

Superiority Conditions of Shrinkage Estimator in Laplacian Class of Elliptical Models

M. Arashi and A. Bekker

Department of Statistics, Faculty of Natural and Agricultural Sciences,
University of Pretoria, Pretoria, 0002, South Africa

Email: m_arashi_stat@yahoo.com

Abstract

Shrinkage estimation is a method that a naive or target estimator is improved, in some sense, by combining it with other information. In this paper, we are basically discussing on the performance of a class of Baranchik type shrinkage estimators for the vector location parameter in a location model, when it is assumed that both location and scale parameters are unknown. Since the assumption of normality restricts the range of possible applications, especially in flatter densities, the errors of the location model are further assumed to belong to a sub-class of elliptically contoured distributions, namely Laplacian as an extension to the multivariate normal distribution. Sufficient conditions on dominant class of Baranchik type shrinkage estimators to outperform the usual James-Stein shrinkage estimator are established. It is nicely presented that the dominant properties of the class of estimators are robust truly respect to departures from normality.

Keywords: Elliptically contoured distribution, James-Stein estimator, Jeffreys' prior, Robustness, Schwartz space.

1. Introduction & Some Preliminaries

In this paper, we consider the location model in a more general setup involving dependent errors. Initially let $\mathcal{S}(p)$ denotes the set of all $p \times p$ positive definite matrices. The precise set-up of the problem is as follows: Let \mathbf{Y}_i be an $p \times 1$ response vector with model

$$\mathbf{Y}_i = \boldsymbol{\theta} + \boldsymbol{\epsilon}_i, \quad 1 \leq i \leq N. \quad (1)$$

Here $\boldsymbol{\theta}$ is a $p \times 1$ vector of location parameters and $\boldsymbol{\epsilon}_i$ is a $p \times 1$ error vector such that $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, $Cov(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j) = \boldsymbol{\Sigma} \in \mathcal{S}(p)$, $i, j = 1, \dots, N$, $N > p$. It is assumed, in general, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N)'$ have a joint elliptically contoured

distribution. Typically if it possess a density, it is followed by

$$f(\boldsymbol{\epsilon}|\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{N}{2}} g\left(\text{tr } \boldsymbol{\Sigma} \sum_{i=1}^N \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i'\right), \tag{2}$$

where $g(\cdot)$ is a non-negative function over \mathbb{R}_+ such that $f(\cdot)$ is a density function w.r.t a σ -finite measure μ on \mathbb{R}^p . In this case, notation $\boldsymbol{\epsilon}_i \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}, g)$ would probably be used. Now for the purpose of this study, we represent the *Laplacian class of elliptical models*. Due to Chu (1973), each component of the aforementioned model being proposed in (2), possibly can be presented

$$f_{\boldsymbol{\epsilon}_i}(\mathbf{x}) = \int_0^\infty \mathcal{W}(t) \phi_{\mathcal{N}_p(\mathbf{0}, t^{-1}\boldsymbol{\Sigma})}(\mathbf{x}) dt, \tag{3}$$

where $\phi_{\mathcal{N}_p(\mathbf{0}, t^{-1}\boldsymbol{\Sigma})}(\cdot)$ is the pdf of $\mathcal{N}_p(\mathbf{0}, t^{-1}\boldsymbol{\Sigma})$,

$$\mathcal{W}(t) = (2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} t^{-\frac{p}{2}} \mathcal{L}^{-1}[f(s)], \tag{4}$$

$\mathcal{L}^{-1}[f(s)]$ denotes the inverse Laplace transform of $f(s)$ with $s = t[\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}/2]$.

The mean of $\boldsymbol{\epsilon}_i$ is the zero-vector and the covariance-matrix of $\boldsymbol{\epsilon}_i$ is

$$\int_0^\infty \text{Cov}(\boldsymbol{\epsilon}_i|t) \mathcal{W}(t) dt = \left(\int_0^\infty t^{-1} \mathcal{W}(t) dt \right) \boldsymbol{\Sigma}, \tag{5}$$

provided the above integral exists. Another sub-class of elliptically contoured distributions (ECDs) which includes the above class may be generated by a signed measure \mathcal{W} on the measurable space $(\mathbb{R}^+, \mathbb{B})$ such that the pdf $f(\cdot)$ can be expressed through the following procedures:

$$(i) \quad f(\mathbf{x}) = \int_0^\infty \phi_{\mathcal{N}_p(\mathbf{0}, t^{-1}\boldsymbol{\Sigma})}(\mathbf{x}) \mathcal{W}(dt), \tag{6}$$

$$(ii) \quad \int_0^\infty t^{-1} \mathcal{W}^+(dt) < \infty, \quad (iii) \quad \int_0^\infty t^{-1} \mathcal{W}^-(dt) < \infty,$$

where $\mathcal{W}^+ - \mathcal{W}^-$ is the Jordan decomposition of \mathcal{W} in positive and negative parts (see Srivastava and Bilodeau, 1989). Note that from (ii) – (iii) of (6), $\int_0^\infty t^{-1} \mathcal{W}(dt) < \infty$ and $\text{Cov}(\boldsymbol{\epsilon}_i)$ exists under the sub-class defined above.

Now, under Bayesian framework, suppose $\pi(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \doteq \pi(\boldsymbol{\theta})\pi(\boldsymbol{\Sigma})$. Using the invariant theory, we take $\pi(\boldsymbol{\theta}) \propto \text{constant}$, $\pi(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{p+1}{2}}$.

Lemma1. (Arashi, 2009) Assume in the location model (1), $\boldsymbol{\epsilon}_i \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}, g)$, where $\boldsymbol{\Sigma} \in \mathcal{S}(p)$. Then, the posterior distribution of $\boldsymbol{\theta}$ is multivariate

Student's t distribution, denoted by $\boldsymbol{\theta}|\mathbf{Y} \sim t_p(\bar{\mathbf{Y}}, \mathbf{S}, N - p)$, with the pdf

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \frac{N^{\frac{p}{2}} \Gamma(\frac{N}{2}) |\mathbf{S}|^{-\frac{1}{2}}}{\pi^{\frac{p}{2}} (N-p)^{\frac{N-p}{2}} \Gamma(\frac{N-p}{2})} \left[1 + N(\boldsymbol{\theta} - \bar{\mathbf{Y}})' \mathbf{S}^{-1} (\boldsymbol{\theta} - \bar{\mathbf{Y}}) \right]^{-\frac{N}{2}},$$

where $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_N)$, $\bar{\mathbf{Y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{Y}_i$, $\mathbf{S} = \sum_{i=1}^N (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})'$.

Throughout, for any estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, we shall consider the loss function

$$L(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}) = N(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{7}$$

The Bayes estimator of $\boldsymbol{\theta}$ with respect to the loss (7) is given by $\hat{\boldsymbol{\theta}} = \bar{\mathbf{Y}}$. Also $\hat{\boldsymbol{\theta}} \sim \mathcal{E}_p(\boldsymbol{\theta}, N^{-1}\boldsymbol{\Sigma}, g)$. Under classical viewpoint, we devote a general class of Stein-type shrinkage estimators to the estimator $\hat{\boldsymbol{\theta}}$, given by

$$\boldsymbol{\delta}_r(\hat{\boldsymbol{\theta}}) = \left[1 - \frac{r(\hat{\boldsymbol{\theta}}' \mathbf{S}^{-1} \hat{\boldsymbol{\theta}})}{\hat{\boldsymbol{\theta}}' \mathbf{S}^{-1} \hat{\boldsymbol{\theta}}} \right] \hat{\boldsymbol{\theta}}, \tag{8}$$

where $r : [0, \infty) \rightarrow [0, \infty)$ is an absolutely continuous function.

Furthermore, $r \in \mathcal{S}(\mathbb{R}_+, \mu)$, (the Schwartz space or space of rapidly decreasing functions on \mathbb{R}_+ with the measure μ) where

$$\mathcal{S}(\mathbb{R}_+, \mu) = \{r \in \mathcal{C}^\infty(\mathbb{R}_+, \mu) : \|r\|_{\alpha, \beta} < \infty \quad \forall \alpha, \beta\},$$

α and β are indices, $\mathcal{C}^\infty(\mathbb{R}_+, \mu)$ is the set of all smooth functions from \mathbb{R}_+ to \mathbf{C} (the set of all complex numbers) and $\|r\|_{\alpha, \beta} = \|x^\alpha \mathbf{D}^\beta r\|_\infty = \sup\{|x^\alpha \mathbf{D}^\beta r(x)| : x \in \text{domain of } r\}$. Here \mathbf{D}^β stands for β^{th} derivative of r . Note that for every function such as $r(\cdot)$ belongs to $\mathcal{S}(\mathbb{R}_+, \mu)$, we have

$$\int_0^\infty r'(x) d\mu(x) < \infty, \quad \int_0^\infty r^2(x) d\mu(x) < \infty, \tag{9}$$

More interesting that the Schwartz space is dense in the space of all functions satisfy the above conditions in (9).

Lemma2. If $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\theta}, \alpha \boldsymbol{\Sigma})$, $\alpha > 0$, $\boldsymbol{\Sigma} \in \mathcal{S}(p)$ is independent of $\mathbf{S} \sim W_p(\beta \boldsymbol{\Sigma}, n)$, $\beta > 0$, $n = N - 1$, then

$$E \left[\frac{\mathbf{x}' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\theta}) r(\mathbf{x}' \mathbf{S}^{-1} \mathbf{x})}{\mathbf{x}' \mathbf{S}^{-1} \mathbf{x}} \right] = \beta \alpha (n - p + 1) \left\{ (p - 2) E \left[\frac{r(\mathbf{x}' \mathbf{S}^{-1} \mathbf{x})}{\mathbf{x}' \boldsymbol{\Sigma}^{-1} \mathbf{x}} \right] + 2 E [r'(\mathbf{x}' \mathbf{S}^{-1} \mathbf{x})] \right\}$$

$$E \left[\frac{\mathbf{x}' \boldsymbol{\Sigma}^{-1} \mathbf{x} r^2(\mathbf{x}' \mathbf{S}^{-1} \mathbf{x})}{(\mathbf{x}' \mathbf{S}^{-1} \mathbf{x})^2} \right] = \beta^2 (n - p + 1)(n - p + 3) E \left[\frac{r^2(\mathbf{x}' \mathbf{S}^{-1} \mathbf{x})}{\mathbf{x}' \boldsymbol{\Sigma}^{-1} \mathbf{x}} \right]$$

Lemma3. The risk function of the estimator $\delta_r(\hat{\theta})$ w.r.t. the loss function (7) is give by

$$R(\hat{\theta}; \theta) - 4(N - p) \int_0^\infty E \left[r' \left(\hat{\theta}' \mathbf{S}^{-1} \hat{\theta} \right) \middle| t \right] t^{-2} \mathcal{W}(dt) + \int_0^\infty E \left\{ \frac{(N - p)r \left(\hat{\theta}' \mathbf{S}^{-1} \hat{\theta} \right)}{t \hat{\theta}' (t^{-1} \boldsymbol{\Sigma})^{-1} \hat{\theta}} \right. \\ \left. \times \left[N(N - p + 2)r \left(\hat{\theta}' \mathbf{S}^{-1} \hat{\theta} \right) - 2(p - 2) \right] \middle| t \right\} \mathcal{W}(dt)$$

2. Domination Necessary Conditions

In this section, we develop necessary conditions for which the shrinkage estimator $\delta_r(\hat{\theta})$ dominates the James-Stein type estimator given by

$$\delta_{JS}(\hat{\theta}) = \left[1 - \frac{p - 2}{\hat{\theta}' \mathbf{S}^{-1} \hat{\theta}} \right] \hat{\theta}. \tag{10}$$

The performance of this estimator is discussed in Srivastava and Bilodeau (1989) extensively.

Theorem1. Assume that the function $r(\cdot)$ is bounded and absolutely continuous. Necessary conditions for an estimator $\delta_r(\hat{\theta})$ to dominate $\delta_{JS}(\hat{\theta})$ are that

- (i) for every ω , there exists $\omega_0 (> \omega)$ such that $r'(\omega_0) \geq 0$,
- (ii) if $\omega r'(\omega)$ has a limiting value as ω approaches infinity, it must be 0,
- (iii) if $r(\omega)$ has a limiting value as ω approaches infinity and $\omega r'(\omega)$ converges to 0 as ω approaches infinity, the limit value for $r(\omega)$ must be $\frac{p-2}{N(N-p+2)}$.

Proof: Proof of (i) directly follows from the proof of Corollary 2.1. of Maruyama and Strawderman (2005). Now consider using Lemma 3, one can directly obtain

$$\Delta = R(\delta_{JS}(\hat{\theta}); \theta) - R(\delta_r(\hat{\theta}); \theta) = 4(N - p) \int_0^\infty E \left[r' \left(\hat{\theta}' \mathbf{S}^{-1} \hat{\theta} \right) \middle| t \right] t^{-2} \mathcal{W}(dt) \\ + \int_0^\infty E \left\{ \frac{(N - p)(p - 2)}{t \hat{\theta}' (t^{-1} \boldsymbol{\Sigma})^{-1} \hat{\theta}} \left[N(N - p + 2)(p - 2) - 2(p - 2) \right] \middle| t \right\} \mathcal{W}(dt) \\ - \int_0^\infty E \left\{ \frac{(N - p)r \left(\hat{\theta}' \mathbf{S}^{-1} \hat{\theta} \right)}{t \hat{\theta}' (t^{-1} \boldsymbol{\Sigma})^{-1} \hat{\theta}} \left[N(N - p + 2)r \left(\hat{\theta}' \mathbf{S}^{-1} \hat{\theta} \right) - 2(p - 2) \right] \middle| t \right\} \mathcal{W}(dt) \\ = 4(N - p) \int_0^\infty E \left[r' \left(\mathbf{z}' \mathbf{B}^{-1} \mathbf{z} \right) \middle| t \right] t^{-2} \mathcal{W}(dt) \\ - (N - p) \int_0^\infty E \left[- \frac{\{r \left(\mathbf{z}' \mathbf{B}^{-1} \mathbf{z} \right) - (p - 2)\}^2}{\mathbf{z}' \mathbf{B}^{-1} \mathbf{z}} \frac{\mathbf{z}' \mathbf{B}^{-1} \mathbf{z}}{\mathbf{z}' \mathbf{z}} \middle| t \right] t^{-2} \mathcal{W}(dt) \\ = (N - p) \int_0^\infty E \left[\mathcal{G}_r \left(\mathbf{z}' \mathbf{B}^{-1} \mathbf{z} \right) \middle| t \right] t^{-2} \mathcal{W}(dt),$$

where $\mathbf{z} = \boldsymbol{\Sigma}^{-\frac{1}{2}} \bar{\mathbf{Y}}$, $\mathbf{B} = \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{S} \boldsymbol{\Sigma}^{-\frac{1}{2}}$ and

$$\mathcal{G}_r(\omega) = - \frac{[r(\omega) - (p - 2)]^2}{(n - p - 1)\omega} + 4r'(\omega). \tag{11}$$

For the proofs of (ii) and (iii), using the proofs of (ii) and (iii) of Corollary 2.1. of Maruyama and Strawderman (2005), it is enough to show that if $\delta_r(\hat{\theta})$ dominates $\delta_{JS}(\hat{\theta})$, then, for every ω , there exists $\omega_0(> \omega)$ such that $\mathcal{G}_r(\omega_0) \geq 0$. In this case we follow the proof of Theorem 2.1. of Maruyama and Strawderman (2005). Suppose to the contrary that there exists ω_0 such that $\mathcal{G}_r(\omega) < 0$ for any $\omega \geq \omega_0$. Under the boundedness of $\mathcal{G}_r(\cdot)$, there exists an $M(> 0)$ such that $\mathcal{G}_r(\omega) \leq M$ for any ω . Under the assumption of absolute continuity of $\mathcal{G}_r(\cdot)$ there exists two points $(\omega_0 <)\omega_1 < \omega_2$ and $\epsilon(> 0)$ such that $\mathcal{G}_r(\omega) < -\epsilon$ on $\omega \in [\omega_1, \omega_2]$. Using M and ϵ , we define $\mathcal{G}_{r,\epsilon}(\omega)$ as

$$\mathcal{G}_{r,\epsilon}(\omega) = \begin{cases} M & \omega \leq \omega_0 \\ 0 & \omega_0 < \omega < \omega_1 \\ -\epsilon & \omega_1 \leq \omega \leq \omega_2 \\ 0 & \omega > \omega_2 \end{cases}$$

The inequality $\mathcal{G}_{r,\epsilon}(\omega) \geq \mathcal{G}_r(\omega)$ for any ω and using equation (11) imply

$$\begin{aligned} \Delta &\leq M(N-p) \int_0^\infty P_\theta \left(W \leq \omega_0 \mid t \right) t^{-2} \mathcal{W}(dt) \\ &\quad - \epsilon(N-p) \int_0^\infty P_\theta \left(\omega_1 \leq W \leq \omega_2 \mid t \right) t^{-2} \mathcal{W}(dt), \end{aligned}$$

where $W = \|\mathbf{X}\|^2$ for $\mathbf{X} = (t^{-1}\Sigma)^{-\frac{1}{2}}\hat{\theta}$.

Based on the properties of the model under study, it can be realized that

$$\begin{aligned} \int_0^\infty P_\theta \left(W \leq \omega_0 \mid t \right) t^{-2} \mathcal{W}(dt) &= \int_0^\infty P_\theta \left(W \leq \omega_0 \mid t \right) t^{-2} \mathcal{W}^+(dt) \\ &\quad - \int_0^\infty P_\theta \left(W \leq \omega_0 \mid t \right) t^{-2} \mathcal{W}^-(dt) \geq 0. \end{aligned}$$

This phenomenon is also valid for $\int_0^\infty P_\theta \left(\omega_1 \leq W \leq \omega_2 \mid t \right) t^{-2} \mathcal{W}(dt)$.

Now let \mathbf{a} be a fixed p -dimensional unit vector (see Fig. 1). Then the half plane $\{\mathbf{x} : \mathbf{a}'\mathbf{x} \leq \sqrt{\omega_0}\}$ includes the p -dimensional hyper-ellipsoid $\{\mathbf{x} : \|\mathbf{x}\|^2 \leq \omega_0\}$. For $\boldsymbol{\theta} = (\sqrt{\omega_0} + \lambda)(t^{-1}\Sigma)^{-\frac{1}{2}}\mathbf{a}$, we have

$$\begin{aligned} P_\theta \left(W \leq \omega_0 \mid t \right) &< \int_{\mathbf{a}'\mathbf{x} \leq \sqrt{\omega_0}} \frac{|t^{-1}\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\theta}\|^2}{2}\right) d\mathbf{x} \\ &\leq \exp(\lambda\sqrt{\omega_0}) \exp\left(-\frac{\|\boldsymbol{\theta}\|^2}{2}\right) \times \int_{\mathbf{a}'\mathbf{x} \leq \sqrt{\omega_0}} \frac{|t^{-1}\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2} + \sqrt{\omega_0}\mathbf{a}'\mathbf{x}\right) d\mathbf{x} \\ &\leq \exp(\lambda\sqrt{\omega_0}) \exp\left(-\frac{\|\boldsymbol{\theta}\|^2}{2} + \frac{\omega_0}{2}\right). \end{aligned}$$

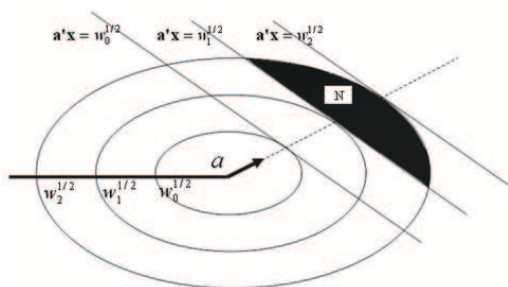


Figure 1: Graph of half planes

For $N = \{\mathbf{x} : \omega_1 \leq \|\mathbf{x}\|^2 \leq \omega_2, \sqrt{\omega_1} \leq \mathbf{a}'\mathbf{x} \leq \sqrt{\omega_2}\}$ and $\boldsymbol{\theta} = (\sqrt{\omega_0} + \lambda)(t^{-1}\boldsymbol{\Sigma})^{-\frac{1}{2}}\mathbf{a}$,

$$P_{\boldsymbol{\theta}} \left(\omega_1 \leq W \leq \omega_2 \mid t \right) > \int_N \frac{|t^{-1}\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \exp \left(-\frac{\|\mathbf{x} - \boldsymbol{\theta}\|^2}{2} \right) d\mathbf{x} \\ \geq \exp(\lambda\sqrt{\omega_1}) \exp \left(-\frac{\|\boldsymbol{\theta}\|^2}{2} \right) \times \int_N \frac{|t^{-1}\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \exp \left(-\frac{\|\mathbf{x}\|^2}{2} + \sqrt{\omega_0}\mathbf{a}'\mathbf{x} \right) d\mathbf{x}.$$

By making use of the above equations, we can obtain

$$\Delta \leq (N - p) \int_0^\infty c_1 \exp \left(\sqrt{\omega_0}\lambda - \frac{\|\boldsymbol{\theta}\|^2}{2} \right) \left(1 - c_2 \exp \left[(\sqrt{\omega_1} - \sqrt{\omega_0})\lambda \right] \right) t^{-2} \mathcal{W}(dt),$$

where $c_1 = M \exp \left(\frac{\omega_0}{2} \right)$ and $c_2 = \frac{\epsilon}{M} \exp \left(\frac{\omega_0}{2} \right) \int_N \frac{|t^{-1}\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \exp \left(-\frac{\|\mathbf{x}\|^2}{2} + \sqrt{\omega_0}\mathbf{a}'\mathbf{x} \right) d\mathbf{x}$.

Since c_1 and c_2 do not depend on λ , Δ is negative for sufficiently large λ . This completes the proof. ■

Acknowledgements

This work is based upon research supported by the University of Pretoria Vice-chancellor’s post-doctoral fellowship programme.

References

Arashi, M. (2009). Constrained Bayes estimator in Elliptical Models, *Proc. 57th Sess. Int. Statist. Inst.*, Durban, South Africa.
 Chu, K. C., (1973). Estimation and decision for linear systems with elliptically random process. *IEEE Trans. Autom. Cont.*, 18, 499-505.
 Maruyama, Y. and Strawderman, W. E. (2005). Necessary conditions for dominating the James-Stein estimator, *Ann. Inst. Statist. Math.*, 57, 157-165.
 Srivastava, M. and Bilodeau, M., (1989). Stein estimation under elliptical distribution, *J. Mult. Annal.*, 28, 247-259.