Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 353–368

Estimation of Kurtosis Parameters of Multivariate Populations

Nighat Zahra[1], Muhammad Kashif Ali Shah[1] and S. Ejaz Ahmed[2]
[1] *Department of Statistics, GC University Lahore, Pakistan*
[2] *Department of Mathematics and Statistics, Brock University, St. Catharines, ON, Canada*

Received: 01 June 2020; Revised: 06 September 2020; Accepted: 25 September 2020

Abstract

This paper is based on the estimation of the kurtosis parameters of several multivariate populations under the assumption that all the kurtosis parameters are equal. The shrinkage and preliminary test estimators are suggested for the estimation of the vector of kurtosis parameters. Asymptotic properties of the suggested estimators are presented analytically and compared on the basis of their asymptotic distributional bias and asymptotic quadratic risk. Monte-Carlo simulations are performed in order to explain the analytical results numerically. A real data example is also given to demonstrate the application of the suggested estimators. From the results it can be observed that the Stein-type estimators perform better than all other estimators when the number of populations is greater than four and also when the assumption of homogeneity is suspicious.

Key words: Kurtosis; Shrinkage; Preliminary test; Asymptotic quadratic bias; Asymptotic quadratic risk; Stein-type estimators.

AMS Subject Classifications: 62E20, 62F10, 62F12

1. Introduction

Kurtosis and skewness are often considered as the shape parameters of a probability distribution. First introduced by Karl Pearson in 1905, kurtosis can be defined as a measurement with which to represent the size of a distribution's tails in contrast to a normal distribution. In a contemporary context, kurtosis is widely used in different areas of research, such as finance, space science, economics and signal processing. See Kim and White (2004), Liang *et al.* (2008), Nita and Gary (2010), Lai (2012), Araújo *et al.* (2012), and Echer and Bolzan (2016) for detailed examples.

Mardia (1970) defined the kurtosis parameter of a p-dimensional random variable $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as

$$\beta = \mathbb{E}\left[\left\{(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\}^2\right]. \tag{1}$$

Corresponding Author: Nighat Zahra
Email: nighatzahra@gcu.edu.pk

For a random sample of n observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ taken from a multivariate population, let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ be the estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively, then the estimator of kurtosis parameter proposed by Mardia (1970) can be defined as

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \left\{ (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}) \, \hat{\boldsymbol{\Sigma}}^{-1} \, (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T \right\}^2 , \tag{2}$$

where $\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i$ and $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}) (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T$.

Ahmed *et al.* (2012) worked on the improved estimation of kurtosis parameter β of a multivariate population using uncertain prior information (UPI). They proposed the linear shrinkage and the preliminary test estimators for β and developed the large sample theory for these estimators. Zahra *et al.* (2017 b) presented the improved estimation of kurtosis parameters for two multivariate populations and suggested that the shrinkage pretest estimator performs better when the null hypothesis $\beta_1 = \beta_2$ is uncertain; otherwise, the restricted estimator performs well. In this paper, we have extended their work on multivariate populations to q sample case using the UPI that kurtosis parameters of all populations are homogeneous. The UPI can be presented in the form of null hypothesis as

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = \beta_0, \tag{3}$$

where β_0 is unknown.

Let $\boldsymbol{X}_i^{(l)} = \left(\boldsymbol{X}_{1i}^{(l)}, \boldsymbol{X}_{2i}^{(l)}, \cdots, \boldsymbol{X}_{pi}^{(l)} \right)$ where $i = 1, 2, \cdots, n_l$ and $l = 1, 2, \cdots, q$ be a multivariate random sample of size n_l from a p-variate normal distribution with mean vector $\boldsymbol{\mu}_l$ and covariance matrix $\boldsymbol{\Sigma}_l$. We want to estimate the parameter vector of kurtosis coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_q)^T$ using their maximum likelihood estimator (MLE) $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_q)^T$ with sample sizes $(n_1, n_2, \cdots, n_q)$ under the UPI given in equation (3). Our multivariate sampled data $\boldsymbol{X}_i^{(l)}$ may have collected at q different times or spaces and there is natural tendency to combine the data to get efficient estimation results. One has to integrate both the sample information (SI) available in the form of $\boldsymbol{X}_i^{(l)}, l = 1, 2, \cdots, q$ and the UPI in such a way that estimators having optimal properties in terms of smallest risk, can be developed. Such integrated estimation strategies are based on the preliminary test (pretest) and Stein's shrinkage methodologies and are proved to be superior for data polling purposes. Many of these estimation strategies under different contexts have been discussed by Zahra *et al.* (2017 a), Shah *et al.* (2017), Lisawadi *et al.* (2019), Shah *et al.* (2020) and references therein.

The organization of the paper is as follows: Several estimation strategies are given in Section 2 that utilized either SI or the combination of both SI and UPI. Section 3 is concerned with the expressions of the asymptotic distributional quadratic bias (ADQB) and asymptotic distributional quadratic risk (ADQR) of the stated estimators. The results of the Monte-Carlo simulations are given in Section 4. An empirical example is given in Section 5 and the concluding remarks are made in the last section. All computations are done with the latest version of freeware R, and mathematical proofs are given in the Appendix. The matrices and vectors are represented with boldface symbols, while script letters $\mathbb{E}$ and $\mathbb{V}$ are reserved for the operators of expectation and variance.

2. Estimation Strategies

In this section, some improved estimation strategies for the kurtosis parameter vector $\boldsymbol{\beta}$ for q multivariate populations are described. At the first place, Using Mardia's estimator given in equation (2), the unrestricted estimator (UE) of $\boldsymbol{\beta}$ using only the SI is defined as

$$\hat{\boldsymbol{\beta}}^{UE} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_q)^T. \tag{4}$$

The restricted estimator (RE) of the parameter vector $\boldsymbol{\beta}$ under UPI is given as

$$\hat{\boldsymbol{\beta}}^{RE} = \left(\hat{\beta}_0, \hat{\beta}_0, \cdots, \hat{\beta}_0\right)^T = \hat{\beta}_0 \mathbf{1}_q. \tag{5}$$

The point estimator of β_0 denoted by $\hat{\beta}_0$ is calculated as $\hat{\beta}_0 = \hat{\omega}^{-1}\sum_{l=1}^{q}\lambda_{l,n}\hat{\beta}_l^{UE}$, where $\hat{\omega} = \sum_{l=1}^{q}\lambda_{l,n}$ and $\lambda_{l,n} = n_l/n$. An alternative form of $\hat{\boldsymbol{\beta}}^{RE}$, which is further used for the derivation of the mathematical results, is given as

$$\hat{\boldsymbol{\beta}}^{RE} = \hat{\omega}^{-1}\boldsymbol{J}_q\boldsymbol{V}_n^{-1}\hat{\boldsymbol{\beta}}^{UE} = \boldsymbol{H}_n\hat{\boldsymbol{\beta}}^{UE}, \tag{6}$$

where $\boldsymbol{J}_q = \mathbf{1}_q\mathbf{1}_q^T$, $\boldsymbol{H}_n = \hat{\omega}^{-1}\boldsymbol{J}_q\boldsymbol{V}_n^{-1}$, $\boldsymbol{V}_n = \text{diag}\left(\frac{v}{\lambda_{1,n}}, \frac{v}{\lambda_{2,n}}, \cdots, \frac{v}{\lambda_{q,n}}\right)$, and $v = 8p(p+2)$. Assuming $\lim(\lambda_{l,n}) = \lambda_l$ is fixed for $l = 1, 2, \cdots, q$ then $\boldsymbol{V}_n$ converges in probability to $\boldsymbol{V} = \text{diag}\left(\frac{v}{\lambda_1}, \frac{v}{\lambda_2}, \cdots, \frac{v}{\lambda_q}\right)$ as $n \to \infty$ where $n = n_1 + n_2 + \cdots + n_q$. The linear shrinkage (LS) estimator of $\boldsymbol{\beta}$ may be defined as

$$\hat{\boldsymbol{\beta}}^{LS} = \hat{\boldsymbol{\beta}}^{UE} - \pi(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}); \quad \pi \in (0, 1), \tag{7}$$

where π is the degree of trust in the null hypothesis (3). The linear shrinkage estimator becomes a restricted estimator when π is one and an unrestricted estimator when it is zero.

The preliminary test or simply pretest (PT) estimator of the population parameter vector $\boldsymbol{\beta}$ is stated as

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}^{UE} - \left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) I(\mathfrak{L}_n < c_{n,\alpha}), \tag{8}$$

where $I(\cdot)$ is an indicator function and $\mathfrak{L}_n$ is a Wald-type test statistic computed as

$$\mathfrak{L}_n = n\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right)^T \boldsymbol{V}_n^{-1}\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right). \tag{9}$$

Under the null hypothesis given in equation (3), $\mathfrak{L}_n$ converges in distribution to a χ^2 distribution with $(q-1)$ degrees of freedom. Thus, upper α-level critical values of $\mathfrak{L}_n$ defined by $c_{n,\alpha}$ are approximated by a $\chi^2_{(q-1)}$ distribution. The shrinkage pretest (SP) estimator which incorporates π into equation (8) is given as

$$\hat{\boldsymbol{\beta}}^{SP} = \hat{\boldsymbol{\beta}}^{UE} - \pi\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) I(\mathfrak{L}_n < c_{n,\alpha}). \tag{10}$$

It is interesting to note that for $\pi = 1$, the shrinkage preliminary test estimator $\hat{\boldsymbol{\beta}}^{SP}$ is reduced to the preliminary test estimator $\hat{\boldsymbol{\beta}}^{PT}$.

The Stein-type shrinkage (SS) estimator is defined as

$$\hat{\boldsymbol{\beta}}^{SS} = \hat{\boldsymbol{\beta}}^{UE} - (q-3)\mathfrak{L}_n^{-1}\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right), \quad q \geq 4, \tag{11}$$

and an improved Stein-type shrinkage (S+) estimator for $q \geq 4$ is given as

$$\hat{\boldsymbol{\beta}}^{S+} = \hat{\boldsymbol{\beta}}^{UE} - (q-3)\mathfrak{L}_n^{-1}\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) - \left[1-(q-3)\mathfrak{L}_n^{-1}\right] I(\mathfrak{L}_n < q-3)\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right). \tag{12}$$

Secondly, it is to be noted that the test-statistic given in equation (9) is consistent against fixed $\boldsymbol{\beta}$ such that $\boldsymbol{\beta} \notin H_0$, hence all the estimators involving $\mathfrak{L}_n$ are equivalent to the bench-mark estimator $\hat{\boldsymbol{\beta}}^{UE}$ for the fixed alternatives in a large sample setup. Interested readers are referred to Ahmed (2002) for further details. Therefore, we consider a sequence of local alternatives $\{\mathcal{H}_{(n)}\}$ as

$$\mathcal{H}_{(n)} : \boldsymbol{\beta} = \boldsymbol{\beta}_{(n)}, \tag{13}$$

where $\boldsymbol{\beta}_{(n)} = \beta_0 \mathbf{1}_q + n^{-1}\boldsymbol{\delta}$, and $\boldsymbol{\delta} \in \mathbb{R}^q$ is a fixed real vector. It is to be noted that we are not making any assumption for local alternatives setup either. By virtue of Ahmed *et al.* (2012) and under the sequence of local alternatives defined in equation (13) in the univariate sense, following result holds:

$$\sqrt{n_l}(\hat{\beta}_l - \beta_l) \sim N(\delta, v). \tag{14}$$

Following Appendix A.1 in Zahra *et al.* (2017 a) with aforementioned local alternative and using equation (14), the asymptotic distribution of $\hat{\boldsymbol{\beta}}^{UE}$ is

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}) \sim N(\mathbf{0}, \boldsymbol{V}). \tag{15}$$

3. Asymptotic Results

In this section, the analytical results regarding the asymptotic properties of the aforementioned estimators are presented. Using the asymptotic framework cited in Section 3 of Shah *et al.* (2020), and under the sequence of local alternatives given in equation (13), following evaluation criterion are used to assess the performance of the estimators under consideration. We have omitted the special cases of estimators (as mentioned above) from discussions in order to save space, as one can deduce their results by using their relations.

3.1. Asymptotic distributional bias

The vector of asymptotic distributional bias (ADB) of an estimator $\hat{\boldsymbol{\beta}}^*$ is calculated as

$$\mathfrak{B}(\hat{\boldsymbol{\beta}}^*) = \lim_{n\to\infty} \mathbb{E}\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)\right). \tag{16}$$

Theorem 1: Expressions for ADB of various estimators under sequence of local alternatives are given as

$$\mathfrak{B}(\hat{\boldsymbol{\beta}}^{LS}) = -\pi\boldsymbol{\delta}^\star,$$
$$\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SP}) = -\pi\boldsymbol{\delta}^\star\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta),$$
$$\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) = -(q-3)\boldsymbol{\delta}^\star\mathbb{E}\left(\chi^{-2}_{q+1}(\Delta)\right),$$
$$\mathfrak{B}(\hat{\boldsymbol{\beta}}^{S+}) = -\boldsymbol{\delta}^\star[\Phi_{q+1}(q-3;\Delta) + (q-3)\mathbb{E}\{\chi^{-2}_{q+1}(\Delta)I(\chi^2_{q+1}(\Delta) > (q-3))\}],$$

where $\boldsymbol{\delta}^\star = \boldsymbol{C}_0\boldsymbol{\delta}$, $\boldsymbol{C}_0 = \boldsymbol{I}_q - \boldsymbol{H}_0$, and $\Phi_{q+1}(\cdot;\Delta)$ is the cumulative distribution function (CDF) of a non-central chi-square distribution with $q+1$ degrees of freedom and non-centrality parameter Δ.

Proof: See Appendix 1 for the proof. □

3.2. Asymptotic distributional quadratic bias

The expressions given in Theorem 1 are in vector form and for comparison purposes, we needed expressions in scalar form. Thus, we applied following quadratic transformation that yielded asymptotic distributional quadratic bias (ADQB) of the competing estimators:

$$\mathfrak{B}^\star(\hat{\boldsymbol{\beta}}^*) = \left(\mathfrak{B}(\hat{\boldsymbol{\beta}}^*)\right)^T \boldsymbol{V}^{-1}\mathfrak{B}(\hat{\boldsymbol{\beta}}^*). \tag{17}$$

Theorem 2: Under the sequence of local alternatives, expressions for ADQB of various estimators are given as

$$\begin{aligned}
\mathfrak{B}^\star(\hat{\boldsymbol{\beta}}^{LS}) &= \pi^2\Delta,\\
\mathfrak{B}^\star(\hat{\boldsymbol{\beta}}^{SP}) &= \pi^2\Delta\left[\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta)\right]^2,\\
\mathfrak{B}^\star(\hat{\boldsymbol{\beta}}^{SS}) &= (q-3)^2\Delta\left[\mathbb{E}\left\{\chi^{-2}_{q+1}(\Delta)\right\}\right]^2,\\
\mathfrak{B}^\star(\hat{\boldsymbol{\beta}}^{S+}) &= \Delta\left[\Phi_{q+1}(q-3;\Delta) + (q-3)\mathbb{E}\left\{\chi^{-2}_{q+1}(\Delta)I(\chi^2_{q+1}(\Delta) > (q-3))\right\}\right]^2.
\end{aligned}$$

Proof: Using the transformation mentioned in equation (17) and with the help of following Lemma, the proof is straightforward.

Lemma 3: The test statistic given in equation (9) converges to a non-central χ^2 distribution with $(q-1)$ degrees of freedom and non-centrality parameter $\Delta = (\boldsymbol{\delta}^\star)^T\boldsymbol{V}^{-1}\boldsymbol{\delta}^\star$ as $n\to\infty$. □

Note that the estimators based on the strategies of preliminary test and Stein-type shrinkage methodologies are biased.

3.3. Asymptotic mean square error matrix

The asymptotic mean-square error matrices are needed for the computations of asymptotic distributional quadratic risk (ADQR) expressions of the various estimators mentioned above. Under the sequence of local alternative, the general expression for such matrices is given as

$$\mathfrak{S}(\hat{\boldsymbol{\beta}}^*) = \lim_{n\to\infty}\mathbb{E}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)^T\right]. \tag{18}$$

Theorem 4: The expressions for the asymptotic mean square error matrices for various estimators, under the sequence of local alternatives and for $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{C}_0^T$, are given as follows:

$$\begin{aligned}
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{LS}) &= \boldsymbol{V} - \pi(2-\pi)\boldsymbol{C} + \pi^2\boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T, \\
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{SP}) &= \boldsymbol{V} - \pi(2-\pi)\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta)\boldsymbol{C} \\
&\quad + \pi\boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T\left[2\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta) - (2-\pi)\Phi_{q+3}(\chi^2_{q-1,\alpha};\Delta)\right], \\
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{SS}) &= \boldsymbol{V} - (q-3)\boldsymbol{C}\Big[2\mathbb{E}\left[\chi^{-2}_{q+1}(\Delta)\right] - (q-3)\mathbb{E}\left[\chi^{-4}_{q+1}(\Delta)\right]\Big] \\
&\quad + (q-3)(q+1)\boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T\mathbb{E}\left[\chi^{-4}_{q+3}(\Delta)\right], \\
\mathfrak{S}(\hat{\boldsymbol{\beta}}^{S+}) &= \mathfrak{S}(\hat{\boldsymbol{\beta}}^{SS}) - \boldsymbol{C}\mathbb{E}\Big[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\}^2 I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\Big] \\
&\quad + \boldsymbol{\delta}^\star(\boldsymbol{\delta}^\star)^T\begin{bmatrix} 2\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\} I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\right] \\ -\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+3}(\Delta)\right\}^2 I\left(\chi^2_{q+3}(\Delta) < (q-3)\right)\right]\end{bmatrix},
\end{aligned}$$

Proof: Following Appendix 7 of Zahra *et al.* (2017 a), the proof can be completed. □

3.4. Asymptotic distributional quadratic risk

The asymptotic distributional quadratic risk (ADQR) of an estimator $\hat{\boldsymbol{\beta}}^\star$ of the parameter vector $\boldsymbol{\beta}$ is defined as

$$\begin{aligned}
\mathfrak{R}(\hat{\boldsymbol{\beta}}^\star;\boldsymbol{W}) &= \lim_{n\to\infty}\mathbb{E}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)^T\boldsymbol{W}\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)\right] \\
&= \lim_{n\to\infty}\operatorname{tr}\left[\boldsymbol{W}\mathbb{E}\left\{\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)\sqrt{n}\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_{(n)}\right)^T\right\}\right] = \operatorname{tr}\left[\boldsymbol{W}\mathfrak{S}(\hat{\boldsymbol{\beta}}^*)\right], \qquad (19)
\end{aligned}$$

where $\boldsymbol{W}$ is a $(q\times q)$ positive semi-definite (psd) weight matrix.

Theorem 5: Under the sequence of local alternatives, expressions of ADQR for the various estimators are given as

$$\begin{aligned}
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{LS};\boldsymbol{W}) &= \operatorname{tr}(\boldsymbol{W}\boldsymbol{V}) - \pi(2-\pi)\operatorname{tr}(\boldsymbol{W}\boldsymbol{C}) + \pi^2\Delta_W, \\
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{SP};\boldsymbol{W}) &= \operatorname{tr}(\boldsymbol{W}\boldsymbol{V}) - \pi(2-\pi)\operatorname{tr}(\boldsymbol{W}\boldsymbol{C})\,\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta) \\
&\quad + \pi\Delta_W\left[2\Phi_{q+1}(\chi^2_{q-1,\alpha};\Delta) - (2-\pi)\Phi_{q+3}(\chi^2_{q-1,\alpha};\Delta)\right], \\
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS};\boldsymbol{W}) &= \operatorname{tr}(\boldsymbol{W}\boldsymbol{V}) - (q-3)\operatorname{tr}(\boldsymbol{W}\boldsymbol{C})\left[2\mathbb{E}\left[\chi^{-2}_{q+1}(\Delta)\right] - (q-3)\mathbb{E}\left[\chi^{-4}_{q+1}(\Delta)\right]\right] \\
&\quad + (q-3)(q+1)\Delta_W\mathbb{E}\left[\chi^{-4}_{q+3}(\Delta)\right], \\
\mathfrak{R}(\hat{\boldsymbol{\beta}}^{S+};\boldsymbol{W}) &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS};\boldsymbol{W}) \\
&\quad - \operatorname{tr}(\boldsymbol{W}\boldsymbol{C})\,\mathbb{E}\Big[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\}^2 I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\Big] \\
&\quad + \Delta_W\begin{bmatrix} 2\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+1}(\Delta)\right\} I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\right] \\ -\mathbb{E}\left[\left\{1-(q-3)\chi^{-2}_{q+3}(\Delta)\right\}^2 I\left(\chi^2_{q+3}(\Delta) < (q-3)\right)\right]\end{bmatrix},
\end{aligned}$$

where $\Delta_W = (\boldsymbol{\delta}^\star)^T \boldsymbol{W} \boldsymbol{\delta}^\star$ is the non-centrality parameter involving weight matrix $\boldsymbol{W}$.

Proof: Using the asymptotic mean square matrices given in Theorem 4 and the definition of ADQR in equation (19), the proof is trivial. □

Corollary 6: For the choice $\boldsymbol{W} = \boldsymbol{V}^{-1}$, ADQR expressions given above are simplified and are given as

$$\begin{aligned}
R_1 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{V}^{-1}) = q, \\
R_2 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{RE}; \boldsymbol{V}^{-1}) = 1 + \Delta, \\
R_3 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{LS}; \boldsymbol{V}^{-1}) = q - \pi(2-\pi)(q-1) + \pi^2\Delta, \\
R_4 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{PT}; \boldsymbol{V}^{-1}) = q - (q-1)\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) \\
&\qquad + \Delta\left[2\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) - \Phi_{q+3}(\chi^2_{q-1,\alpha}; \Delta)\right], \\
R_5 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SP}; \boldsymbol{V}^{-1}) = q - \pi(2-\pi)(q-1)\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) \\
&\qquad + \pi\Delta\left[2\Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta) - (2-\pi)\Phi_{q+3}(\chi^2_{q-1,\alpha}; \Delta)\right], \\
R_6 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS}; \boldsymbol{V}^{-1}) = q - (q-1)(q-3)\Big[2\mathbb{E}\left[\chi^{-2}_{q+1}(\Delta)\right] - (q-3)\mathbb{E}\left[\chi^{-4}_{q+1}(\Delta)\right]\Big] \\
&\qquad + (q-3)(q+1)\Delta\mathbb{E}\left[\chi^{-4}_{q+3}(\Delta)\right], \\
R_7 &= \mathfrak{R}(\hat{\boldsymbol{\beta}}^{S+}; \boldsymbol{V}^{-1}) = \mathfrak{R}(\hat{\boldsymbol{\beta}}^{SS}; \boldsymbol{V}^{-1}) \\
&\qquad - (q-1)\mathbb{E}\Big[\left\{1 - (q-3)\chi^{-2}_{q+1}(\Delta)\right\}^2 I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\Big] \\
&\qquad + \Delta\begin{bmatrix} 2\mathbb{E}\left[\left\{1 - (q-3)\chi^{-2}_{q+1}(\Delta)\right\} I\left(\chi^2_{q+1}(\Delta) < (q-3)\right)\right] \\ -\mathbb{E}\left[\left\{1 - (q-3)\chi^{-2}_{q+3}(\Delta)\right\}^2 I\left(\chi^2_{q+3}(\Delta) < (q-3)\right)\right] \end{bmatrix}.
\end{aligned}$$

Proof: This proof of Corollary 6 can be completed by replacing $\boldsymbol{W} = \boldsymbol{V}^{-1}$ and noting that $\text{tr}(\boldsymbol{W}\boldsymbol{V}) = \text{tr}(\boldsymbol{I}_q) = q$, $\text{tr}(\boldsymbol{W}\boldsymbol{C}) = \text{tr}(\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{C}_0^T) = \text{tr}(\boldsymbol{C}_0^T) = \text{tr}(\boldsymbol{I}_q - \boldsymbol{H}_0)^T = \text{tr}(\boldsymbol{I}_q - \boldsymbol{H}_0) = q-1$, and $\Delta_W = (\boldsymbol{\delta}^\star)^T \boldsymbol{V} \boldsymbol{\delta}^\star = \Delta$. □

3.5. Risk comparison

In this section, the performance of restricted (RE), linear shrinkage (LS), both preliminary test (PT and SP) and Stein-type shrinkage estimators (SS and S+) is compared with the benchmark unrestricted estimator (UE) using the aforementioned simplified ADQR expressions of Corollary 6. For this purpose, we have defined the notion of asymptotic relative efficiency (AREFF) of an estimator $\hat{\boldsymbol{\beta}}^*$ with reference to $\hat{\boldsymbol{\beta}}^{UE}$ as

$$\text{AREFF}(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\beta}}^{UE}) = \frac{\mathfrak{R}(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{V}^{-1})}{\mathfrak{R}(\hat{\boldsymbol{\beta}}^*; \boldsymbol{V}^{-1})} = \frac{R_1}{R_j}; \quad j \leq 7. \tag{20}$$

An estimator is considered to be more efficient, in asymptotic terms, if AREFF exceeds 1, and vice versa. Since all the risk expressions are the function of a drift parameter Δ, we have plotted the AREFF of all the competing estimators against Δ in order to compare their performance in Figures 1–4, while fixing $q = 4, 8, 10$, $\pi = 0.50$, and $\alpha = 0.01, 0.05, 0.10, 0.20$.

From Figure 1, it can be seen that both $\hat{\boldsymbol{\beta}}^{RE}$ and $\hat{\boldsymbol{\beta}}^{LS}$ have higher efficiencies than $\hat{\boldsymbol{\beta}}^{UE}$ (which is constant), and that the AREFF decreases as Δ increases but the decay of $\hat{\boldsymbol{\beta}}^{RE}$ is much faster than that of $\hat{\boldsymbol{\beta}}^{LS}$.

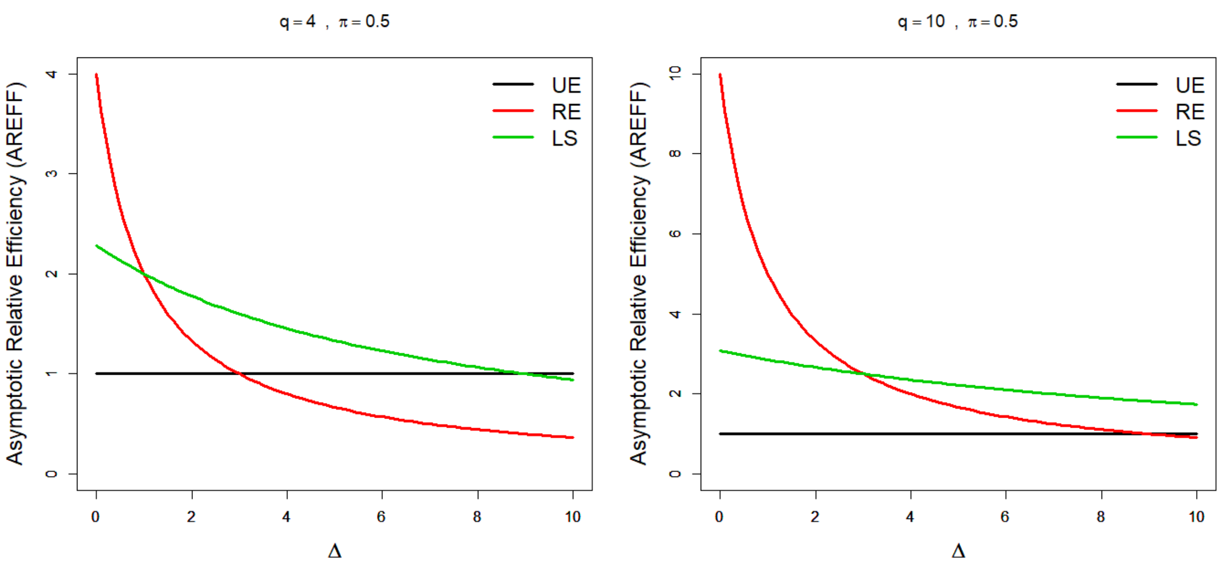

Figure 1: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{UE}$, $\hat{\boldsymbol{\beta}}^{RE}$, and $\hat{\boldsymbol{\beta}}^{LS}$

AREFFs of the unrestricted estimator and the preliminary test estimators are compared in Figure 2. It can be observed that the $\hat{\boldsymbol{\beta}}^{PT}$ dominates the $\hat{\boldsymbol{\beta}}^{SP}$ in a region where the drift parameter Δ is smaller *i.e.*, when the null hypothesis (3) is true or nearly true. But for larger values of Δ, the situation is reversed, and it is $\hat{\boldsymbol{\beta}}^{SP}$ that outperforms $\hat{\boldsymbol{\beta}}^{PT}$ for all choices of α. Moreover, curves of AREFFs of $\hat{\boldsymbol{\beta}}^{PT}$ are approaching $\hat{\boldsymbol{\beta}}^{SP}$ for larger values of α. This means that for smaller values of α, the region where $\hat{\boldsymbol{\beta}}^{PT}$ dominates $\hat{\boldsymbol{\beta}}^{SP}$ is more spacious.

From Figure 3, it is evident that $\hat{\boldsymbol{\beta}}^{S+}$ is far superior to $\hat{\boldsymbol{\beta}}^{SS}$ and $\hat{\boldsymbol{\beta}}^{UE}$, uniformly dominating both estimators. Figure 4 establishes the dominance of the restricted estimator $\hat{\boldsymbol{\beta}}^{RE}$ over the Stein-type shrinkage estimator $\hat{\boldsymbol{\beta}}^{SS}$ and shrinkage preliminary test estimator $\hat{\boldsymbol{\beta}}^{SP}$ for smaller values of Δ. Even the shrinkage preliminary test estimator $\hat{\boldsymbol{\beta}}^{SP}$ performs better than $\hat{\boldsymbol{\beta}}^{SS}$ in a reasonable region over Δ.

4. Monte-Carlo Simulations

In this section, we conducted extensive Monte-Carlo simulations to examine the performance of the various estimators discussed earlier for $\boldsymbol{\beta}$, which incorporates UPI into the estimation procedure. The performance of the estimators is investigated by comparing their simulated relative efficiencies (SRE). The SRE of an estimator $\hat{\boldsymbol{\beta}}^{*}$ to a benchmark unre-

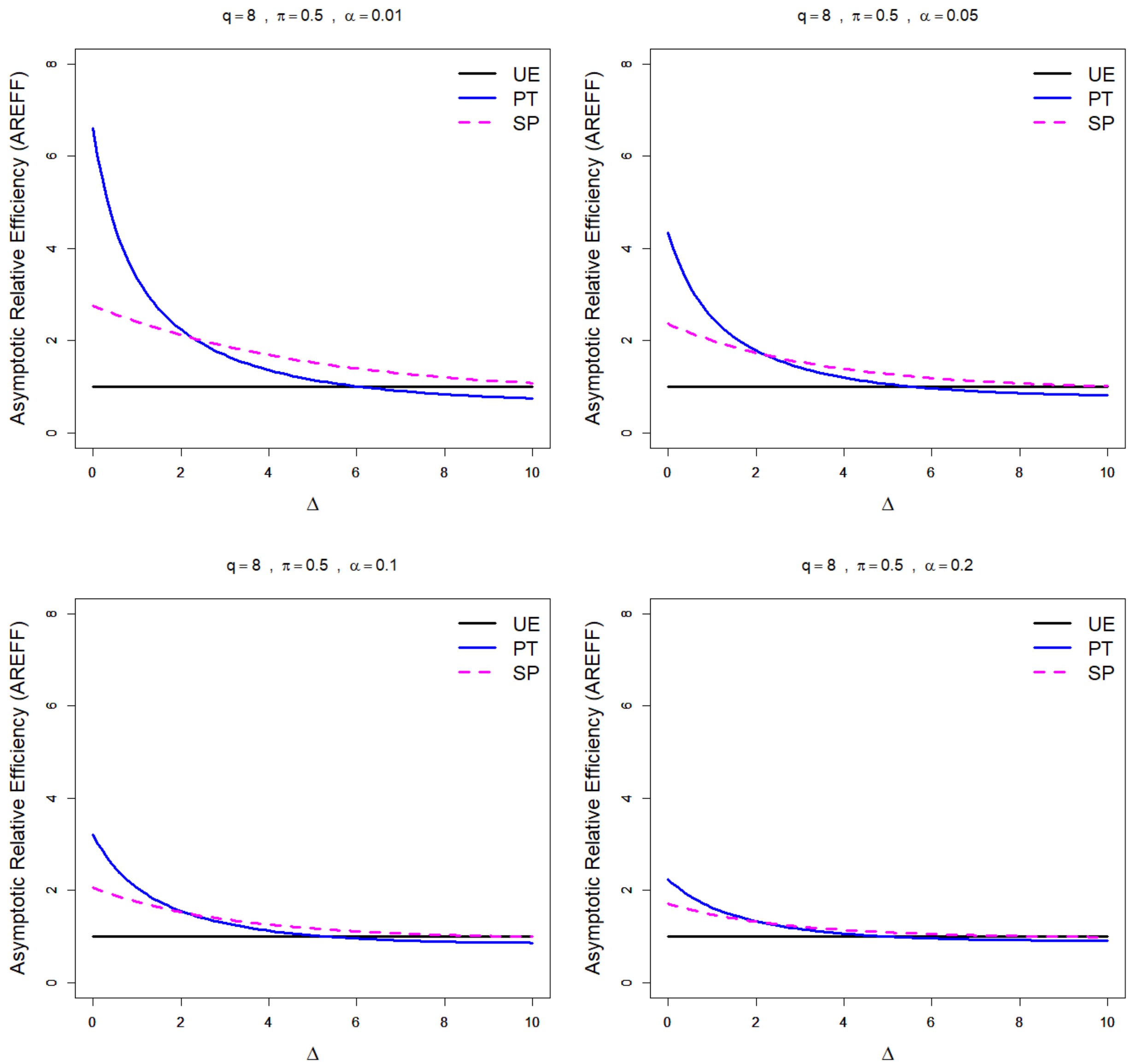

Figure 2: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{UE}$, $\hat{\boldsymbol{\beta}}^{PT}$, and $\hat{\boldsymbol{\beta}}^{SP}$

stricted estimator $\hat{\boldsymbol{\beta}}^{UE}$ is defined by the ratio of their simulated risks as

$$\mathrm{SRE}(\hat{\boldsymbol{\beta}}^{*}, \hat{\boldsymbol{\beta}}^{UE}) = \frac{\text{Simulated Risk}(\hat{\boldsymbol{\beta}}^{UE})}{\text{Simulated Risk}(\hat{\boldsymbol{\beta}}^{*})}. \tag{21}$$

The value of an SRE greater than 1 indicates that $\hat{\boldsymbol{\beta}}^{*}$ is superior to $\hat{\boldsymbol{\beta}}^{UE}$. Furthermore, we have defined a parameter Δ^* (which is essentially a measure of how far away we deviate from the hypothesized common kurtosis vector $\boldsymbol{\beta}_{\mathbf{0}}$) as $\Delta^* = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)$.

Ahmed *et al.* (2012) mentioned comparisons for kurtosis of a multivariate normal distribution with t-distribution with degrees of freedom ranging from $\nu = 5$ to $\nu = 60$. Fol-

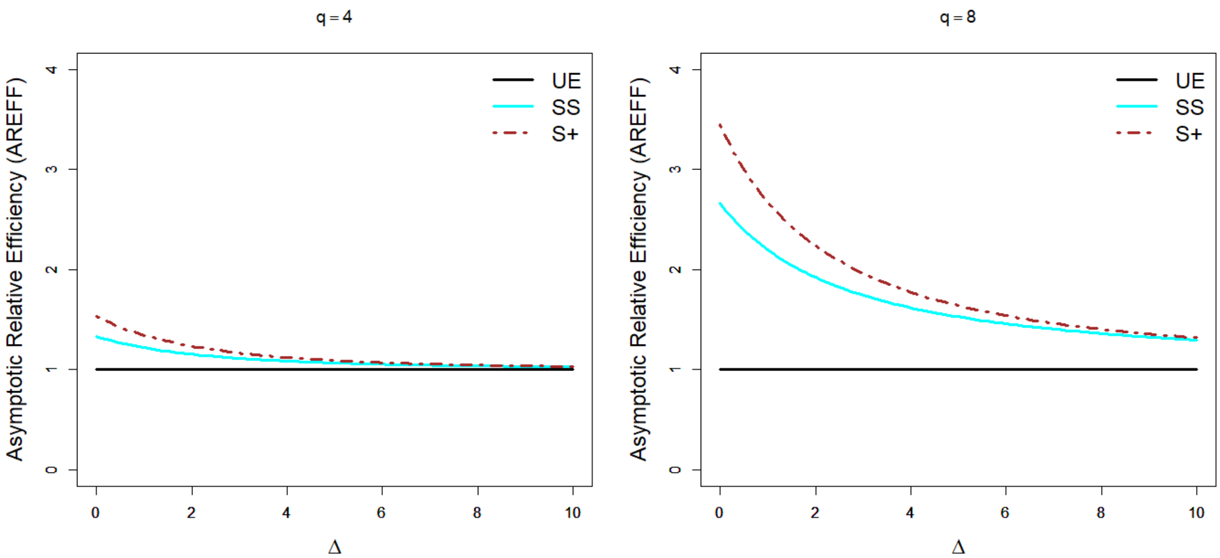

Figure 3: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{UE}$, $\hat{\boldsymbol{\beta}}^{SS}$, and $\hat{\boldsymbol{\beta}}^{S+}$

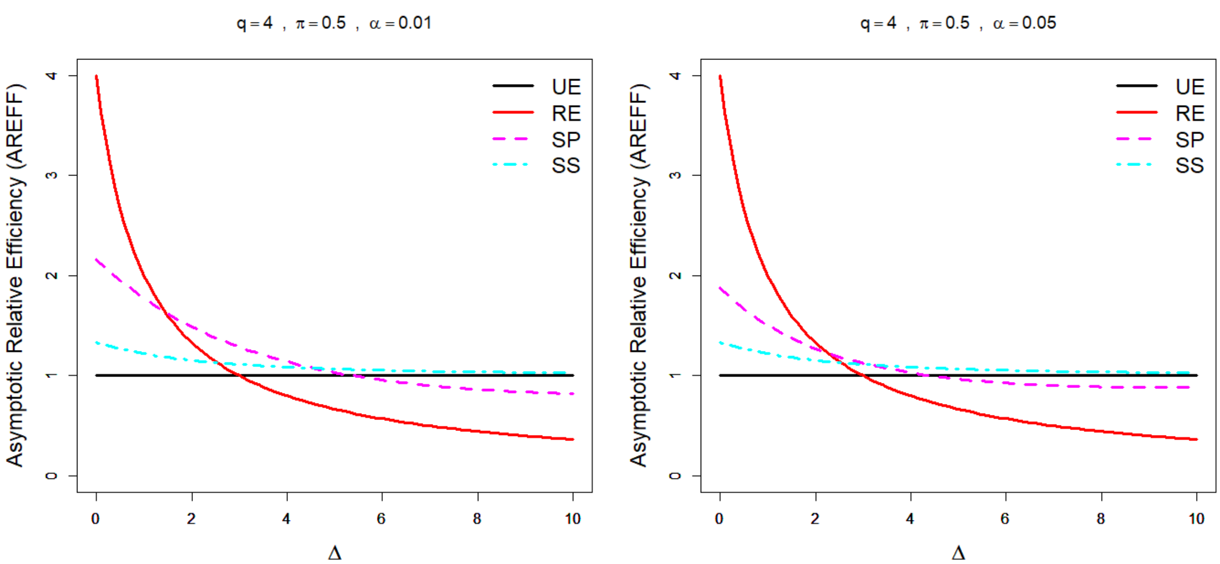

Figure 4: Asymptotic relative efficiencies of $\hat{\boldsymbol{\beta}}^{RE}$, $\hat{\boldsymbol{\beta}}^{SP}$, and $\hat{\boldsymbol{\beta}}^{SS}$

lowing the same idea, we have considered multivariate normal distributions with dimensions $p = 2$ and 4 initially, when the null hypothesis is assumed to be true. In order to study the deviation of the data from the null hypothesis, further samples are taken from multivariate t-distribution with various degrees of freedom between 5 and 60.

Simulation processes are repeated $N = 5000$ times for different choices of $q = 4$, 6, 8, 10 and n_l = 80, 100, 150, 200. The choices of α and π are assumed to be 0.05, 0.10 and 0.25, 0.50, 0.75, respectively. Simulated relative efficiencies of the suggested estimators for different configurations of p, q, n_l, α, and π against various values of $\Delta^* \geq 0$ were computed; results are reported in Table 1 for only $p = 4$, $\pi = 0.50$ and $n_l = 100$ to conserve space. A graphical representation is shown in Figure 5.

Table 1: Simulated relative efficiencies of estimators when $p = 4$, $\pi = 0.50$ and $n_l = 100$

q	Δ^*	$\hat{\boldsymbol{\beta}}^{RE}$	$\hat{\boldsymbol{\beta}}^{LS}$	$\hat{\boldsymbol{\beta}}^{PT}$		$\hat{\boldsymbol{\beta}}^{SP}$		$\hat{\boldsymbol{\beta}}^{SS}$	$\hat{\boldsymbol{\beta}}^{S+}$
				$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$		
4	**0.00**	4.00	2.29	2.41	2.00	1.78	1.60	1.35	1.50
	0.58	2.52	2.02	1.60	1.43	1.46	1.34	1.23	1.32
	1.30	1.22	1.55	1.02	1.01	1.12	1.09	1.09	1.13
	1.89	0.82	1.25	0.86	0.89	0.99	0.99	1.02	1.04
	3.46	0.51	0.89	0.81	0.86	0.91	0.94	0.98	0.99
	10.39	0.45	0.69	0.97	0.98	0.99	0.99	0.99	0.99
6	**0.00**	5.91	2.65	3.03	2.39	2.01	1.77	1.86	2.33
	0.50	3.17	2.28	1.97	1.74	1.68	1.54	1.58	1.90
	1.12	1.58	1.86	1.16	1.12	1.24	1.18	1.29	1.37
	1.63	0.96	1.42	0.92	0.93	1.06	1.04	1.09	1.14
	2.98	0.53	0.93	0.80	0.84	0.92	0.94	0.99	0.99
	8.94	0.43	0.69	0.97	0.98	0.98	0.99	0.99	0.99
8	**0.00**	7.84	2.89	3.57	2.77	2.17	1.92	2.43	3.15
	0.44	2.98	2.08	1.62	1.45	1.42	1.32	1.72	1.79
	0.99	1.66	1.85	1.27	1.22	1.34	1.26	1.39	1.55
	1.44	1.07	1.52	0.98	0.98	1.12	1.09	1.18	1.24
	2.65	0.57	0.99	0.81	0.85	0.93	0.95	0.99	0.99
	7.94	0.42	0.68	0.96	0.97	0.98	0.99	0.99	0.99
10	**0.00**	9.91	3.07	3.85	2.97	2.25	1.99	3.05	3.94
	0.40	4.02	2.41	1.88	1.63	1.57	1.43	2.14	2.28
	0.90	2.23	2.05	1.36	1.27	1.32	1.24	1.68	1.74
	1.31	1.30	1.72	1.08	1.06	1.20	1.15	1.34	1.41
	2.40	0.61	1.04	0.83	0.86	0.95	0.96	0.99	0.99
	7.20	0.42	0.68	0.96	0.97	0.98	0.99	0.99	0.99

The restricted estimator $\hat{\boldsymbol{\beta}}^{RE}$ performed better when homogeneity assumption of kurtosis parameters holds, but when it is not true, the SRE of the restricted estimator declined rapidly, and approaches zero for larger values of Δ^*. The SRE of the linear shrinkage estimator $\hat{\boldsymbol{\beta}}^{LS}$ declined slowly and its performance is comparable to shrinkage estimators only when q and Δ^* are small. The SRE of the pretest $\hat{\boldsymbol{\beta}}^{PT}$ and shrinkage pretest $\hat{\boldsymbol{\beta}}^{SP}$ estimators declined as Δ^* increased, but after reaching a minimum value, it attained a value of one again. Both pretest estimators performed well for smaller values of $q \leq 4$ and Δ^*. The Stein-type estimators $\hat{\boldsymbol{\beta}}^{SS}$ and $\hat{\boldsymbol{\beta}}^{S+}$ performed better than all other suggested estimators in the wider range of Δ^*, especially as q increases. In short, simulation study endorsed the analytical findings.

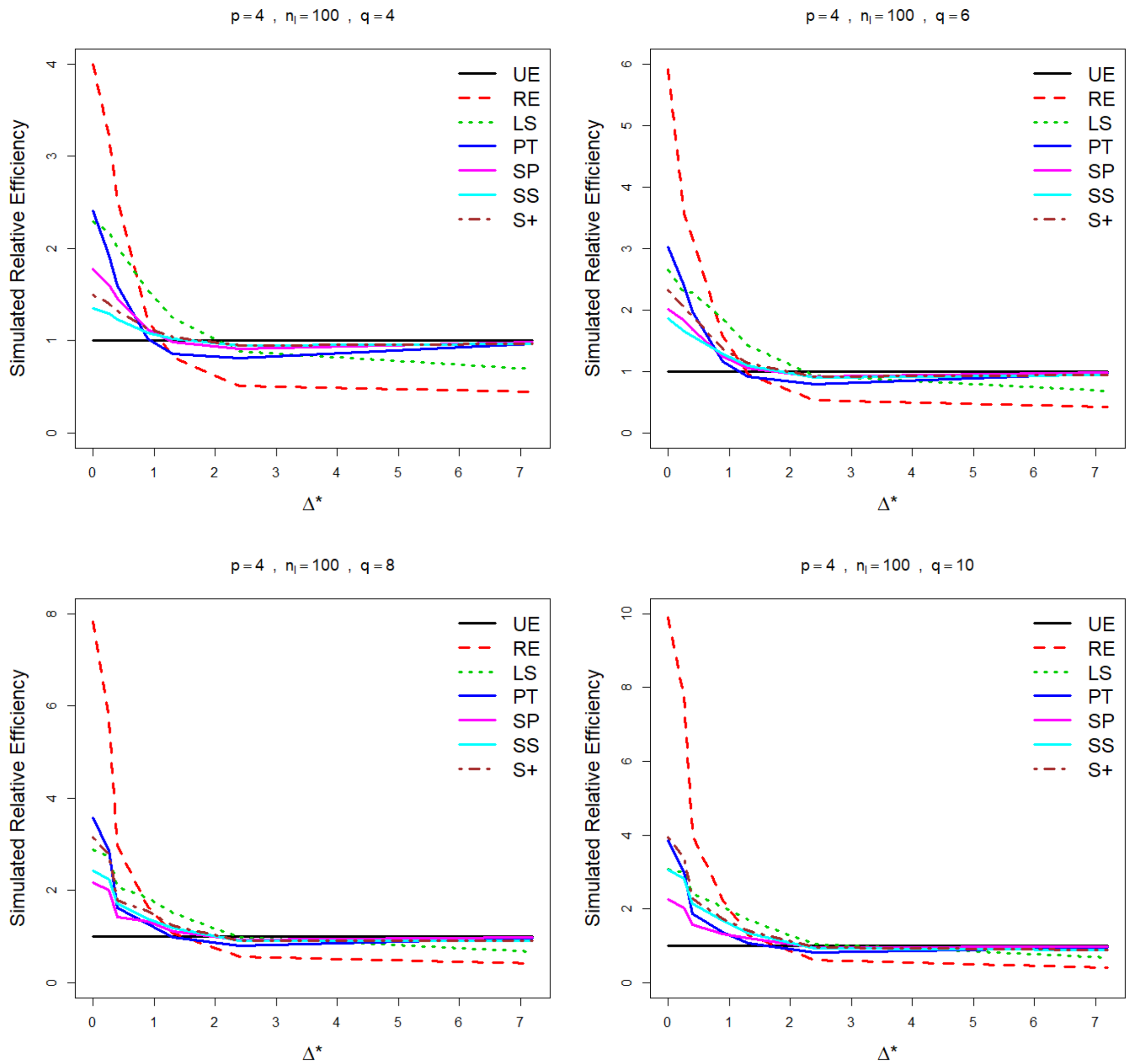

Figure 5: Simulated relative efficiency of the estimators

5. Data Application

Our real-data application considered the four-dimensional multivariate data of geographical regions of Europe, based on monthly long-term interest rates in percentages. Four countries of each region are considered; Central Europe: Austria, Germany, Hungry, and Poland; Southern Europe: Spain, Italy, Portugal, and Slovenia; Western Europe: Belgium, France, Netherlands, and Switzerland; Northern Europe: United Kingdom, Norway, Denmark, and Sweden. The data comprises of 100 observations for each country from August 2007 to November 2015, as reported by the Organisation for Economic Co-operation and Development (OECD) website.

The unrestricted estimator is obtained as $\hat{\boldsymbol{\beta}}^{UE} = (25.16, 27.20, 28.17, 20.45)^T$. Next, we want to test the hypothesis that the kurtosis parameters are same for all four regions, against the alternative that at least one of them is different from others. The test statistic $\mathfrak{L}_n$ is 18.42, therefore, we reject the null hypothesis at $\alpha = 0.05$. The suggested estimators are calculated for $\pi = 0.50$ and $\alpha = 0.05$ as $\hat{\boldsymbol{\beta}}^{RE} = (25.24, 25.24, 25.24, 25.24)^T$; the pretest, shrinkage pretest and Stein-type estimators are equal to the unrestricted estimator, while linear shrinkage and improved Stein-type shrinkage estimators are given as $\hat{\boldsymbol{\beta}}^{LS} = (25.20, 26.22, 26.71, 22.85)^T$ and $\hat{\boldsymbol{\beta}}^{S+} = (25.17, 27.10, 28.01, 20.71)^T$.

Bootstrap methodology is used to assess the performance of the suggested estimators. Samples of equal size $n_l = 100$ are selected from each country with replacement; this process is repeated $N = 5000$ times. All the suggested estimators are computed for $\pi = 0.25, 0.50, 0.75$ and $\alpha = 0.05, 0.10, 0.30$. Simulated relative efficiencies of the estimators under consideration relative to the unrestricted estimator are computed and reported in the following table:

Table 2: Relative efficiencies of estimators for long-term interest rates data based on multivariate bootstrap samples

α	π	$\hat{\boldsymbol{\beta}}^{RE}$	$\hat{\boldsymbol{\beta}}^{LS}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{SP}$	$\hat{\boldsymbol{\beta}}^{SS}$	$\hat{\boldsymbol{\beta}}^{S+}$
	0.25	0.54	1.23	0.93	0.99	1.01	1.01
0.05	0.50	0.54	1.11	0.93	0.97	1.01	1.01
	0.75	0.54	0.80	0.93	0.95	1.01	1.01

It is revealed from the above table that the SRE of $\hat{\boldsymbol{\beta}}^{RE}$ is less than 1, which is in line with the analytical and simulated results; this confirms when the null hypothesis is not true, the restricted estimator performs inferior to all other estimators. The SRE of $\hat{\boldsymbol{\beta}}^{LS}$ declines as the value of π increases. $\hat{\boldsymbol{\beta}}^{PT}$ and $\hat{\boldsymbol{\beta}}^{SP}$ both have SREs smaller than 1 and their SREs remain below 1 as π increases. However, we recommend using the positive part of the Stein-type shrinkage estimator, since its performance is not drastically impacted by departure from the null hypothesis.

6. Concluding Remarks

In this paper, we have discussed the asymptotic theory of simultaneous estimation of kurtosis parameters for q multivariate normal distributions using the UPI that all kurtosis parameters are homogeneous. It is concluded that the performance of restricted and pretest estimators is better when the null hypothesis of equal kurtosis parameters holds, while the risk of the restricted estimator becomes unbounded as we move away from the null hypothesis. The pretest estimator performs better than restricted and Stein-type shrinkage estimators but only in a certain region of the parametric space. Stein-type estimators, outperform the unrestricted estimator in the entire parametric space. The improved Stein-type shrinkage estimator is strongly recommended when the equality of parameters is uncertain for $q \geq 4$, while for small dimensions, shrinkage pretest estimator is a better choice.

Acknowledgements

The research work of Professor S. Ejaz Ahmed was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Ahmed, S. E. (2002). Simultaneous estimation of coefficients of variation. *Journal of Statistical Planning and Inference*, **104(10)**, 31–51.

Ahmed, S. E., Omar, M. H. and Joarder, A. H. (2012). Stabilizing the performance of kurtosis estimator of multivariate data. *Communications in Statistics - Simulation and Computation*, **41(10)**, 1860–1871.

Araújo, T., Dias, J., Eleutério, S. and Louçã, F. (2012). *How Fama Went Wrong: Measures of Multivariate Kurtosis for the Identification of the Dynamics of a N-Dimensional Market*. Working Papers WP21/2012/DE/UECE, Department of Economics, School of Economics and Management, Technical University of Lisbon, Portugal.

Echer, E. and Bolzan, M. J. A. (2016). A comparative study of solar wind and foreshock turbulence near Uranus orbit. *Planetary and Space Science*, **120**, 70–77.

Kim, T. H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, **1(1)**, 56–73.

Lai, J. Y. (2012). An empirical study of the impact of skewness and kurtosis on hedging decisions. *Quantitative Finance*, **12(12)**, 1827–1837.

Liang, Z., Wei, J., Zhao, J., Liu, H., Li, B., Shen, J. and Zheng, C. (2008). The statistical meaning of kurtosis and its new application to identification of persons based on seismic signals. *Sensors*, **8(8)**, 5106–5119.

Lisawadi, S., Ahmed, S. E., Reangsephet, O. and Shah, M. K. A. (2019). Simultaneous estimation of Cronbach's alpha coefficients. *Communications in Statistics - Theory and Methods*, **48(13)**, 3236–3257.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

Nita, G. M. and Gary, D. E. (2010). Statistics of the spectral kurtosis estimator. *Publications of the Astronomical Society of the Pacific*, **122(891)**, 595–607.

Shah, M. K. A., Lisawadi, S. and Ahmed, S. E. (2017). Merging data from multiple sources: pretest and shrinkage perspectives. *Journal of Statistical Computation and Simulation*, **87(8)**, 1577–1592.

Shah, M. K. A., Zahra, N. and Ahmed, S. E. (2020). On the simultaneous estimation of Weibull reliability functions. In: *Proceedings of the Thirteenth International Conference on Management Science and Engineering Management. ICMSEM 2019*, *Advances in Intelligent Systems and Computing*, **1001**, 85–108. Springer, Cham.

Zahra, N., Lisawadi, S. and Ahmed, S. E. (2017 a). Meta-analysis, pretest, and shrinkage estimation of kurtosis parameters. *Communications in Statistics - Simulation and Computation*, **46(10)**, 7986–8004.

Zahra, N., Lisawadi, S. and Ahmed, S. E. (2017 b). Improved estimation of kurtosis parameters for two multivariate populations. *Lobachevskii Journal of Mathematics*, **38(1)**, 110–115.

APPENDIX

Mathematical Proofs

Two important results cited in Lemma 3 of Shah *et al.* (2020) are used, as well as some distributional results crucial for the derivation of mathematical results of appendices, are given in the following theorem:

Theorem 7: Under the sequence of local alternatives $\{\mathcal{H}_n\}$ and assuming independence among q components, following distributional result holds:

$$\begin{aligned}
\boldsymbol{\eta}_{1,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}\right) \xrightarrow{D} \boldsymbol{\eta}_1 \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{V}),\\
\boldsymbol{\eta}_{2,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}_0\right) \xrightarrow{D} \boldsymbol{\eta}_2 \sim \mathcal{N}_q(\boldsymbol{\delta}, \boldsymbol{V}),\\
\boldsymbol{\eta}_{3,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta}_0\right) \xrightarrow{D} \boldsymbol{\eta}_3 \sim \mathcal{N}_q(\mathbf{0}, v\boldsymbol{J}_q),\\
\boldsymbol{\eta}_{4,n} &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}\right) \xrightarrow{D} \boldsymbol{\eta}_4 \sim \mathcal{N}_q(\boldsymbol{\delta}^\star, \boldsymbol{C}),\\
\begin{pmatrix}\boldsymbol{\eta}_{2,n}\\ \boldsymbol{\eta}_{4,n}\end{pmatrix} &\xrightarrow{D} \begin{pmatrix}\boldsymbol{\eta}_2\\ \boldsymbol{\eta}_4\end{pmatrix} \sim \mathcal{N}_{2q}\left\{\begin{pmatrix}\boldsymbol{\delta}\\ \boldsymbol{\delta}^\star\end{pmatrix}, \begin{pmatrix}\boldsymbol{V} & \boldsymbol{C}\\ \boldsymbol{C}^T & \boldsymbol{C}\end{pmatrix}\right\},\\
\begin{pmatrix}\boldsymbol{\eta}_{3,n}\\ \boldsymbol{\eta}_{4,n}\end{pmatrix} &\xrightarrow{D} \begin{pmatrix}\boldsymbol{\eta}_3\\ \boldsymbol{\eta}_4\end{pmatrix} \sim \mathcal{N}_{2q}\left\{\begin{pmatrix}\mathbf{0}\\ \boldsymbol{\delta}^\star\end{pmatrix}, \begin{pmatrix}v\boldsymbol{J}_q & \mathbf{0}\\ \mathbf{0} & \boldsymbol{C}\end{pmatrix}\right\},
\end{aligned}$$

where $\xrightarrow{D}$ means convergence in distribution as $n \to \infty$.

Proof: See Appendices A1–A6 of Zahra *et al.* (2017 a) for detailed proof with some adjustments in notations. □

A1. Proof of Theorem 1

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{LS}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{LS} - \boldsymbol{\beta}_{(n)})\right] = \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{UE} - \pi(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}_0 - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right]\\
&= \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{2,n} - \boldsymbol{\delta} - \pi\boldsymbol{\eta}_{4,n}\right] = \mathbb{E}(\boldsymbol{\eta}_2) - \boldsymbol{\delta} - \pi\mathbb{E}(\boldsymbol{\eta}_4) = -\pi\boldsymbol{\delta}^\star.
\end{aligned}$$

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SP}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{SP} - \boldsymbol{\beta}_{(n)})\right]\\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{UE} - \pi(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) I(\mathfrak{L}_n < c_{n,\alpha}) - \boldsymbol{\beta}_0 - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right]\\
&= \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{2,n} - \boldsymbol{\delta} - \pi\boldsymbol{\eta}_{4,n} I(\mathfrak{L}_n < c_{n,\alpha})\right]\\
&= \mathbb{E}(\boldsymbol{\eta}_2) - \boldsymbol{\delta} - \pi\mathbb{E}[\boldsymbol{\eta}_4 I(\chi^2_{q-1}(\Delta) < \chi^2_{q-1,\alpha})] = -\pi\mathbb{E}[\boldsymbol{\eta}_4 I(\chi^2_{q-1}(\Delta) < \chi^2_{q-1,\alpha})]\\
&= -\pi\boldsymbol{\delta}^\star \Phi_{q+1}(\chi^2_{q-1,\alpha}; \Delta).
\end{aligned}$$

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{SS} - \boldsymbol{\beta}_{(n)})\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{UE} - (q-3)\mathfrak{L}_n^{-1}(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}_0 - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{2,n} - \boldsymbol{\delta} - (q-3)\boldsymbol{\eta}_{4,n}\mathfrak{L}_n^{-1}\right] = \mathbb{E}(\boldsymbol{\eta}_2) - \boldsymbol{\delta} - (q-3)\mathbb{E}[\boldsymbol{\eta}_4\chi_{q-1}^{-2}(\Delta)] \\
&= -(q-3)\boldsymbol{\delta}^\star\mathbb{E}[\chi_{q+1}^{-2}(\Delta)].
\end{aligned}$$

$$\begin{aligned}
\mathfrak{B}(\hat{\boldsymbol{\beta}}^{S+}) &= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^{S+} - \boldsymbol{\beta}_{(n)})\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{SS} - \left(1-(q-3)\mathfrak{L}_n^{-1}\right)I(\mathfrak{L}_n < (q-3))(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \frac{1}{\sqrt{n}}\boldsymbol{\delta}\right\}\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\sqrt{n}\left\{\hat{\boldsymbol{\beta}}^{SS} - \boldsymbol{\beta}_{(n)} - \left(1-(q-3)\mathfrak{L}_n^{-1}\right)I(\mathfrak{L}_n < (q-3))(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\right\}\right] \\
&= \mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) - \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{4,n}I(\mathfrak{L}_n < (q-3))\right] + (q-3)\lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\eta}_{4,n}\mathfrak{D}_n^{-1}I(\mathfrak{D}_n < (q-3))\right] \\
&= \mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) - \mathbb{E}\left[\boldsymbol{\eta}_4 I(\chi_{q-1}^2(\Delta) < (q-3))\right] + (q-3)\mathbb{E}\left[\boldsymbol{\eta}_4\chi_{q-1}^{-2}(\Delta)I(\chi_{q-1}^2(\Delta) < (q-3))\right] \\
&= \mathfrak{B}(\hat{\boldsymbol{\beta}}^{SS}) - \boldsymbol{\delta}^\star\Phi_{q+1}(q-3;\Delta) + (q-3)\boldsymbol{\delta}^\star\mathbb{E}\left[\chi_{q+1}^{-2}(\Delta)I(\chi_{q+1}^2(\Delta) < (q-3))\right] \\
&= -\boldsymbol{\delta}^\star\left[\Phi_{q+1}(q-3;\Delta) + (q-3)\mathbb{E}[\chi_{q+1}^{-2}(\Delta)I(\chi_{q+1}^2(\Delta) > (q-3))]\right].
\end{aligned}$$