012
Available online 29 July 2013

AMS subject classifications:
62H12
62F25

Keywords:
Bias function
Elliptically contoured distribution
Kronecker-product
Matrix estimation
Risk function
Stein rules

ABSTRACT

In this paper, we are interested in an estimation problem concerning the mean parameter of a random matrix whose distribution is elliptically contoured. We derive two general formulas for the bias and risk functions of a class of multidimensional shrinkage-type estimators. As a by product, we generalize some recent identities established in Gaussian sample cases for which the shrinking random part is a single Kronecker-product. Here, the variance–covariance matrix of the shrinking random part is the sum of two Kronecker-products.

1. Introduction

Over the years, a very rich literature has evolved on modeling diverse phenomena by using a class of distributions which is more robust and more realistic than the Gaussian distribution. In particular, as discussed for example in [1], elliptically contoured distributions are a useful alternative to the multivariate Gaussian paradigm, not only in statistics, but in several areas of applications such as actuarial science (see [7,11]), economics and finance (see [5]). Also, as explained in [18], many test statistics and optimality properties associated with the Gaussian distribution case remain unchanged for elliptically contoured distributions. For recent references about the advantages of elliptically contoured distributions, we quote [13] and the references therein.

In this paper, we study an estimation problem of the mean parameter matrix of the random $q \times k$-matrix whose distribution is elliptically contoured. In particular, we consider the case where the target parameter $\boldsymbol{\theta}$ represents the mean of the random $q \times k$-matrix $\boldsymbol{X}$ whose distribution is elliptically contoured with a known covariance–variance. Further, we study the case where some imprecise knowledge about the target parameter is available. By combining the sample information and uncertain prior knowledge, we propose a class of shrinkage-type estimators for the parameter $\boldsymbol{\theta}$. Also, we establish the bias and risk functions of the proposed class of estimators. Let $\boldsymbol{X} \sim \mathcal{E}_{q\times k}(\boldsymbol{\theta},\ \boldsymbol{\Lambda} \otimes \boldsymbol{\Omega}; g)$ be a matrix random variate elliptically contoured distributions with mean $\boldsymbol{\theta}$ and covariance–variance $\boldsymbol{\Lambda} \otimes \boldsymbol{\Omega}$, where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ are known positive definite matrices of rank q and k respectively, $A \otimes B$ denote the Kronecker-product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and g is the probability density function (pdf) generator.

Without further assumption, the problem considered is (to our best knowledge) insoluble. To this end, we concentrate our study to the subclass of scale mixtures of normals. The subclass under consideration includes for example multivariate

* Corresponding author.
E-mail address: severien@uwindsor.ca (S. Nkurunziza).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.07.005

Table 1
Examples of pdf with the respective weighting functions.

Distribution	pdf "$f_U(\boldsymbol{x})$"	The function "$\boldsymbol{\omega}(t)$"
Gaussian	$(2\pi)^{-qk/2}\,\lvert\boldsymbol{\Lambda}\rvert^{-\frac{k}{2}}\,\lvert\boldsymbol{\Omega}\rvert^{-\frac{q}{2}}\,\exp\left[-\frac{1}{2}g_0(\boldsymbol{x})\right]$	$\delta(t-1)$
Pearson type VII	$\kappa\,(m,\ qk,\ q_0)\,[1+g_0(\boldsymbol{x})/q_0]^{-m}, m > qk/2$	$\frac{t^{m-qk/2-1}\exp(-q_0t/2)}{(q_0/2)^{qk/2-m}\Gamma(m-qk/2)}, t > 0$
t with q_0 d.f.	$\kappa\left(\frac{q_0+qk}{2},\ qk,\ q_0\right)[1+g_0(\boldsymbol{x})/q_0]^{-\frac{q_0+qk}{2}}$	$\frac{\left(\frac{q_0t}{2}\right)^{\frac{q_0}{2}}e^{-\frac{q_0t}{2}}}{t\,\Gamma(\frac{q_0}{2})}, t > 0$

Gaussian, t, Pearson type II and VII as well as Kotz. Formally, as in [6], the pdf of $U = \mathrm{Vec}(\boldsymbol{X})$ is assumed to be written as

$$f_U(x) = \int_0^\infty f_{\mathcal{N}\left(Vec(\boldsymbol{\theta}),\, z^{-1}\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}\right)}(x)\boldsymbol{\omega}(z)dz, \tag{1.1}$$

where $f_{\mathcal{N}(\boldsymbol{\mu},\,\boldsymbol{\Sigma})}$ denotes the pdf of a random vector which follows a normal distribution with mean $\boldsymbol{\mu}$ and variance–covariance $\boldsymbol{\Sigma}$, with $\boldsymbol{\omega}(.)$ a weighting function as defined for example in [6] and the references therein. For the convenience of the reader, we present in Table 1 examples of pdfs which satisfy the condition in (1.1) along with their corresponding weighting function $\boldsymbol{\omega}(.)$. To introduce some notations used in Table 1, let $\delta(.)$ and $\Gamma(.)$ denote the Dirac delta function and the gamma function respectively i.e.

$$\int_0^\infty \delta(z)dz = 1 \quad \text{and} \quad \int_{-\infty}^\infty f(z)\delta(z)dz = f(0) \text{ for every Borel measurable function } f(.),$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\exp(-t)dt, \quad \alpha > 0.$$

Also, let

$$g_0(\boldsymbol{x}) = \mathrm{trace}\left(\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\theta})'\right), \qquad \kappa\,(m,\ qk,\ q_0) = \frac{\lvert\boldsymbol{\Lambda}\rvert^{-\frac{k}{2}}\,\lvert\boldsymbol{\Omega}\rvert^{-\frac{q}{2}}\,\Gamma\,(m)}{(q_0\pi)^{qk/2}\,\Gamma\left(m-\frac{qk}{2}\right)}.$$

It should be noticed that the proposed statistical model, for which the random sample satisfies the condition in (1.1), is more general than that in [15] for which the matrix $\boldsymbol{X}$ is assumed to be Gaussian. Also, we generalize some identities which are given in [10, Theorems 1 and 2], as well as their extension given in [15]. In particular, the derived results are useful in computing the bias and the risk functions of the proposed class of shrinkage-type estimators.

Before presenting the class of estimators under consideration, let us point out that, in the absence of restrictions on the parameter, $\boldsymbol{X}$ is the least squares estimator as well as the maximum likelihood estimator provided that similar conditions as for example in [3] hold. However, when the parameter matrix satisfies some linear constraints, the unrestricted maximum likelihood estimator (UMLE) is dominated by the restricted maximum likelihood estimator (RMLE). In intermediate situations of uncertain constraints, the above estimators may perform poorly. More specifically, for the case where $\boldsymbol{X}$ is a q-column random vector, it has been shown that if $q \geqslant 3$, the James–Stein estimator dominates in mean-square sense the UMLE $\boldsymbol{X}$. In these scenarios, shrinkage estimators dominate the UMLE over the whole parameter space. Further, as we move away from the hypothesized restriction, shrinkage estimators also dominate the RMLE.

More specifically, consider the estimation problem of the parameter matrix $\boldsymbol{\theta}$ when the parameter may satisfy the following restriction:

$$\boldsymbol{L}_1\boldsymbol{\theta} = \boldsymbol{d}_1, \qquad \boldsymbol{\theta}\boldsymbol{L}_2 = \boldsymbol{d}_2 \tag{1.2}$$

with $\boldsymbol{L}_1\boldsymbol{d}_2 = \boldsymbol{d}_1\boldsymbol{L}_2$, where $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are respectively $p \times q$ and $k \times m$-known full rank matrices with $p < q$ and $m \leqslant k$, and $\boldsymbol{d}_i$, $i = 1, 2$, are known, respectively $p \times k$ and $q \times m$-matrices. Note that the relation $\boldsymbol{L}_1\boldsymbol{d}_2 = \boldsymbol{d}_1\boldsymbol{L}_2$ is not an additional restriction since this follows directly from (1.2); thus, this is the unavoidable consequence of the two previous constraints. Accordingly, the constraint in (1.2) is more general than that given in [8, p. 168] in the context of multivariate linear models. Following the interpretation given in the quoted work, the constraint in (1.2) may correspond for example to the case where the population treatment mean profiles are parallel or rather identical. Below we give an example of application context and two explicit motivating examples on the multivariate regression model for which the above constraint is useful.

1.1. Application context and motivating example

1.1.1. Application context

Consider the following multivariate regression $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{Y}$ is the response $n \times m$-matrix, $\boldsymbol{X}$ is a known (non-random) $n \times k$ matrix, and $\boldsymbol{\epsilon}$ is an unobserved noise $n \times m$-random matrix which is assumed to follow an elliptically contoured distribution with mean $\boldsymbol{0}$. Several authors studied the inference problem concerning the parameter matrix $\boldsymbol{\beta}$ (see [12,13,17] among others). As an illustrative application of the proposed methodology, we consider the estimation problem of $\boldsymbol{\beta}$ when this may or may not satisfy the restriction

$$\boldsymbol{L}_1^*\boldsymbol{\beta}=\boldsymbol{d}_1^*, \qquad \boldsymbol{\beta}\boldsymbol{L}_2^*=\boldsymbol{d}_2^* \tag{1.3}$$

where $\boldsymbol{L}_1^*$, $\boldsymbol{L}_2^*$, $\boldsymbol{d}_1^*$, $\boldsymbol{d}_2^*$ are similar to $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{d}_1$, $\boldsymbol{d}_2$ given in the constraint in (1.2). In this context, as explained in [17], the constraint in (1.3) can be interpreted by viewing the first relation $\boldsymbol{L}_1^*\boldsymbol{\beta}=\boldsymbol{d}_1^*$ as a constraint which sets p independent linear combinations of the rows of $\boldsymbol{\beta}$. Statistically, this restriction takes into account the correlation among the k explanatory variables. In a similar way, the second relation $\boldsymbol{\beta}\boldsymbol{L}_2^*=\boldsymbol{d}_2^*$ can be viewed as a restriction which defines q independent linear combinations of the columns of $\boldsymbol{\beta}$, and this takes care of the correlation among the m dependent variables.

As far as the estimation problem in (1.3) is concerned, [17] proposed some shrinkage estimators which dominate the unrestricted estimator. The estimators proposed by the quoted authors are members of the class of estimators given here in the context of elliptically contoured distribution. Further, the asymptotic distributional risk and bias given in the quoted paper can be derived by applying the results given in this paper. For our paper to be self-contained, we recall below the estimators of $\boldsymbol{\beta}$ which are given in [17]. To this end, let $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ denote, respectively, the unrestricted and the restricted estimators. Further, let $\widehat{\boldsymbol{\beta}}^S$ and $\widehat{\boldsymbol{\beta}}^{S+}$ denote, respectively the shrinkage and positive-part shrinkage estimators. Briefly, we have (for more details, see [17]),

$$\widehat{\boldsymbol{\beta}}=\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}, \qquad \widetilde{\boldsymbol{\beta}}=\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{J}^*\boldsymbol{L}_1^*\widehat{\boldsymbol{\beta}}+\boldsymbol{J}^*\boldsymbol{d}_1^*\right)\left(\boldsymbol{I}_m-\boldsymbol{L}_2^*\boldsymbol{P}^*\right)+\boldsymbol{d}_2^*\boldsymbol{P}^*,$$

where $\boldsymbol{J}^*=\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{L}_1^{*'}\left(\boldsymbol{L}_1^*\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{L}_1^{*'}\right)^{-1}$ and $\boldsymbol{P}^*=\left(\boldsymbol{L}_2^{*'}\boldsymbol{L}_2^*\right)^{-1}\boldsymbol{L}_2^{*'}$. Further, we have

$$\widehat{\boldsymbol{\beta}}^s=\widetilde{\boldsymbol{\beta}}+\{1-(pm-2)\varphi_n^{*-1}\}(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}}), \qquad \widehat{\boldsymbol{\beta}}^{s+}=\widetilde{\boldsymbol{\beta}}+\max\{0,\ 1-(pm-2)\varphi_n^{*-1}\}(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}}),$$

where

$\boldsymbol{\varphi}_n^*=n\mathrm{trace}\left[\left(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}}\right)'\boldsymbol{L}_1^{*'}\left(\boldsymbol{L}_1^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}_1^{*'}\right)^{-1}\boldsymbol{L}_1^*\left(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}}\right)\left(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^{-1}\right]$. In the normal sample case, the mean-square error (MSE) of the estimators of $\boldsymbol{\beta}$ correspond to the asymptotic distributional risk given in [17]. Further, by using the functions given in Examples 3.2–3.3 (see Section 3), one can get the MSE of the estimators $\widehat{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}^s$, $\widehat{\boldsymbol{\beta}}^{s+}$. To save the space of this paper, these expressions are omitted.

1.1.2. Motivating examples

In this subsection, we present two motivating examples which show the interest of the constraint (1.2) in the context of multivariate regression model. The first motivating example is given in [17] where a similar constraint is studied. Also, the second motivating example is described and analyzed in the above quoted paper. In order to save the space of this paper, we do not report the analysis of the data set of these examples. Nevertheless, for more details and analysis of these motivating examples, the reader is referred to [17].

1.1.2.1. The first motivating example. We consider the data set described in [9] which consists of measurements made on specimens of the birds Martes Americana. Briefly, they consist of 4 (i.e. $k=4$) explanatory variables and 2 (i.e. $m=2$) dependent variables.

Namely, the explanatory variables are:

X_1: length of humerus; X_2: width of humerus; X_3: length of femur; X_4: width of femur. The explanatory variables in this example are considered to be fixed (i.e. non-random).

Further, the response variables are weight (Y_1) and volume (Y_2) and, these variables are assumed to be standardized in order to be unit free.

By using a principal component analysis of the logarithms of the Xs, [9] established that X_2 varied as the power 1.5 of X_1, so that $\log(X_2)=a_1+1.5\log(X_1)$, and that X_3, X_4 were both proportional to X_1, so that $\log(X_3)=a_2+\log(X_1)$ and $\log(X_4)=a_3+\log(X_1)$. Here, a_1, a_2, a_3 are taken as the means of $\log(X_2)-1.5\log(X_1)$, $\log(X_3)-\log(X_1)$ and $\log(X_4)-\log(X_1)$ respectively.

Thus, if the logarithms of the X_i are used as the explanatory variables, then the joint prediction of Y_1 and Y_2 can be modeled by using the above multivariate regression model with $m=2$ and $k=4$. Also, to reflect the relationships between the X_i, one considers that the rows of $\boldsymbol{L}_1^*$ may be taken as $(1,-1.5,0,0)$, $(1,0,-1,0)$ and $(1,0,0,-1)$ with $\boldsymbol{d}_1^*=(1,1)\otimes(a_1,a_2,a_3)'$. Further, to reflect the high positive correlation that is expected between Y_1 and Y_2, one can set $\boldsymbol{L}_2^*=(1,-1)'$ and $\boldsymbol{d}_2^*=0$.

1.1.2.2. The second motivating example. The second motivating example is based on a data set which is given in [4, pp. 357–360], and described in [8, pp. 193] as well as in [17]. The data consists of 8 measurements on each of the four response variates taken on 13 different types of root-stocks of apple trees. The 4 response variables are trunk girth in mm (Y_1); extension growth (cm) (Y_2) at 4 years after planting; trunk girth (mm) (Y_3) at 15 years after planting; and weight (lb) of tree above ground (Y_4) at 15 years after planting. As described for example [8,17], the explanatory variables are categorical so that the design matrix $\boldsymbol{X}=\boldsymbol{I}_{13}\otimes\boldsymbol{e}_8$ where $\boldsymbol{e}_n$ denotes an n-column vector with all entries equal to 1. Further, as justified in [17], the first restriction of the interest is $\mathbf{L}_1^*\,\boldsymbol{\beta}=\mathbf{0}$ with $\boldsymbol{L}_1^*=(\boldsymbol{I}_{12},-\boldsymbol{e}_{12})$, and to incorporate the prior information about the Ys, we use the fact that Y_1 and Y_2 are expected to be highly correlated, and Y_3 is expected to be highly correlated with Y_4. To this end, we set $\boldsymbol{\beta L}_2^* = \mathbf{0}$ where $\boldsymbol{L}_2^* = \left[(1, -1, 0, 0)', (0, 1, 0, 0)', (0, 0, 1, -1)'\right]'$.

1.2. The class of shrinkage estimators and some notations

In this subsection, we define some notations used throughout the paper and we present the restricted estimator of the parameter matrix $\boldsymbol{\theta}$ along with a class of shrinkage-type estimators for which the bias and risk functions are established.

Let $\boldsymbol{J} = \boldsymbol{\Lambda L}_1' \left(\boldsymbol{L}_1 \boldsymbol{\Lambda L}_1'\right)^{-1}$ and let $\boldsymbol{P} = \left(\boldsymbol{L}_2' \boldsymbol{\Omega L}_2\right)^{-1} \boldsymbol{L}_2' \boldsymbol{\Omega}$. As given in Proposition A.1 in the Appendix, under the constraint in (1.2), the RMLE is given by

$$\widetilde{\boldsymbol{\theta}} = (\boldsymbol{X} - \boldsymbol{J L}_1 \boldsymbol{X} + \boldsymbol{J d}_1)(\boldsymbol{I}_k - \boldsymbol{L}_2 \boldsymbol{P}) + \boldsymbol{d}_2 \boldsymbol{P}. \tag{1.4}$$

Following the notations in [15], let $\boldsymbol{A}$ be a matrix and let $\|\boldsymbol{A}\|^2_{\boldsymbol{\Xi}_1, \boldsymbol{\Xi}_2} = \text{trace}\left(\boldsymbol{A}' \boldsymbol{\Xi}_1 \boldsymbol{A} \boldsymbol{\Xi}_2\right)$ where $\boldsymbol{\Xi}_1$, $\boldsymbol{\Xi}_2$ are known nonnegative definite matrices. Further, let h be a known Borel measurable and real-valued integrable function, and let us consider the following class of estimators:

$$\widehat{\boldsymbol{\theta}}(h) = \widetilde{\boldsymbol{\theta}} + h\left(\left\|\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right\|^2_{\boldsymbol{\Xi}_1, \boldsymbol{\Xi}_2}\right)\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right). \tag{1.5}$$

With respect to the family of the distributions of the estimator $\widehat{\boldsymbol{\theta}}(h)$ in (1.5), note that the family in [15] is based on a Gaussian family which is a special case of that considered here. Also, the uncertain constraint in (1.2) is less restrictive and thus more versatile than that given in [15]. Indeed, if the constraint in (1.2) holds, then $\boldsymbol{L}_1 \boldsymbol{\theta L}_2 = \boldsymbol{d}$ where $\boldsymbol{d} = (\boldsymbol{L}_1 \boldsymbol{d}_2 + \boldsymbol{d}_1 \boldsymbol{L}_2)/2$. Further, note that the restricted and unrestricted estimators are members of the class of estimators in (1.5). Indeed, the restricted and the unrestricted estimators can be obtained by taking $h \equiv 0$ and $h \equiv 1$ respectively. Finally, Stein-type estimators are members of the class of estimators in (1.5). Indeed, set $h(x) = (1 - a/x)$, $x > 0$ for some constant $a > 0$. The estimator in (1.5) becomes $\widehat{\boldsymbol{\theta}}^{s} = \widetilde{\boldsymbol{\theta}} + \left[1 - a/\,\text{trace}\left(\boldsymbol{\Xi}_2\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)' \boldsymbol{\Xi}_1\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)\right)\right]\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)$ which is a well known shrinkage estimator for $\boldsymbol{\theta}$. Further, taking $h(x) = \max(0, 1 - a/x)$, $x > 0$ for some constant $a > 0$, we get $\widehat{\boldsymbol{\theta}}^{s+} = \widetilde{\boldsymbol{\theta}} + \max\left\{0, 1 - a/\,\text{trace}\left(\boldsymbol{\Xi}_2\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)' \boldsymbol{\Xi}_1\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)\right)\right\}\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)$ which is well known as a positive-part shrinkage estimator for $\boldsymbol{\theta}$. Also, the class in (1.5) includes the preliminary test estimators as studied for example in [14] and the references therein. Briefly, the above class of the estimators combines both sample information and imprecise prior knowledge from the uncertain constraint in (1.2). Thus, the proposed class of estimators can be seen as a shrinkage-type estimator of $\boldsymbol{\theta}$.

As a common practice in point estimation, the bias and risk functions are needed in order to evaluate the performance of the proposed class of estimators. For the statistical model studied here, the derivation of these quantities is mathematically complex. The major difficulty consists in the fact that the distribution of the shrinking random part $\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}$ is not Gaussian. In addition, the covariance–variance matrix of the shrinking random part is a sum of two Kronecker-products.

In the sequel, let $B\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}\right)$ denote the bias function of $\widehat{\boldsymbol{\theta}}$, and let $R\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right)$ denote the risk function of $\widehat{\boldsymbol{\theta}}$ with $\boldsymbol{W}$ a nonnegative definite matrix. Recall that

$$B\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}\right) = E\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\right].$$

For computing the risk function $R\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right)$, we use the following quadratic loss function $L\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = \text{trace}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \boldsymbol{W} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\right]$. Thus, we have

$$R\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = E\left[\text{trace}\left(\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \boldsymbol{W} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\right)\right].$$

The remainder of this paper is organized as follows. Section 2 presents the mathematical background of this paper. Namely, it gives three theorems which are used in deriving the bias and risk functions of the proposed class of estimators. Section 3 gives the risk and bias function formulas, that is the second main contribution of this paper. Section 4 gives concluding remarks. Finally, the proofs of the main identities given in Section 3 and related details are given in the Appendix.

2. Some useful identities and their extension

In this section, we present the results which constitute the first contribution of this paper. In particular, we derive some mathematical results which generalize the identities given in literature for the Gaussian random matrix case. More specifically, our identities generalize those of [15] which in turn are also extensions of Theorems 1 and 2 in [10]. The established identities play an important role in deriving the bias and risk functions which constitute the second contribution of this paper.

Let $\mathcal{G}am(\alpha,\beta;\lambda)$ denote a random variable which follows noncentral gamma distribution with parameters α,β with noncentrality parameter λ. Let $\boldsymbol{\omega}(.)$ denote the weighting function as introduced in (1.1), and let h^i denote the ith-power of h, for some i nonnegative integer i.e. $h^0=1$ and $h^i(x)=\underbrace{h(x)\times h(x)\times\cdots\times h(x)}_{i\text{ terms}}$, $i\in\{1,2,3,\ldots\}$. Further, for $i=0,1,2,\ldots,$ $n=1,2,3,\ldots,$ let

$$\boldsymbol{\psi}^{(1)}_{i,n}(x,y)=\int_0^\infty E\left[h^i\left(\mathcal{G}am\left(\frac{n}{2},\frac{2y}{t};\frac{tx}{y}\right)\right)\right]\boldsymbol{\omega}(t)dt,$$
$$\boldsymbol{\psi}^{(2)}_{i,n}(x,y)=y\int_0^\infty t^{-1}E\left[h^i\left(\mathcal{G}am\left(\frac{n}{2},\frac{2y}{t};\frac{tx}{y}\right)\right)\right]\boldsymbol{\omega}(t)dt,\quad x\geqslant 0,\ y>0. \tag{2.1}$$

Theorem 2.1. *Let Λ_i and Υ_i, $i=1,2$, be respectively $q\times q$ and $k\times k$ positive semi-definite matrices, with rank$(\Lambda_1)=q_1\leqslant q$ and rank$(\Upsilon_1)=p\leqslant k$. Let $\boldsymbol{X}\sim\mathcal{E}_{q\times k}\left(\boldsymbol{M},\ \sum_{i=1}^2(\boldsymbol{\Upsilon}_i\otimes\boldsymbol{\Lambda}_i);g\right)$, and let Ξ_3 and Ξ_4 be symmetric and positive definite matrices such that $\Xi_3^{\frac12}\Upsilon_1\Xi_3^{\frac12}$, $\Xi_3^{\frac12}\Upsilon_2\Xi_3^{\frac12}$ and $\Xi_4^{\frac12}\Lambda_1\Xi_4^{\frac12}$ are idempotent matrices and $\Xi_3\Upsilon_1\Xi_3M=\Xi_3M$. Then, for any h Borel measurable and real-valued integrable function, we have*

$$E\left[h\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\boldsymbol{X}\right]=\boldsymbol{\psi}^{(1)}_{1,pq_1+2}\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{M}'\boldsymbol{\Xi}_3\boldsymbol{M}\right),\ 1\right)\boldsymbol{M},$$

where $\boldsymbol{\psi}^{(1)}_{1,pq_1+2}(.,\ .)$ is defined in (2.1). □

The proof of the theorem is given in the Appendix.

Remark 2.1. For the Gaussian matrix variate case, the quantities $\boldsymbol{\psi}^{(1)}_{i,n}$ and $\boldsymbol{\psi}^{(2)}_{i,n}$ become

$$\boldsymbol{\psi}^{(1)}_{i,n}(x,\ 1)=\boldsymbol{\psi}^{(2)}_{i,n}(x,\ 1)=E\left[h^i\left(\chi^2_n(x)\right)\right],\quad x\geqslant 0. \tag{2.2}$$

Thus, Theorem 2.1 generalizes Theorem 2.1 in [15] for which only a single Kronecker-product is considered for a Gaussian random sample.

By using Theorem 2.1, we deduce the following theorem that is useful in deriving $R\left(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}\right)$.

Theorem 2.2. *Let $\boldsymbol{\Lambda}_{1i}$ and let $\boldsymbol{\Upsilon}_{1i}$, $i=1,2$, be respectively $q\times q$ and $k\times k$ nonnegative definite matrices with rank$(\boldsymbol{\Lambda}_{11})=q_1\leqslant q$, rank$(\boldsymbol{\Upsilon}_{11})=p\leqslant k$. Further, let $\boldsymbol{\Xi}_3$ and $\boldsymbol{\Xi}_4$ be respectively $k\times k$ and $q\times q$ symmetric and positive definite matrices such that $\Xi_3^{\frac12}\Upsilon_{11}\Xi_3^{\frac12}$, $\Xi_3^{\frac12}\Upsilon_{12}\Xi_3^{\frac12}$ and $\Xi_4^{\frac12}\Lambda_{11}\Xi_4^{\frac12}$ are idempotent matrices and $\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_{11}\boldsymbol{\Xi}_3M_1=\boldsymbol{\Xi}_3M_1$. Also, let $\boldsymbol{\Upsilon}_{2i}$, $\boldsymbol{\Lambda}_{2i}$, $i=1,2,\ldots,m$, be nonnegative definite matrices and let $\left(\boldsymbol{X}',\boldsymbol{Y}'\right)'\sim\mathcal{E}_{2q\times k}\left((\boldsymbol{M}_1,\boldsymbol{M}_2),\begin{pmatrix}\sum_{i=1}^2(\boldsymbol{\Upsilon}_{1i}\otimes\boldsymbol{\Lambda}_{1i}) & \boldsymbol{0}\\ \boldsymbol{0} & \sum_{j=1}^m(\boldsymbol{\Upsilon}_{2j}\otimes\boldsymbol{\Lambda}_{2j})\end{pmatrix};g\right)$. Then, for any h Borel measurable and real-valued integrable function, and any positive semi-definite matrix $\boldsymbol{A}$, we have*

$$E\left[h\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_{11}\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{X}\right]=\boldsymbol{\psi}^{(1)}_{1,pq_1+2}\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{M}_1'\boldsymbol{\Xi}_3\boldsymbol{M}_1\right),\ 1\right)\boldsymbol{M}_2'\boldsymbol{A}\boldsymbol{M}_1,$$

where $\boldsymbol{\psi}^{(1)}_{1,pq_1+2}(.,\ .)$ is defined in (2.1). □

Proof. Since h is a real-valued function, we have

$$\begin{aligned}E\left[h\left(\mathrm{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_{11}\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{X}\right]&=E\{(E(\boldsymbol{Y}|\boldsymbol{X}))'\boldsymbol{A}\ h\left(\mathrm{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_{11}\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\boldsymbol{X}\},\\&=\boldsymbol{M}_2'\boldsymbol{A}E\left[h\left(\mathrm{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_{11}\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\boldsymbol{X}\right],\end{aligned}$$

and then, the proof is completed by applying the above Theorem 2.1. □

Further, in establishing the risk function $R\left(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta};\boldsymbol{W}\right)$ we use the following theorem.

Theorem 2.3. *Suppose that the assumptions of Theorem 2.1 hold and let $\boldsymbol{A}$ be a nonnegative definite matrix. Then, for any h Borel measurable and real-valued integrable function, we have*

$$\begin{aligned}&E\left[h\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)trace\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right)\right]=\boldsymbol{\psi}^{(2)}_{1,pq_1+2}\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{M}'\boldsymbol{\Xi}_3\boldsymbol{M}\right),\ 1\right)\\&\quad\times trace\left(\boldsymbol{A}\boldsymbol{\Upsilon}_1\right)trace\left(\boldsymbol{\Xi}_4^{-1}\right)+\boldsymbol{\psi}^{(1)}_{1,pq_1+4}\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{M}'\boldsymbol{\Xi}_3\boldsymbol{M}\right),\ 1\right)trace\left(\boldsymbol{M}'\boldsymbol{A}\boldsymbol{M}\right)\\&\quad+\boldsymbol{\psi}^{(2)}_{0,p_1q}(0,\ 1)\,\boldsymbol{\psi}^{(1)}_{1,pq_1}\left(trace\left(\boldsymbol{\Xi}_4\boldsymbol{M}'\boldsymbol{\Xi}_3\boldsymbol{M}\right),\ 1\right)trace\left(\boldsymbol{A}\left(\boldsymbol{\Xi}_3^{-1}-\boldsymbol{\Upsilon}_1\right)\right)trace\left(\boldsymbol{\Lambda}_2\right).\end{aligned}$$ □

The proof is given in the Appendix. Note that for the special case where the random sample is Gaussian with $\boldsymbol{\Lambda}_2 = \mathbf{0}$, Theorem 2.3 gives Theorem 1.2 in [15] which was an extension of Theorem 2 given in [10]. Indeed, this last result can be deduced from Theorem 2.3 by taking $k = 1$, $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Xi} = 1$, $\boldsymbol{\Lambda}_1 = \boldsymbol{I}_q$.

Using Theorems 2.2 and 2.3, we establish in the next section the bias and risk functions of the estimator $\widehat{\boldsymbol{\theta}}(h)$ as given in (1.5). As intermediate result, we present below a proposition which gives the joint distribution of the unrestricted and restricted estimators.

Proposition 2.1. *We have*

$$\left(\boldsymbol{X}' - \boldsymbol{\theta}', \boldsymbol{X}' - \widetilde{\boldsymbol{\theta}}'\right)' \sim \mathscr{E}_{2q\times k}\left(\left(\mathbf{0}, -\boldsymbol{\delta}'\right)', \begin{pmatrix} \boldsymbol{\Lambda}\otimes\boldsymbol{\Omega} & \boldsymbol{\Omega}_{22} \\ \boldsymbol{\Omega}_{22} & \boldsymbol{\Omega}_{22}\end{pmatrix}; g\right), \quad \text{and}$$

$$\left(\boldsymbol{X}' - \widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}' - \boldsymbol{\theta}'\right)' \sim \mathscr{E}_{2q\times k}\left(\left(-\boldsymbol{\delta}', \boldsymbol{\delta}'\right)', \begin{pmatrix} \boldsymbol{\Omega}_{22} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Upsilon}_{22}\end{pmatrix}; g\right),$$

where $\boldsymbol{\Omega}_{22} = \boldsymbol{\Upsilon}^*\otimes\boldsymbol{\Omega} - \boldsymbol{\Upsilon}^*\otimes\boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P} + \boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}$, $\boldsymbol{\Upsilon}^* = \boldsymbol{J}\boldsymbol{L}_1\boldsymbol{\Lambda}$, $\boldsymbol{\Upsilon}_{22} = \boldsymbol{\Lambda}\otimes\boldsymbol{\Omega} - \boldsymbol{\Omega}_{22}$, $\boldsymbol{\delta} = -\boldsymbol{J}\boldsymbol{\delta}_1 - \boldsymbol{J}\boldsymbol{\delta}_1\boldsymbol{L}_2\boldsymbol{P} - \boldsymbol{\delta}_2\boldsymbol{P}$, *with* $\boldsymbol{\delta}_1 = \boldsymbol{L}_1\boldsymbol{\theta} - \boldsymbol{d}_1$ *and* $\boldsymbol{\delta}_2 = \boldsymbol{\theta}\boldsymbol{L}_2 - \boldsymbol{d}_2$. □

The proof follows directly from the properties of elliptically contoured distributions (see for example [2]) along with some algebraic computations. To simplify the notation, let $\boldsymbol{\rho} = \boldsymbol{X} - \boldsymbol{\theta}$, $\boldsymbol{\xi} = \boldsymbol{X} - \widetilde{\boldsymbol{\theta}}$, and let $\boldsymbol{\zeta} = \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$ where $\widetilde{\boldsymbol{\theta}}$ is the RE as defined in (1.4). Also, let $\Delta = \text{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{\delta}'\boldsymbol{\Xi}_3\boldsymbol{\delta}\right)$.

3. The bias and risk functions

In this section, we give another contribution of this paper. In particular, we apply Theorems 2.1–2.3 in order to derive the bias and the risk functions of the proposed class of estimators. More specifically, the bias and risk functions of the class of estimators in (1.5) for which the matrices $\boldsymbol{\Xi}_1$ and $\boldsymbol{\Xi}_2$ are respectively taken as $\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}^*\boldsymbol{\Xi}_3$ and $\boldsymbol{\Xi}_4$ for all symmetric positive definite matrices $\boldsymbol{\Xi}_3$ and $\boldsymbol{\Xi}_4$ such that $\boldsymbol{\Xi}_3^{\frac12}\boldsymbol{\Upsilon}^*\boldsymbol{\Xi}_3^{\frac12}$, $\boldsymbol{\Xi}_3^{\frac12}(\boldsymbol{\Lambda} - \boldsymbol{\Upsilon}^*)\boldsymbol{\Xi}_3^{\frac12}$ and $\boldsymbol{\Xi}_4^{\frac12}\left(\boldsymbol{\Omega} - \boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right)\boldsymbol{\Xi}_4^{\frac12}$ are idempotent. As an example, note that $\boldsymbol{\Xi}_3 = \boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{\Xi}_4 = \boldsymbol{\Omega}^{-1}$ satisfy the conditions.

Theorem 3.1. *If the conditions of Proposition 2.1 hold, then the bias of* $\widehat{\boldsymbol{\theta}}(h)$ *is*

$$B\left(\widehat{\boldsymbol{\theta}}(h), \boldsymbol{\theta}\right) = -\boldsymbol{\delta} + \boldsymbol{\psi}^{(1)}_{1,pk+2}(\Delta, 1)\,\boldsymbol{\delta}. \quad \square$$

Proof. We have $B\left(\widehat{\boldsymbol{\theta}}(h), \boldsymbol{\theta}\right) = E\left\{\boldsymbol{\eta} + h\left(\|\boldsymbol{\xi}\|^2_{\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2}\right)\boldsymbol{\xi}\right\} = -\boldsymbol{\delta} + E\left\{h\left(\|\boldsymbol{\xi}\|^2_{\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2}\right)\boldsymbol{\xi}\right\}$ with $\boldsymbol{\Xi}_1 = \boldsymbol{\Xi}_3\boldsymbol{\Upsilon}^*\boldsymbol{\Xi}_3$ and $\boldsymbol{\Xi}_2 = \boldsymbol{\Xi}_4$. Therefore, using Theorem 2.1, we get $E\left\{h\left(\|\boldsymbol{\xi}\|^2_{\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2}\right)\boldsymbol{\xi}\right\} = -\boldsymbol{\psi}^{(1)}_{1,pk+2}(\Delta, 1)\,\boldsymbol{\delta}$, which completes the proof. □

Theorem 3.2. *If the conditions of Proposition 2.1 hold, then the risk of* $\widehat{\boldsymbol{\theta}}(h)$ *is*

$$\begin{aligned} R\left(\widehat{\boldsymbol{\theta}}(h), \boldsymbol{\theta}; \boldsymbol{W}\right) = {} & \boldsymbol{\psi}^{(2)}_{0,pk}(0, 1)\,\text{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Lambda} - \boldsymbol{\Upsilon}^*\right)\right)\text{trace}\left(\boldsymbol{\Omega} - \boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right) + \text{trace}\left(\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\right) - 2\boldsymbol{\psi}^{(1)}_{1,pk+2} \\ & \times (\Delta, 1)\,\text{trace}\left(\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\right) + \boldsymbol{\psi}^{(2)}_{2,pk+2}(\Delta, 1)\,\text{trace}\left(\boldsymbol{W}\boldsymbol{\Upsilon}^*\right)\text{trace}\left(\boldsymbol{\Xi}_4^{-1}\right) + \boldsymbol{\psi}^{(1)}_{2,pk+4} \\ & \times (\Delta, 1)\,\text{trace}\left(\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\right) + \boldsymbol{\psi}^{(2)}_{0,pk}(0, 1)\,\boldsymbol{\psi}^{(1)}_{2,pk}(\Delta, 1)\,\text{trace}\left(\boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right)\text{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Xi}_3^{-1} - \boldsymbol{\Upsilon}^*\right)\right). \quad \square \end{aligned}$$

Proof. Since h is a real-valued function, we get

$$R\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = E\left\{\text{trace}\left[\boldsymbol{\zeta}'\boldsymbol{W}\boldsymbol{\zeta}\right]\right\} + 2E\left\{h\left(\|\boldsymbol{\xi}\|^2_{\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2}\right)\text{trace}\left[\boldsymbol{\xi}'\boldsymbol{W}\boldsymbol{\zeta}\right]\right\} + E\left\{h^2\left(\|\boldsymbol{\xi}\|^2_{\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2}\right)\text{trace}\left[\boldsymbol{\xi}'\boldsymbol{W}\boldsymbol{\xi}\right]\right\}.$$

Further, by using Proposition 2.1, we have

$$E\left\{\text{trace}\left[\boldsymbol{\zeta}'\boldsymbol{W}\boldsymbol{\zeta}\right]\right\} = \boldsymbol{\psi}^{(2)}_{0,pk}(0, 1)\,\text{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Lambda} - \boldsymbol{\Upsilon}^*\right)\right)\text{trace}\left(\boldsymbol{\Omega} - \boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right) + \text{trace}\left(\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\right).$$

Also, using Theorem 2.2 with $\boldsymbol{\Upsilon}_{11} = \boldsymbol{\Upsilon}^*$, $\boldsymbol{\Lambda}_{11} = \boldsymbol{\Omega}$, $\boldsymbol{\Upsilon}_{12} = \boldsymbol{\Lambda} - \boldsymbol{\Upsilon}^*$, and $\boldsymbol{\Lambda}_{12} = \boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}$, we get

$$E\left\{h\left(\|\boldsymbol{\xi}\|^2_{\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2}\right)\text{trace}\left[\boldsymbol{\xi}'\boldsymbol{W}\boldsymbol{\zeta}\right]\right\} = -\boldsymbol{\psi}^{(1)}_{1,pk+2}(\Delta, 1)\,\text{trace}\left(\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\right),$$

and applying Theorem 2.3, we get

$$\begin{aligned} E\left\{h^2\left(\|\boldsymbol{\xi}\|^2_{\boldsymbol{\Xi}_1,\boldsymbol{\Xi}_2}\right)\text{trace}\left[\boldsymbol{\xi}'\boldsymbol{W}\boldsymbol{\xi}\right]\right\} = {} & \boldsymbol{\psi}^{(2)}_{2,pk+2}(\Delta, 1)\,\text{trace}\left(\boldsymbol{W}\boldsymbol{\Upsilon}^*\right)\text{trace}\left(\boldsymbol{\Xi}_4^{-1}\right) + \boldsymbol{\psi}^{(1)}_{2,pk+4}(\Delta, 1)\,\text{trace}\left(\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\right) \\ & + \boldsymbol{\psi}^{(2)}_{0,pk}(0, 1)\,\boldsymbol{\psi}^{(1)}_{2,pk}(\Delta, 1)\,\text{trace}\left(\boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right)\text{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Xi}_3^{-1} - \boldsymbol{\Upsilon}^*\right)\right). \end{aligned}$$

This completes the proof. □

3.1. Application of Theorems 3.1–3.2 to some distributions

Example 3.1. As mentioned in Section 1.2 the ULSE and RLSE can be viewed as special cases of the estimator in (1.5) for which $h(x) = 1$ and $h(x) = 0$ respectively. Thus, by using Theorem 3.1, we get $B(\boldsymbol{X}, \boldsymbol{\theta}) = B\left(\widehat{\boldsymbol{\theta}}(1), \boldsymbol{\theta}\right) = \mathbf{0}$, $B\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}\right) = B\left(\widehat{\boldsymbol{\theta}}(0), \boldsymbol{\theta}\right) = -\boldsymbol{\delta}$. Further, by taking $h(x) = 1$ and $h(x) = 0$ respectively, Theorem 3.2 gives, as expected, the risk of $\widetilde{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ respectively. We have, $R\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = R\left(\widehat{\boldsymbol{\theta}}(0), \boldsymbol{\theta}; \boldsymbol{W}\right)$ and $R\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = R\left(\widehat{\boldsymbol{\theta}}(1), \boldsymbol{\theta}; \boldsymbol{W}\right)$, and with some algebraic computations, we get

$$R\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = \boldsymbol{\psi}^{(2)}_{0,pk}(0,\ 1)\ \mathrm{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Lambda} - \boldsymbol{\Upsilon}^{*}\right)\right) \mathrm{trace}\left(\boldsymbol{\Omega} - \boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right) + \mathrm{trace}\left(\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\right),$$
$$R\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{W}\right) = \boldsymbol{\psi}^{(2)}_{0,pk}(0,\ 1)\ \mathrm{trace}\left(\boldsymbol{W}\boldsymbol{\Lambda}\right) \mathrm{trace}\left(\boldsymbol{\Omega}\right).$$

Example 3.2. Let $\varphi = \mathrm{trace}\left(\boldsymbol{\Xi}_2\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)' \boldsymbol{\Xi}_1\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)\right)$ and set $h(x) = (1 - a/x), x > 0$ for some $a > 0$. With this function, the estimator in (1.5) is $\widehat{\boldsymbol{\theta}}^{s} = \widetilde{\boldsymbol{\theta}} + (1 - a/\varphi)\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)$ which is known as the shrinkage estimator. By taking the above h and applying Theorems 3.1–3.2 with appropriate $\boldsymbol{\omega}(.)$, one gets the bias and risk functions for the shrinkage estimator. In particular, for the Gaussian sample cases, the bias and risk functions are given by

$$B\left(\widehat{\boldsymbol{\theta}}^{S}, \boldsymbol{\theta}\right) = -\boldsymbol{\delta}\, a\, E\left\{\chi^{-2}_{pk+2}(\Delta)\right\},$$
$$\begin{aligned} R\left(\widehat{\boldsymbol{\theta}}^{s}, \boldsymbol{\theta}; \boldsymbol{W}\right) &= R(\boldsymbol{X}, \boldsymbol{\theta}; \boldsymbol{W}) + a\,\mathrm{trace}\left(\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^{*}\right)(pk+2)\,E\left(\chi^{-4}_{pk+4}(\Delta)\right) \\ &\quad - a\,\mathrm{trace}\left(\boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right)\mathrm{trace}\left(\boldsymbol{W}\boldsymbol{\Lambda}^{*}\right)\left\{2E\left(\chi^{-2}_{pk+2}(\Delta)\right) - (pk-2)E\left(\chi^{-4}_{pk+2}(\Delta)\right)\right\} \\ &\quad + \left[1 - 2aE\left(\chi^{-2}_{pk}(\Delta)\right) + a^2E\left(\chi^{-4}_{pk}(\Delta)\right)\right]\mathrm{trace}\left(\boldsymbol{\Omega}\boldsymbol{L}_2\boldsymbol{P}\right)\mathrm{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Xi}_3^{-1} - \boldsymbol{\Upsilon}^{*}\right)\right). \end{aligned}$$

Further, for the Pearson type VII sample cases, let $\Delta^{*} = q_0/(q_0 + \Delta)$ and let $\mathscr{B}^{-}(s,\ q_1)$ denote a negative binomial random variable with parameters $s = m - qk/2 > 1$, $0 < q_1 < 1$ (where q_1 is the probability of success). We have

$$B\left(\widehat{\boldsymbol{\theta}}^{s+}, \boldsymbol{\theta}\right) = -\boldsymbol{\delta} + \left[1 - \frac{2a\,s}{q_0}E\left(\frac{1}{pk + 2\,\mathscr{B}^{-}(s+1,\ \Delta^{*})}\right)\right]\boldsymbol{\delta}.$$

Similarly, one applies Theorem 3.2 in order to derive the risk function $R\left(\widehat{\boldsymbol{\theta}}^{s}, \boldsymbol{\theta}; \boldsymbol{W}\right)$. To this end, some algebraic computations give $\boldsymbol{\psi}^{(2)}_{0,pk}(0,\ 1) = q_0/(2(s-1))$, and the functions $\boldsymbol{\psi}^{(1)}_{1,pk+2}(\Delta,\ 1)$, $\boldsymbol{\psi}^{(1)}_{2,pk+4}(\Delta,\ 1)$, $\boldsymbol{\psi}^{(1)}_{2,pk}(\Delta,\ 1)$, $\boldsymbol{\psi}^{(2)}_{2,pk+2}(\Delta,\ 1)$ are replaced by

$$\boldsymbol{\psi}^{(1)}_{1,pk+2}(\Delta,\ 1) = 1 - \frac{2as}{q_0}E\left(\frac{1}{pk + 2\,\mathscr{B}^{-}(s+1,\ \Delta^{*})}\right),$$
$$\begin{aligned} \boldsymbol{\psi}^{(1)}_{2,pk+4}(\Delta,\ 1) &= 1 - \frac{4a\,s}{q_0}E\left(\frac{1}{pk+2+2\mathscr{B}^{-}(s+1,\ \Delta^{*})}\right) - \frac{2s(s+1)a^2}{q_0^2}E\left(\frac{1}{pk+2+2\mathscr{B}^{-}(s+2,\ \Delta^{*})}\right) \\ &\quad + \frac{2s(s+1)a^2}{q_0^2}E\left(\frac{1}{pk+2\mathscr{B}^{-}(s+2,\ \Delta^{*})}\right), \end{aligned}$$
$$\begin{aligned} \boldsymbol{\psi}^{(1)}_{2,pk}(\Delta,\ 1) &= 1 - \frac{4a\,s}{q_0}E\left(\frac{1}{pk-2+2\mathscr{B}^{-}(s+1,\ \Delta^{*})}\right) - \frac{2s(s+1)a^2}{q_0^2}E\left(\frac{1}{pk-2+2\mathscr{B}^{-}(s+2,\ \Delta^{*})}\right) \\ &\quad + \frac{2s(s+1)a^2}{q_0^2}E\left(\frac{1}{pk-2+2\mathscr{B}^{-}(s+2,\ \Delta^{*})}\right), \end{aligned}$$
$$\begin{aligned} \boldsymbol{\psi}^{(2)}_{2,pk+2}(\Delta,\ 1) &= \frac{q_0}{2(s-1)} - a\,E\left(\frac{1}{pk+2\,\mathscr{B}^{-}(s,\ \Delta^{*})}\right) - \frac{s\,a^2}{q_0}E\left(\frac{1}{pk+2\,\mathscr{B}^{-}(s+1,\ \Delta^{*})}\right) \\ &\quad + \frac{s\,a^2}{q_0}E\left(\frac{1}{pk-2+2\,\mathscr{B}^{-}(s+1,\ \Delta^{*})}\right). \end{aligned}$$

In order to save the space, we do not write the risk and bias functions for the Student't sample cases. However, these functions follow from the above functions by taking $s = q_0/2$ with $q_0 > 2$.

Example 3.3. Let $h(x) = \max(0, (1 - a/x))$, $x > 0$ for some $a > 0$. Using the relation in (1.5), we get the well known positive-part shrinkage estimator $\widehat{\boldsymbol{\theta}}^{s+} = \widetilde{\boldsymbol{\theta}} + \max(0, (1 - a/\varphi))\left(\boldsymbol{X} - \widetilde{\boldsymbol{\theta}}\right)$. Let $H_\nu(x\,;\varDelta) = P\{\boldsymbol{\chi}_\nu^2(\varDelta) \le x\}$, $x \in \mathbb{R}^+$ where $\boldsymbol{\chi}_\nu^2(\varDelta)$ denotes a chi-square random variate with ν degrees of freedom and the noncentrality parameter $\varDelta$. By applying Theorems 3.1–3.2, we get the bias and risk of $\widehat{\boldsymbol{\theta}}^{s+}$ by replacing in (2.1) the function $h(x)$ with $\max(0, (1 - a/x))$, $x > 0$, for some $a > 0$. In particular, in the case of a Gaussian sample, the bias and risk functions are given by

$$
\begin{aligned}
B\left(\widehat{\boldsymbol{\theta}}^{s+}, \boldsymbol{\theta}\right) &= -\boldsymbol{\delta}\left[H_{pk+2}(a;\varDelta) + aE\left\{\boldsymbol{\chi}_{pk+2}^{-2}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk+2}(\varDelta)>a\right\}}\right\}\right],\\
R\left(\widehat{\boldsymbol{\theta}}^{s+}, \boldsymbol{\theta}; \boldsymbol{W}\right) &= R\left(\widehat{\boldsymbol{\theta}}^{s}, \boldsymbol{\theta}; \boldsymbol{W}\right) + a\,\mathrm{trace}\,(\boldsymbol{\Omega L}_2\boldsymbol{P})\,\mathrm{trace}\left(\boldsymbol{W\Lambda}^*\right)\left[2E\left\{\boldsymbol{\chi}_{pk+2}^{-2}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk+2}(\varDelta)\le a\right\}}\right\}\right.\\
&\quad\left. - aE\left\{\boldsymbol{\chi}_{pk+2}^{-4}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk+2}(\varDelta)\le a\right\}}\right\}\right] - \mathrm{trace}\,(\boldsymbol{\Omega L}_2\boldsymbol{P})\,\mathrm{trace}\,(W\boldsymbol{\Lambda})\,H_{pk+2}(a;\varDelta) + \mathrm{trace}\left(\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*\right)\\
&\quad\times\left\{2H_{pk+2}(a;\varDelta) - H_{pk+4}(a;\varDelta)\right\} - \left[H_{pk}(a;\varDelta) - 2aE\left(\boldsymbol{\chi}_{pk}^{-2}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk}(\varDelta)\le a\right\}}\right)\right]\mathrm{trace}\,(\boldsymbol{\Omega L}_2\boldsymbol{P})\\
&\quad\times\mathrm{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Xi}_3^{-1} - \boldsymbol{\Upsilon}^*\right)\right) - a^2E\left(\boldsymbol{\chi}_{pk}^{-4}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk}(\varDelta)\le a\right\}}\right)\mathrm{trace}\,(\boldsymbol{\Omega L}_2\boldsymbol{P})\\
&\quad\times\mathrm{trace}\left(\boldsymbol{W}\left(\boldsymbol{\Xi}_3^{-1} - \boldsymbol{\Upsilon}^*\right)\right) - a\,\mathrm{trace}\left(\boldsymbol{\delta}^{*\prime}\mathbf{W}\boldsymbol{\delta}^*\right)\left[2E\left\{\boldsymbol{\chi}_{pk+2}^{-2}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk+2}(\varDelta)\le a\right\}}\right\}\right.\\
&\quad\left. - 2\,E\left\{\boldsymbol{\chi}_{pk+4}^{-2}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk+4}(\varDelta)\le a\right\}}\right\} + a\,E\left\{\boldsymbol{\chi}_{pk+4}^{-4}(\varDelta)\mathbb{I}_{\left\{\chi^2_{pk+4}(\varDelta)\le a\right\}}\right\}\right].
\end{aligned}
$$

Further, for the Pearson type VII sample cases, the application of Theorem 3.1 gives

$$
\begin{aligned}
B\left(\widehat{\boldsymbol{\theta}}^{s+}, \boldsymbol{\theta}\right) &= -\frac{a\,s}{q_0}\,E\left(\frac{2}{pk + 2\mathscr{B}^-(s+1,\ \varDelta^*)}\right)\boldsymbol{\delta}\\
&\quad - E\left\{E\left[H_{pk+2+2\mathscr{B}^-(s,\ \varDelta^*)}\left(\mathscr{G}am\left(s + \mathscr{B}^-(s,\ \varDelta^*),\ \frac{2a\varDelta^*}{q_0}\right);\ 0\right)\Big|\mathscr{B}^-(s,\ \varDelta^*)\right]\right\}\boldsymbol{\delta}\\
&\quad + \frac{a\,s}{q_0}E\left\{E\left[H_{pk+2\mathscr{B}^-(s+1,\ \varDelta^*)}\left(\mathscr{G}am\left(s + \mathscr{B}^-(s+1,\ \varDelta^*),\ \frac{2a\varDelta^*}{q_0}\right);\ 0\right)\Big|\mathscr{B}^-(s+1,\ \varDelta^*)\right]\right\}\boldsymbol{\delta}.
\end{aligned}
$$

Also, in the similar way as in Example 3.2, the risk function $R\left(\widehat{\boldsymbol{\theta}}^{s}, \boldsymbol{\theta}; \boldsymbol{W}\right)$ is obtained by applying Theorem 3.2 with the functions $\boldsymbol{\psi}_{1,pk+2}^{(1)}(\varDelta)$, $\boldsymbol{\psi}_{2,pk+4}^{(1)}(\varDelta)$, $\boldsymbol{\psi}_{2,pk}^{(1)}(\varDelta,\ 1)$, $\boldsymbol{\psi}_{2,pk+2}^{(2)}(\varDelta)$ replaced by

$$
\begin{aligned}
\boldsymbol{\psi}_{1,pk+2}^{(1)}(\varDelta,\ 1) &= 1 - E\left\{E\left[H_{pk+2+2\mathscr{B}^-(s,\ \varDelta^*)}\left(\mathscr{G}am\left(s + \mathscr{B}^-(s,\ \varDelta^*),\ \frac{2a\varDelta^*}{q_0}\right);\ 0\right)\Big|\mathscr{B}^-(s,\ \varDelta^*)\right]\right\}\\
&\quad + \frac{a\,s}{q_0}E\left\{E\left[H_{pk+2\mathscr{B}^-(s+1,\ \varDelta^*)}\left(\mathscr{G}am\left(s + \mathscr{B}^-(s+1,\ \varDelta^*),\ \frac{2a\varDelta^*}{q_0}\right);\ 0\right)\Big|\mathscr{B}^-(s+1,\ \varDelta^*)\right]\right\}\\
&\quad - \frac{a\,s}{q_0}\,E\left(\frac{2}{pk + 2\mathscr{B}^-(s+1,\ \varDelta^*)}\right),\\
\boldsymbol{\psi}_{2,pk+4}^{(1)}(\varDelta) &= 1 - E\left\{E\left[H_{pk+4+2\mathscr{B}^-(s,\ \varDelta^*)}\left(\mathscr{G}am\left(s + \mathscr{B}^-(s,\ \varDelta^*),\ \frac{2a\varDelta^*}{q_0}\right);\ 0\right)\Big|\mathscr{B}^-(s,\ \varDelta^*)\right]\right\}\\
&\quad + \frac{a\,s}{q_0}E\left\{E\left[H_{pk+2+2\mathscr{B}^-(s+1,\ \varDelta^*)}\left(\mathscr{G}am\left(s + 1 + \mathscr{B}^-(s+1,\ \varDelta^*),\ \frac{2a\varDelta^*}{q_0}\right);\ 0\right)\Big|\mathscr{B}^-(s+1,\ \varDelta^*)\right]\right\}\\
&\quad - \frac{4a^2\,s(s+1)}{q_0^2}E\left\{E\left[H_{pk+2\mathscr{B}^-(s+2,\ \varDelta^*)}\left(\mathscr{G}am\left(s + 2 + \mathscr{B}^-(s+2,\ \varDelta^*),\ \frac{2a\varDelta^*}{q_0}\right);\ 0\right)\Big|\mathscr{B}^-(s+2,\ \varDelta^*)\right]\right\}\\
&\quad - \frac{4a}{q_0}\,E\left(\frac{s}{pk + 2 + 2\mathscr{B}^-(s+1,\ \varDelta^*)}\right) - \frac{2\,a^2}{q_0^2}\,E\left(\frac{s(s+1)}{pk + 2 + 2\mathscr{B}^-(s+2,\ \varDelta^*)}\right)\\
&\quad + \frac{2\,a^2}{q_0^2}\,E\left(\frac{s(s+1)}{pk + 2\mathscr{B}^-(s+2,\ \varDelta^*)}\right),
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\psi}^{(1)}_{2,pk}(\Delta) &= 1 - E\left\{E\left[H_{pk+2\mathscr{B}^-(s,\,\Delta^*)}\left(\mathscr{G}am\left(s+\mathscr{B}^-(s,\,\Delta^*),\,\frac{2a\Delta^*}{q_0}\right);\,0\right)\Big|\mathscr{B}^-(s,\,\Delta^*)\right]\right\}\\
&+\frac{a\,s}{q_0}E\left\{E\left[H_{pk-2+2\mathscr{B}^-(s+1,\,\Delta^*)}\left(\mathscr{G}am\left(s+1+\mathscr{B}^-(s+1,\,\Delta^*),\,\frac{2a\Delta^*}{q_0}\right);\,0\right)\Big|\mathscr{B}^-(s+1,\,\Delta^*)\right]\right\}\\
&-\frac{4a^2\,s(s+1)}{q_0^2}E\left\{E\left[H_{pk-4+2\mathscr{B}^-(s+2,\,\Delta^*)}\left(\mathscr{G}am\left(s+2+\mathscr{B}^-(s+2,\,\Delta^*),\,\frac{2a\Delta^*}{q_0}\right);\,0\right)\Big|\mathscr{B}^-(s+2,\,\Delta^*)\right]\right\}\\
&-\frac{4\,a}{q_0}E\left(\frac{s}{pk-2+2\mathscr{B}^-(s+1,\,\Delta^*)}\right)-\frac{2\,a^2}{q_0^2}E\left(\frac{s(s+1)}{pk-2+2\mathscr{B}^-(s+2,\,\Delta^*)}\right)\\
&+\frac{2\,a^2}{q_0^2}E\left(\frac{s(s+1)}{pk-4+2\mathscr{B}^-(s+2,\,\Delta^*)}\right),\\
\boldsymbol{\psi}^{(2)}_{2,pk+2}(\Delta) &= \frac{q_0}{2(s-1)}+a\,E\left\{E\left[H_{pk+2\mathscr{B}^-(s,\,\Delta^*)}\left(\mathscr{G}am\left(s+\mathscr{B}^-(s,\,\Delta^*),\,\frac{2a\Delta^*}{q_0}\right);\,0\right)\Big|\mathscr{B}^-(s,\,\Delta^*)\right]\right\}\\
&-\frac{q_0}{2(s-1)}E\left\{E\left[H_{pk+2+2\mathscr{B}^-(s-1,\,\Delta^*)}\left(\mathscr{G}am\left(s-1+\mathscr{B}^-(s-1,\,\Delta^*),\,\frac{2a\Delta^*}{q_0}\right);\,0\right)\Big|\mathscr{B}^-(s-1,\,\Delta^*)\right]\right\}\\
&-\frac{a^2\,s}{2q_0}E\left\{E\left[H_{pk-2+2\mathscr{B}^-(s+1,\,\Delta^*)}\left(\mathscr{G}am\left(s+1+\mathscr{B}^-(s+1,\,\Delta^*),\,\frac{2a\Delta^*}{q_0}\right);\,0\right)\Big|\mathscr{B}^-(s+1,\,\Delta^*)\right]\right\}\\
&-a\,E\left(\frac{2}{pk+2\mathscr{B}^-(s,\,\Delta^*)}\right)-\frac{a^2}{q_0}E\left(\frac{s}{pk+2\mathscr{B}^-(s+1,\,\Delta^*)}\right)+\frac{a^2}{q_0}E\left(\frac{s}{pk-2+2\mathscr{B}^-(s+1,\,\Delta^*)}\right).
\end{aligned}
$$

Once again, in order to save the space, the risk and bias functions for the Student't sample cases are omitted. Indeed as justified in Example 3.2, the results from the above functions by taking $s = q_0/2$ with $q_0 > 2$.

Following the results given in Examples 3.2–3.3, note that, in general, the bias and the risk functions of the estimators $\widehat{\boldsymbol{\theta}}^S$ and $\widehat{\boldsymbol{\theta}}^{S+}$ do not have a closed form. However, by using similar techniques as used in the Gaussian sample case, (see for example the technique used for the Proof of Corollary 2.1 in [16]), one can establish, for an appropriate choice of $\boldsymbol{W}$, the dominance of the shrinkage estimator over the unrestricted estimator. In particular, for $\boldsymbol{W} = \boldsymbol{\Lambda}^{-1}$, one can verify that $R\left(\widehat{\boldsymbol{\theta}}^S, \boldsymbol{\theta}; \boldsymbol{\Lambda}^{-1}\right) < R\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}; \boldsymbol{\Lambda}^{-1}\right)$, for all $\Delta \geqslant 0$. Further, one can establish that $R\left(\widehat{\boldsymbol{\theta}}^{S+}, \boldsymbol{\theta}; \boldsymbol{W}\right) \leqslant R\left(\widehat{\boldsymbol{\theta}}^S, \boldsymbol{\theta}; \boldsymbol{W}\right)$, for all $\Delta \geqslant 0$.

4. Conclusion

In this paper, we provided two general formulas for the bias and risk functions of a class of shrinkage-type estimators for the mean of matrix-valued, elliptically contoured random variate. To this end, we extended some recent identities which are only applicable in the Gaussian sample cases in which the shrinking random part is a single Kronecker-product. Also, the established identities are applicable to the case in which the variance–covariance matrix of the shrinking random part is the sum of two Kronecker-products. Another interesting feature of this paper is that the matrices involved in the two Kronecker-products do not need to be invertible.

Acknowledgments

The author would like to acknowledge the financial support received from Natural Sciences and Engineering Research Council of Canada. Further, the author would like to thank Dr. A. Hussein and Dr. S. Froda as well as the anonymous referees for useful comments and suggestions.

Appendix. Derivation of technical results

In this section, we present some technical details that are used in establishing the main result. We start by a proposition which gives the RMLE. Let $\boldsymbol{W} \sim \mathscr{N}_n(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$ stand for an n-column random vector normally distributed with mean $\boldsymbol{\mu}$ and covariance–variance matrix $\boldsymbol{\Sigma}$.

Proposition A.1. *If the constraint in (1.2) holds, then the RMLE is given by*

$$\widetilde{\boldsymbol{\theta}} = (\boldsymbol{X} - \boldsymbol{J}\boldsymbol{L}_1\boldsymbol{X} + \boldsymbol{J}\boldsymbol{d}_1)(\boldsymbol{I}_k - \boldsymbol{L}_2\boldsymbol{P}) + \boldsymbol{d}_2\boldsymbol{P}. \quad \square \tag{A.1}$$

Proof. The restricted estimator of $\boldsymbol{\theta}$ is the solution of the minimization problem $\min_{\boldsymbol{\theta}}\left\{\text{trace}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{\theta})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\theta})\right]\right\}$ subjected to $\boldsymbol{L}_1\boldsymbol{\theta}=\boldsymbol{d}_1,\ \boldsymbol{\theta}\boldsymbol{L}_2=\boldsymbol{d}_2$. To this end, let $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ be $p\times k$ and $q\times m$-Lagrangian matrices respectively, and thus, let $\mathfrak{L}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\theta})$ denote a Lagrangian function. We have

$$\mathfrak{L}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\theta})=\text{trace}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{\theta})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\theta})\right]+\text{trace}\left[(\boldsymbol{L}_1\boldsymbol{\theta}-\boldsymbol{d}_1)\boldsymbol{\lambda}_1'\right]+\text{trace}\left[(\boldsymbol{\theta}\boldsymbol{L}_2-\boldsymbol{d}_2)\boldsymbol{\lambda}_2'\right].$$

Then, by applying Theorem A.95 in [19, p. 522], one differentiates $\mathfrak{L}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. Hence, equating to zero the obtained derivatives, one gets a system of equations in $\boldsymbol{\theta}$, $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, and then, solving the obtained system of equations, one gets the result stated in the proposition. $\square$

Below, we give two lemmas which correspond to Theorems 1 and 2 of [10] for the case of a random vector which follows an elliptically contoured distribution. The two lemmas are useful in deriving Theorems 2.1 and 2.3 given in Section 2.

Lemma A.1. *Let $\boldsymbol{U}\sim\mathscr{E}_n(\boldsymbol{\mu},\boldsymbol{\sigma}^2\boldsymbol{I}_n;g)$, $\boldsymbol{\sigma}>0$. Then, for all h, Borel measurable and real-valued integrable functions, $E\left[h(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{U}\right]=\boldsymbol{\psi}^{(1)}_{1,n+2}\left(\boldsymbol{\mu}'\boldsymbol{\mu},\ \boldsymbol{\sigma}^2\right)\boldsymbol{\mu}$, where $\boldsymbol{\psi}^{(1)}_{i,n+2}(.,\ .)$ is defined in (2.1).* $\square$

Proof. From the pdf's representation of an elliptically contoured variate (see for example, [6]), we have

$$E\left[h(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{U}\right]=\int_0^\infty E_z\left[h(\boldsymbol{W}'\boldsymbol{W})\boldsymbol{W}\right]\omega(z)dz, \tag{A.2}$$

where $\boldsymbol{W}$ denotes an n-column random vector normally distributed with mean $\boldsymbol{\mu}$ and covariance–variance matrix $z^{-1}\boldsymbol{\sigma}^2\boldsymbol{I}_n$, $0<z<\infty$, $\boldsymbol{\sigma}>0$. Further,

$$E_z\left[h(\boldsymbol{W}'\boldsymbol{W})\boldsymbol{W}\right]=(\boldsymbol{\sigma}/\sqrt{z})\times E_z\left[h(z\,\boldsymbol{\sigma}^{-2}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{\sigma}^2z^{-1})\frac{\sqrt{z}}{\boldsymbol{\sigma}}\boldsymbol{W}\right], \tag{A.3}$$

and then, using Theorem 1 in [10], we have

$$E_z\left[h(z\,\boldsymbol{\sigma}^{-2}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{\sigma}^2z^{-1})\frac{\sqrt{z}}{\boldsymbol{\sigma}}\boldsymbol{W}\right]=E\left[h(z^{-1}\boldsymbol{\sigma}^2\chi^2_{n+2}(z\boldsymbol{\sigma}^{-2}\boldsymbol{\mu}'\boldsymbol{\mu}))\right]\frac{\sqrt{z}}{\boldsymbol{\sigma}}\boldsymbol{\mu}, \tag{A.4}$$

and then, combining relations (A.2)–(A.4), we get

$$E\left[h(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{U}\right]=\left(\int_0^\infty E\left[h\left(\mathscr{G}am\left(\frac{n+2}{2},\ \frac{2\boldsymbol{\sigma}^2}{z};\ \frac{z\boldsymbol{\mu}'\boldsymbol{\mu}}{\boldsymbol{\sigma}^2}\right)\right)\right]\omega(z)dz\right)\boldsymbol{\mu};$$

this completes the proof. $\square$

Lemma A.2. *Let $\boldsymbol{U}\sim\mathscr{E}_n(\boldsymbol{\mu},\boldsymbol{\sigma}^2\boldsymbol{I}_n;g)$, $\boldsymbol{\sigma}>$, and let $\boldsymbol{A}$ be an $n\times n$-nonnegative definite matrix. Then, for all h, Borel measurable and real-valued integrable functions,*

$$E\left[h(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{U}'\boldsymbol{A}\boldsymbol{U}\right]=\mathit{trace}(\boldsymbol{A})\boldsymbol{\psi}^{(2)}_{1,n+2}\left(\boldsymbol{\mu}'\boldsymbol{\mu},\ \boldsymbol{\sigma}^2\right)+\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\psi}^{(1)}_{1,n+4}\left(\boldsymbol{\mu}'\boldsymbol{\mu},\ \boldsymbol{\sigma}^2\right),$$

where $\boldsymbol{\psi}^{(1)}_{i,n+2}(.,\ .)$ and $\boldsymbol{\psi}^{(2)}_{i,n+2}(.,\ .)$ are defined in (2.1). $\square$

Proof. As in the Proof of Lemma A.1, we have

$$E\left[h(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{U}'\boldsymbol{A}\boldsymbol{U}\right]=\int_0^\infty E_z\left[h(\boldsymbol{W}'\boldsymbol{W})\boldsymbol{U}'\boldsymbol{A}\boldsymbol{U}\right]\omega(z)dz, \tag{A.5}$$

where $\boldsymbol{W}$ denotes an n-column random vector normally distributed with mean $\boldsymbol{\mu}$ and covariance–variance matrix $z^{-1}\boldsymbol{\sigma}^2\boldsymbol{I}_n$, $0<z<\infty$. Further,

$$E_z\left[h(\boldsymbol{W}'\boldsymbol{W})\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\right]=\left(\frac{\boldsymbol{\sigma}^2}{z}\right)\times E_z\left[h\left(\frac{z}{\boldsymbol{\sigma}^2}\boldsymbol{W}'\boldsymbol{W}\frac{\boldsymbol{\sigma}^2}{z}\right)\left(\frac{\sqrt{z}}{\boldsymbol{\sigma}}\boldsymbol{W}\right)'\boldsymbol{A}\left(\frac{\sqrt{z}}{\boldsymbol{\sigma}}\boldsymbol{W}\right)\right], \tag{A.6}$$

and then, using Theorem 2 in [10], we have

$$\begin{aligned}E_z\left[h\left(\frac{z}{\boldsymbol{\sigma}^2}\boldsymbol{W}'\boldsymbol{W}\frac{\boldsymbol{\sigma}^2}{z}\right)\left(\frac{\sqrt{z}}{\boldsymbol{\sigma}}\boldsymbol{W}\right)'\boldsymbol{A}\left(\frac{\sqrt{z}}{\boldsymbol{\sigma}}\boldsymbol{W}\right)\right]&=\text{trace}(\boldsymbol{A})E\left[h\left(\frac{\boldsymbol{\sigma}^2}{z}\chi^2_{n+2}\left(\frac{z}{\boldsymbol{\sigma}^2}\boldsymbol{\mu}'\boldsymbol{\mu}\right)\right)\right]\\&\quad+\frac{z}{\boldsymbol{\sigma}^2}\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}E\left[h\left(\frac{\boldsymbol{\sigma}^2}{z}\chi^2_{n+4}\left(\frac{z}{\boldsymbol{\sigma}^2}\boldsymbol{\mu}'\boldsymbol{\mu}\right)\right)\right].\end{aligned} \tag{A.7}$$

Therefore, using relations (A.5)–(A.7) along with some algebraic computations, we get

$$E\left[h(\boldsymbol{U}'\boldsymbol{U})\boldsymbol{U}'\boldsymbol{A}\boldsymbol{U}\right] = \operatorname{trace}(\boldsymbol{A})\left(\sigma^2\int_0^\infty z^{-1}E\left[h\left(\mathscr{G}am\left(\frac{n+2}{2},\frac{2\sigma^2}{z};\frac{z\boldsymbol{\mu}'\boldsymbol{\mu}}{\sigma^2}\right)\right)\right]\omega(z)dz\right) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}\left(\int_0^\infty E\left[h\left(\mathscr{G}am\left(\frac{n+4}{2},\frac{2\sigma^2}{z};\frac{z\boldsymbol{\mu}'\boldsymbol{\mu}}{\sigma^2}\right)\right)\right]\omega(z)dz\right);$$

this completes the proof. □

Proof of Theorem 2.1. Since $\boldsymbol{\Xi}_3^{\frac{1}{2}}\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}$ and $\boldsymbol{\Xi}_4^{\frac{1}{2}}\boldsymbol{\Lambda}_2\boldsymbol{\Xi}_4^{\frac{1}{2}}$ are symmetric idempotent matrices, there exist orthogonal matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ such that

$$\boldsymbol{Q}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\boldsymbol{Q}_1' = \begin{pmatrix}\boldsymbol{I}_{p_1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix} \quad\text{and}\quad \boldsymbol{Q}_2\boldsymbol{\Xi}_4^{\frac{1}{2}}\boldsymbol{\Lambda}_2\boldsymbol{\Xi}_4^{\frac{1}{2}}\boldsymbol{Q}_2' = \begin{pmatrix}\boldsymbol{I}_{q_1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}, \tag{A.8}$$

with q_1 the rank of $\boldsymbol{\Lambda}_2$. Moreover, let $\boldsymbol{V} = \boldsymbol{Q}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\boldsymbol{X}\boldsymbol{\Xi}_4^{\frac{1}{2}}\boldsymbol{Q}_2'$; then,

$$\operatorname{Vec}(\boldsymbol{V}) = \left(\boldsymbol{Q}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\otimes\boldsymbol{Q}_2\boldsymbol{\Xi}_4^{\frac{1}{2}}\right)\operatorname{Vec}(\boldsymbol{X}) \text{ and hence,}$$

$$\operatorname{Vec}(\boldsymbol{V}) = \begin{pmatrix}\boldsymbol{V}_1\\ \boldsymbol{V}_2\end{pmatrix} \sim \mathscr{E}_{q\times k}\left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{0}\end{pmatrix},\ \boldsymbol{\Sigma}_v;\ g\right), \tag{A.9}$$

with

$$\boldsymbol{\mu}_1 = \left[\boldsymbol{I}_{p_1q},\ \boldsymbol{0}\right]E\left(\operatorname{Vec}(\boldsymbol{V})\right) = \left(\left[\boldsymbol{I}_{p_1},\ \boldsymbol{0}\right]\otimes\boldsymbol{I}_q\right)\left(\boldsymbol{Q}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\otimes\boldsymbol{Q}_2\boldsymbol{\Xi}_4^{\frac{1}{2}}\right)\operatorname{Vec}(\boldsymbol{M})$$

$$\boldsymbol{\Sigma}_v = \begin{pmatrix}\boldsymbol{I}_{p_1q} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix} + \begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{k-p_1}\end{pmatrix}\otimes\begin{pmatrix}\boldsymbol{I}_{q_1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}. \tag{A.10}$$

Then, using the relation in (A.9), we have

$$\operatorname{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{X}\right) = \operatorname{trace}\left(\boldsymbol{V}'\boldsymbol{Q}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\boldsymbol{Q}_1'\boldsymbol{V}\right) = \boldsymbol{V}_1'\boldsymbol{V}_1,$$

and then,

$$\operatorname{Vec}\left(E\left[h\left(\operatorname{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\boldsymbol{X}\right]\right) = \left(\boldsymbol{\Xi}_3^{-\frac{1}{2}}\boldsymbol{Q}_1'\otimes\boldsymbol{\Xi}_4^{-\frac{1}{2}}\boldsymbol{Q}_2'\right)\left(E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1'\right)\boldsymbol{V}_1\right],\ E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\boldsymbol{V}_2'\right]\right)'. \tag{A.11}$$

Combining (A.9) and (A.10), we get $E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\boldsymbol{V}_2\right] = \boldsymbol{0}$, and by using Lemma A.1, we have

$$E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\boldsymbol{V}_1\right] = \boldsymbol{\mu}_1\boldsymbol{\psi}^{(1)}_{1,p_1q+2}\left(\boldsymbol{\mu}_1'\boldsymbol{\mu}_1,\ 1\right), \tag{A.12}$$

where $\boldsymbol{\mu}_1$ is given by (A.10). Further, combining (A.11) and (A.12), we have

$$\operatorname{Vec}\left(E\left[h\left(\operatorname{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\boldsymbol{X}\right]\right) = \operatorname{Vec}\left(\boldsymbol{\psi}^{(1)}_{1,p_1q+2}\left(\boldsymbol{\mu}_1'\boldsymbol{\mu}_1,\ 1\right)\boldsymbol{M}\right).$$

Further, using (A.8)–(A.10) and following similar steps as in [15], we have

$$\boldsymbol{\mu}_1'\boldsymbol{\mu}_1 = \operatorname{trace}\left(\boldsymbol{M}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{M}\boldsymbol{\Xi}_4\right) = \operatorname{trace}\left(\boldsymbol{M}'\boldsymbol{\Xi}_3\boldsymbol{M}\boldsymbol{\Xi}_4\right), \tag{A.13}$$

which completes the proof. □

Remark A.1. For the particular case where the random matrix is Gaussian with $\boldsymbol{\Lambda}_2 = \boldsymbol{0}$, Theorem 2.1 leads to Theorem A.1 established in [15]. Thus, the above theorem extends also Theorem 1 in [10] for the Gaussian distribution case with $k = 1$, $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Xi} = 1$, $\boldsymbol{\Lambda}_1 = \boldsymbol{I}_q$.

Proof of Theorem 2.3. As established in the Proof of Theorem 2.1, we have

$$h\left(\operatorname{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right) = h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right),$$

where $\boldsymbol{V}_1$ is given by (A.9). Further, we have $\boldsymbol{V} = \boldsymbol{Q}_1\boldsymbol{\Xi}_3^{\frac{1}{2}}\boldsymbol{X}\boldsymbol{\Xi}_4^{\frac{1}{2}}\boldsymbol{Q}_2'$ where $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are the same as in (A.8). Also, as in [15], let

$$\left(\boldsymbol{Q}_1\boldsymbol{\Xi}_3^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Xi}_3^{-\frac{1}{2}}\boldsymbol{Q}_1'\otimes\boldsymbol{Q}_2\boldsymbol{\Xi}_4^{-1}\boldsymbol{Q}_2'\right) = \boldsymbol{G} = \begin{pmatrix}\boldsymbol{G}_{11} & \boldsymbol{G}_{12}\\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22}\end{pmatrix}. \tag{A.14}$$

After some algebra, we get

$$E\left[h\left(\operatorname{trace}\left(\boldsymbol{\Xi}_4\boldsymbol{X}'\boldsymbol{\Xi}_3\boldsymbol{\Upsilon}_1\boldsymbol{\Xi}_3\boldsymbol{X}\right)\right)\operatorname{trace}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right)\right]=E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\boldsymbol{V}_1'\boldsymbol{G}_{11}\boldsymbol{V}_1\right]+E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\right]E\left[\boldsymbol{V}_2'\boldsymbol{G}_{22}\boldsymbol{V}_2\right]. \tag{A.15}$$

Also, from (A.9), we have

$$E\left[\boldsymbol{V}_2'\boldsymbol{G}_{22}\boldsymbol{V}_2\right]=\boldsymbol{\psi}^{(2)}_{0,p_1q}(0,\ 1)\operatorname{trace}\left[\boldsymbol{G}_{22}\left(\left[\boldsymbol{0},\ \boldsymbol{I}_{k-p_1}\right]\otimes\boldsymbol{I}_q\right)\boldsymbol{\Sigma}_v\left(\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{I}_{k-p_1}\end{pmatrix}\otimes\boldsymbol{I}_q\right)\right], \tag{A.16}$$

where $\boldsymbol{\Sigma}_v$ is given in (A.10), and by combining (A.8) and (A.16) along with the fact that $\boldsymbol{G}_{22}=\left[\boldsymbol{0},\ \boldsymbol{I}_{kq-p_1q}\right]\boldsymbol{G}\left(\left[\boldsymbol{0},\ \boldsymbol{I}_{kq-p_1q}\right]\right)'$, we get

$$\begin{aligned}E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\right]E\left[\boldsymbol{V}_2'\boldsymbol{G}_{22}\boldsymbol{V}_2\right]&=E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\right]\boldsymbol{\psi}^{(2)}_{0,p_1q}(0,\ 1)\operatorname{trace}\left[\boldsymbol{G}\left(\begin{pmatrix}\boldsymbol{0}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{I}_{k-p_1}\end{pmatrix}\otimes\boldsymbol{I}_q\right)\right]\\&=\boldsymbol{\psi}^{(2)}_{0,p_1q}(0,\ 1)\boldsymbol{\psi}^{(1)}_{1,p_1q}\left(\operatorname{trace}\left(\boldsymbol{M}'\boldsymbol{\Xi}_3\boldsymbol{M}\boldsymbol{\Xi}_4\right),\ 1\right)\operatorname{trace}\left(\boldsymbol{A}\left(\boldsymbol{\Xi}_3^{-1}-\boldsymbol{\Upsilon}_1\right)\right)\operatorname{trace}\left(\boldsymbol{\Lambda}_2\right).\end{aligned} \tag{A.17}$$

Further, as in [15], using Lemma A.2, we have

$$E\left[h\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)\boldsymbol{V}_1'\boldsymbol{G}_{11}\boldsymbol{V}_1\right]=\boldsymbol{\psi}^{(2)}_{1,p_1q+2}\left(\boldsymbol{\mu}_1'\boldsymbol{\mu}_1,\ 1\right)\operatorname{trace}\left(\boldsymbol{G}_{11}\right)+\boldsymbol{\psi}^{(1)}_{1,p_1q+4}\left(\boldsymbol{\mu}_1'\boldsymbol{\mu}_1,\ 1\right)\left(\boldsymbol{\mu}_1'\boldsymbol{G}_{11}\boldsymbol{\mu}_1\right), \tag{A.18}$$

with $\boldsymbol{\mu}_1'\boldsymbol{\mu}_1$ given in (A.13), and

$$\boldsymbol{\mu}_1'\boldsymbol{G}_{11}\boldsymbol{\mu}_1=\operatorname{trace}\left(\boldsymbol{M}'\boldsymbol{A}\boldsymbol{M}\right),\qquad\operatorname{trace}\left(\boldsymbol{G}_{11}\right)=\operatorname{trace}\left(\boldsymbol{A}\boldsymbol{\Upsilon}_1\right)\operatorname{trace}\left(\boldsymbol{\Xi}_4^{-1}\right). \tag{A.19}$$

The proof is completed by combining the relations in (A.15)–(A.19). □

References

[1] B. Abdous, C. Genest, B. Rémillard, Dependence properties of meta-elliptical distributions, in: P. Duchesne, B. Rémillard (Eds.), Statistical Modeling and Analysis for Complex Data Problems, Kluwer, 2004.
[2] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, third ed., in: Wiley Series in Probability and Statistics, New York, 2003.
[3] T.W. Anderson, K.T. Fang, H. Hsu, Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions, The Canadian Journal of Statistics 14 (1986) 55–59.
[4] D.F. Andrews, A.M. Herzberg, Data, Springer, New York, 1985.
[5] N.H. Bingham, R. Kiesel, Semi-parametric modelling in finance: theoretical foundations, Quantitative Finance 1 (2001) 1–10.
[6] M.A. Chmielewski, Elliptically symmetric distributions: a review and bibliography, International Statistical Review 49 (1) (1981) 67–74.
[7] E. Furman, Z. Landsman, Tail variance premium with applications for elliptical portfolio of risks, Astin Bulletin 36 (2) (2006) 433–462.
[8] A.J. Izenman, Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning, Springer Science+Business Media, LLC, 2008.
[9] P. Jolicoeur, The degree of generality of robustness in Martes Americana, Growth 27 (1963) 1–27.
[10] G.G. Judge, M.E. Bock, The Statistical Implication of Pre-Test and Stein-Rule Estimators in Econometrics, North Holland, Amsterdam, 1978.
[11] Z.M. Landsman, E.A. Valdez, Tail conditional expectations for elliptical distributions, North American Actuarial Journal 7 (4) (2003) 55–123.
[12] S. Liu, Local influence in multivariate elliptical linear regression models, Linear Algebra and its Applications 354 (2002) 159–174.
[13] J.S. Liu, W. Cheung, H. Wong, Predictive inference for singular multivariate elliptically contoured distributions, Journal of Multivariate Analysis 100 (2009) 1440–1446.
[14] K. Mukherjee, On preliminary test and shrinkage estimation in linear models with long memory errors, Journal of Statistical Planning and Inference 69 (1998) 319–328.
[15] S. Nkurunziza, The risk of pre-test and shrinkage estimators, Statistics: A Journal of Theoretical and Applied Statistics 46 (3) (2012) 305–312.
[16] S. Nkurunziza, Shrinkage strategies in some multiple multi-factor dynamical systems, ESAIM: Probability and Statistics 16 (2012) 139–150.
[17] S. Nkurunziza, E.S. Ahmed, Estimation strategies for the regression coefficient parameter matrix in multivariate multiple regression, Statistica Neerlandica 65 (4) (2011) 387–406.
[18] S.B. Provost, Y.H. Cheong, On the distribution of linear combinations of the components of a Dirichlet random vector, The Canadian Journal of Statistics 28 (2) (2000) 417–425.
[19] C.R. Rao, H. Toutenburg, S. Shalabh, C. Heumann, Linear Models and Generalizations: Least Squares and Alternatives, 3rd ed., Springer, Berlin–Heidelberg, 2008.