Sinopfbd, 2024; 9(2): 551-571 https://doi.org/10.33484/sinopfbd.1485489

Sinop Üniversitesi Fen Bilimleri Dergisi
*Sinop Uni J Nat Sci*

E-ISSN: 2564-7873

https://dergipark.org.tr/tr/pub/sinopfbd

# The Impact of Prior Based Loss Function For Elliptical Regression Models

**Mohammad ARASHI[1], Fatma Sevinç KURNAZ[2] and Naushad MAMODEKHAN[3]**

How to cite: Arashi, M., Kurnaz, F. S., & MamodeKhan, N. (2024). The impact of prior based loss function for elliptical regression models. *Sinop Üniversitesi Fen Bilimleri Dergisi*, *9*(2), 551-571. https://doi.org/10.33484/sinopfbd.1485489

**Research Article**

**Corresponding Author**
Fatma Sevinç KURNAZ
fskurnaz@yildiz.edu.tr

**ORCID of the Authors**
M.A: 0000-0002-5881-9241
F.S.K: 0000-0002-5958-7366
N.M: 0000-0002-9125-7121

**Received:** 16.05.2024
**Accepted:** 27.12.2024

## Abstract

In the paper we consider a multiple regression model with elliptically contoured errors. In the Bayesian view, a prior information is taken for the weight under a prior based balanced-type loss function in order to avoid making redundant assumptions. This is the essence of the Bayesian inference with vague prior information in regression analysis. It directly impacts on the performance of the quasi empirical Bayesian shrinkage estimators through the inclusion of a reciprocal weight related to the dimension of parameter space. The shrinkage factor of the estimator is also robust to outliers and the unknown density generator of elliptical models. Finally, this result is supported by an application.

**Keywords:** Balanced loss function, Elliptically contoured distribution, Prior based loss function, Quasi-empirical Bayes estimator, Shrinkage estimator

# Eliptik Regresyon Modelleri Için Kayıp Fonksiyona Dayalı Önselliğin Etkisi

[1]Ferdowsi University of Mashhad, Department of Statistics, Mashhad, Iran

[2]Yildiz Technical University, Faculty of Mathematical Sciences and Arts, Istanbul, Türkiye

[3]University of Mauritius, Faculty of Mathematical Sciences, Mauritius



## Öz

Bu çalışmada eliptik konturlu hatalara sahip olan çoklu regresyon modeli ele alınmıştır. Bayesyen bakıştan, gereksiz varsayımlarda bulunmaktan kaçınmak için, ağırlık için dengeli tipteki kayıp fonksiyonuna dayalı bir önsellik altında önsellik bilgisi gözönüne alınmıştır. Bu, regresyon analizindeki muğlak öncelik bilgisiyle Bayesyen çıkarımın özüdür. Parametre uzayının boyutuyla ilgili karşılıklı ağırlığın dahil edilmesi yoluyla yarı ampirik Bayesyen büzücü tahmincilerin performansı bundan etkilenir. Tahmin edicinin büzülme faktörünün, verilerdeki aykırı değerlere ve eliptik modellerin bilinmeyen yoğunluk yaratıcı fonksiyonuna karşı dayanıklı olduğu gösterilmiştir. Son olarak, bir uygulama ile desteklenmiştir.

**Anahtar Kelimeler:** Dengeli kayıp fonksiyonu, Eliptik konturlu dağılım, Kayıp fonksiyona dayalı önsellik, Yarı-ampirik Bayes tahmin edici, Büzücü tahmin edici

**Introduction**

Multiple regression model is commonly used statistical tool applied in many disciplines of the modern area. The estimation of parameters of the multiple regression model is a common interest in many studies. Let us consider the following multiple regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where

- $\boldsymbol{y}$ is an $n \times 1$ response vector, representing the observed outcomes or dependent variable.
- $\boldsymbol{X}$ is an $n \times p$ design matrix, which contains the independent variables or predictors. It is assumed to be non-stochastic (i.e., not random) and of full rank $p$ with $n > p$ (meaning the number of observations is greater than the number of predictors).
- $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ is a $p \times 1$ vector of unknown regression coefficients.
- $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ is the $n \times 1$ error vector, representing the random disturbances or noise in the model. The errors are assumed to follow an elliptically contoured distribution (ECD), specifically $\boldsymbol{\epsilon} \sim \mathcal{E}_n(\mathbf{0}, \sigma^2\mathbf{V}, g_n)$, where: $\sigma^2$ is the variance of the error; $\mathbf{V} \in S(n)$ is a known, positive definite matrix (i.e., all its eigenvalues are positive) that defines the covariance structure of the errors; $g_n$ is the density generator function, which specifies the shape of the distribution.

Precisely the density of $\boldsymbol{\epsilon}$ is given by

$$f(\boldsymbol{\epsilon}) = \phi|\sigma^2\mathbf{V}|^{-\frac{1}{2}} g_n\left[\frac{1}{2\sigma^2}\boldsymbol{\epsilon}'\mathbf{V}^{-1}\boldsymbol{\epsilon}\right], \tag{2}$$

where $\phi$ is the normalizing constant given by

$$\phi^{-1} = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_{\mathbb{R}+} y^{\frac{n}{2}-1} g_n(y)\mathrm{d}y \tag{3}$$

for some density generator function $g_n(.)$. Here, $\Gamma(.)$ is the Gamma function, which generalizes the factorial function to continuous values. The integral ensures that the distribution is properly normalized. The condition in [1]

$$\int_0^\infty x^{\frac{n}{2}-1} g_n(x)\mathrm{d}x < \infty \tag{4}$$

guarantees that $g_n(x)$ is a density generator. If the function $g_n(.)$ does not depend on $n$, we use the notation $g$ instead of $g_n(x)$.

Lower risk improves the performance of the estimator. In such a case, the loss function plays as a decision mechanism. It is quite significant to work with reasonable and practical losses in order to improve the performance of the estimator.

Let $\boldsymbol{\beta}^*$ be any estimator of $\boldsymbol{\beta}$, then the quadratic loss function is described by $(\boldsymbol{X}\boldsymbol{\beta}^* - \boldsymbol{y})'(\boldsymbol{X}\boldsymbol{\beta}^* - \boldsymbol{y})$, which shows the goodness of fit. Then, the precision of the estimator $\boldsymbol{\beta}^*$ is computed using the weighted loss function $(\boldsymbol{\beta}^* - \boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta})$. In order to show the performances of an estimator, both of the above criteria are taken into consideration. In this paper, we take into consideration the issue of the estimator with following *balanced loss function* (BLF)

$$L^{\boldsymbol{W}}_{\omega,\boldsymbol{\delta}_0}(\boldsymbol{\beta}^*;\boldsymbol{\beta}) = \omega r\left(\|\boldsymbol{\beta}\|^2\right)(\boldsymbol{\beta}^* - \boldsymbol{\delta}_0)'\boldsymbol{W}(\boldsymbol{\beta}^* - \boldsymbol{\delta}_0) + (1-\omega) r\left(\|\boldsymbol{\beta}\|^2\right)(\boldsymbol{\beta}^* - \boldsymbol{\beta})'\boldsymbol{W}(\boldsymbol{\beta}^* - \boldsymbol{\beta}), \quad (5)$$

where the weight shows with $\omega \in [0,1]$, $\boldsymbol{W}$ shows the matrix including weights, a pd weight function is $r(.)$ and $\boldsymbol{\delta}_0$ is the estimation of unknown vector $\boldsymbol{\beta}$. This loss function was proposed by Jozani [2] based on Zellner's balanced loss function [3]. Gómez-Déniz called it as weighted balanced loss function (WBLF) and generalized that idea to the credibility theory [4]. Both measures for goodness of fit and estimator error are taken using the WBLF. The first term in Eq. (5), $\omega r\left(\|\boldsymbol{\beta}\|^2\right)(\boldsymbol{\beta}^* - \boldsymbol{\delta}_0)'(\boldsymbol{\beta}^* - \boldsymbol{\delta}_0)$, is similar to the penalty term for being without smoothness in non-parametric regression. The weight $\omega$ in Eq. (5) calibrates the relative importance of these two measures. Dey et al. [5] worked the issues of admissibility and dominance under the loss function in Eq. (5) with $r = 1$ and $\boldsymbol{W} = \boldsymbol{I}_p$. For $\omega = 0$, $L^{\boldsymbol{W}}_0(\boldsymbol{\delta};\boldsymbol{\beta})$ shows the quadratic loss function. Bayesian context is used to explain the the weight function $r(.)$. To the best of our knowledge, the studies that use vague prior information (roughly speaking, objective Bayesian) for the parameter space seem to provide no more than a classical approach. A possible idea to solve this issue is to consider a prior information for the weight function in the loss function in Eq. (5). This approach is different from the recent study of Evans and Jang [6], where they derived the least relative surprise Bayes estimator using a prior based loss function.

The main focus of this study is to improve objective Bayesian inference considering a prior based loss function. We estimate the unknown regression coefficients $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ under the condition that we may not know whether $\boldsymbol{\beta}$ belongs to the subspace defined by $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$. This implies the presence of certain restrictions on the parameter space, which may influence the estimation process. Specifically, the matrix $\boldsymbol{H}$ shows a $q \times p$ matrix of constants, and $\boldsymbol{h}$ is a $q$-dimensional vector containing known constants. These restrictions imply that certain linear combinations of the regression coefficients $\boldsymbol{\beta}$ are fixed or constrained, and this affects how the coefficients are estimated.

The study pays special attention to the use of Stein-type shrinkage estimators and the preliminary test (PT) estimator of $\boldsymbol{\beta}$. The key restriction here is that the prior information regarding the weights in the balanced loss function (Eq. (3)) directly influences the risk associated with these estimators. By incorporating prior knowledge about the relationship between $\boldsymbol{\beta}$ and $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, the shrinkage estimator is able to reduce the risk, but the extent of this reduction depends on both the dimensionality of the parameter vector and the specific choice of prior models. Saleh [7] provides an overview for this issue under normal and non-parametric theory covering many standard models. For more information, see [8–19].

Our contribution has the following specific highlights:

1. We propose shrinkage estimators with robust performance concerning.
2. We introduce a flexible prior-based balanced loss function that can calibrate variability and risk.
3. We propose a class of minimax estimators under the balanced loss function.
4. We proposed the preliminary and Stein-type estimators for the class of elliptically contoured

distribution.

Apart from the highlights mentioned above, as an advantage of the proposed shrinkage estimators, they outperform the Bayes estimator using the BLF, under some mild conditions.
The summary of this paper is here: In the next section we take the vague prior on the entire parameter space $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}^+$, in order to show that the results are not unrelated to the Bayesian analysis, and then introduce the prior information that plays a key role for enabling a real Bayesian analysis. The following section introduces various estimators. The bias function and the risk function for the estimators are obtained in the next section. Then the performances of the five estimators are compared theoretically and then supported the results with a numerical example Last section presents out conclusions and remarks.

## The Bayesian Setup

Under standard assumptions, the least squares (LS) estimator of $\boldsymbol{\beta}$ is

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{y} = \boldsymbol{C}^{-1}\mathbf{X}'\boldsymbol{V}^{-1}\boldsymbol{y}, \text{ where } \boldsymbol{C} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}. \tag{6}$$

Similarly the LS estimator of $\sigma^2$ is

$$\tilde{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \tag{7}$$

It is straightforward to show that

$$S^2 = \frac{1}{n-p}\left[(\boldsymbol{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})\right] \tag{8}$$

is an unbiased estimator of the true error variance $\sigma^2_{\boldsymbol{\epsilon}}$. Under the assumption of elliptical errors, the true error variance $\sigma^2_{\boldsymbol{\epsilon}}$ is related to the characteristic generator $\psi'(0)$ of the elliptical distribution:

$$\sigma^2_{\boldsymbol{\epsilon}} = -2\psi'(0)\sigma^2 \tag{9}$$

where $\psi'(0)$ denotes the first derivative of the characteristic generator of the elliptical distribution at zero. Thus, the estimator $S^2$, derived from the residuals weighted by $\mathbf{V}^{-1}$, provides an unbiased estimate of the error variance $\sigma^2_{\boldsymbol{\epsilon}}$, which is consistent with the true variance structure of the model. See Chapt. 4 of [1] for more details.
From the Bayesian perspective, firstly, it is assumed to be a little about the parameters and further the elements of $\boldsymbol{\beta}$ are independent of $\sigma^2$. Therefore, the joint prior distribution has form

$$\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta})\pi(\sigma^2), \tag{10}$$

where $\pi(\cdot)$ is a prior density. At this stage, it is important to consider invariant theory, which ensures that the accuracy of the models is preserved regardless of linear transformations applied to the dataset. Using the invariant theory as in [20], we take the prior knowledge about the parameter space as follows:

$$\pi(\boldsymbol{\beta}) \propto \text{constant}, \quad \pi(\sigma^2) \propto \sigma^{-2} \text{ (that is } \pi(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}), \tag{11}$$

where $\boldsymbol{\beta}$ and $\sigma^2$ are regression model parameters. Additionally, invariant theory enhances the reliability of estimators by ensuring that the error variance remains unchanged under transformations, particularly in special models such as elliptical distributions.

**Lemma 1.** (Arashi, 2010) Assume that $\boldsymbol{\epsilon} \sim \mathcal{E}_n(\mathbf{0}, \sigma^2\boldsymbol{V}, \boldsymbol{g})$, where $\boldsymbol{V} \in S(n)$ in the *multiple regression* model (1). Then, with respect to the prior distribution defined by (11), the posterior distribution of $\boldsymbol{\beta}$ follows a multivariate Student's t-distribution, denoted as $\boldsymbol{\beta}|(\mathbf{X}, \boldsymbol{y}) \sim t_p(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}, m)$, where $\boldsymbol{\Sigma} = S^2\boldsymbol{C}^{-1}$, with the following density

$$f(\boldsymbol{\beta}|\mathbf{X}, \boldsymbol{y}) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{C(m,p)\pi^{\frac{p}{2}}} \left[1 + \frac{1}{m}\,(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right]^{-\frac{n}{2}}, \tag{12}$$

where the normalizing constant $C(m,p)$ is given by

$$C(m,p) = \frac{m^{\frac{p}{2}}\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \text{ with } \quad m = n - p. \tag{13}$$

Using Lemma 2 of Jozani et al. [2], the Bayes estimator under the BLF in Eq. (5) is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}^B &= \omega\boldsymbol{\delta}_0 + (1-\omega)\frac{E_\pi\left[\boldsymbol{\beta} r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]}{E_\pi\left[r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]} \\ &= \omega\tilde{\boldsymbol{\beta}} + (1-\omega)\frac{E_\pi\left[\boldsymbol{\beta} r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]}{E_\pi\left[r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]}, \end{aligned} \tag{14}$$

where $\boldsymbol{\delta}_0$ refers to any estimator of $\boldsymbol{\beta}$ according to [2]. $\omega$ is a weight factor between 0 and 1 that controls the balance between the prior and posterior components. $E_\pi\left[.|(\boldsymbol{X}, \boldsymbol{y})\right]$ presents the posterior expectation given the data. $r\left(\|\boldsymbol{\beta}\|^2\right)$ is a function of the squared norm of $\boldsymbol{\beta}$. Let $\tilde{\boldsymbol{\beta}}$ be the target estimator. Here the LS estimator of $\boldsymbol{\beta}$ as the target estimator, and $\omega$, $r$, and $\delta_0$ are defined in Eq. (5). When $r\left(\|\boldsymbol{\beta}\|^2\right) = 1$, the Bayes estimator reduces to $\hat{\boldsymbol{\beta}}^B = \tilde{\boldsymbol{\beta}}$, which is similar to considering the Bayes estimator under the quadratic error loss (QEL) function. In this case, since the Bayes estimator is nothing more than the classical LS estimator of $\beta$, one may ask ***what would be the benefit of putting prior on the model?*** As a response, we suggest to take a prior based loss function and consider the role of the prior distribution in the loss function in the form of $r(\|\boldsymbol{\beta}\|^2)$.
For the ECDs, let us take the form of the $r(.)$ function as follows

$$r(\|\boldsymbol{\beta}\|^2) = g_p(\|\boldsymbol{\beta}\|^2), \tag{15}$$

where $g_p$ is defined in Eq. (2). Under the above assumption, the loss function relates to the density generator of the base model and therefore the prior information has direct impact on the model understudy. One should note that $r(.)$ can be independent of the function $g(.)$.
In order to recompute the Bayes estimator, we need to use invariant theory, which refers to computing expectations in two steps: first, calculating the conditional expectation of a quantity given the data

$(\boldsymbol{X}, \boldsymbol{y})$, and then taking the expectation of that conditional expectation with respect to the posterior distribution of the parameters $\boldsymbol{\beta}$. This process ensures that the estimator accounts for both the prior knowledge (through the prior distribution) and the data (through the posterior distribution).
Therefore, using double expectation property and Lemma 1, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^B &= \omega\tilde{\boldsymbol{\beta}} + (1-\omega)\frac{E_\pi\left[\boldsymbol{\beta} r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]}{E_\pi\left[r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]} \\
&= \omega\tilde{\boldsymbol{\beta}} + (1-\omega)\frac{E_g\left\{E_\pi\left[\boldsymbol{\beta} r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]\Big| r\left(\|\boldsymbol{\beta}\|^2\right)\right\}}{E_g\left\{E_\pi\left[r\left(\|\boldsymbol{\beta}\|^2\right)|(\boldsymbol{X}, \boldsymbol{y})\right]\Big| r\left(\|\boldsymbol{\beta}\|^2\right)\right\}} \\
&= \omega\tilde{\boldsymbol{\beta}} + (1-\omega)\frac{\phi^{-1}\tilde{\boldsymbol{\beta}}}{\phi^{-1}} = \tilde{\boldsymbol{\beta}}, \qquad (16)
\end{aligned}$$

where we computed $E_g\left\{r(\|\boldsymbol{\beta}\|^2)\right\}$ by making use of (11) for fixed constant (to 1) as follows

$$\begin{aligned}
E_g\left\{r(\|\boldsymbol{\beta}\|^2)\right\} &= \int_{\mathbb{R}^p} g_p(\|\boldsymbol{\beta}\|^2)\mathrm{d}\boldsymbol{\beta} \\
&= \frac{\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)}\int_{\mathbb{R}^+} y^{\frac{p}{2}-1} g_p(y)\mathrm{d}y \\
&= \phi^{-1}, \qquad (17)
\end{aligned}$$

where $\phi$ corresponds to the normalizing constant as in (2) when $n$ is replaced by $p$.
Note that the Bayes estimator we obtained with the prior based loss function may also be obtained by using the QEL function. The advantage of using the risk function relative to the prior based loss function will be benefited from the effect of prior in Eq. (17). In other words, the prior based BLF enables us to evaluate the risk functions based on this loss, whereas the Bayes estimator is the QEL derived in reality.

**Shrinkage Estimators**

For testing $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$ (where $q < p$) against $H_a : \boldsymbol{H\beta} \neq \boldsymbol{h}$, let us consider the restricted estimator (RE) under $H_0$ designated $\hat{\boldsymbol{\beta}}$ given by

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^B - \boldsymbol{C}^{-1}\boldsymbol{H}'\boldsymbol{V}_1(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h}), \qquad (18)$$

where $\boldsymbol{V}_1 = [\boldsymbol{HC}^{-1}\boldsymbol{H}']^{-1}$, $\boldsymbol{H}$ is a $q \times p$ constant matrix of row rank $q(q \leq p)$ and $\boldsymbol{h}$ is a $q$-vector of pre-specified values. $\boldsymbol{C} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$, $\tilde{\boldsymbol{\beta}}$ is the LS estimator in (6), and $\hat{\boldsymbol{\beta}}^B$ is the Bayes estimator.
From the definition of the elliptical model in Eq. (1), $\hat{\boldsymbol{\beta}} \sim \mathcal{E}_p\left(\boldsymbol{\beta} - \boldsymbol{\Delta}, \sigma^2\boldsymbol{V}_2, \boldsymbol{g}\right)$ for $\boldsymbol{\Delta} = \boldsymbol{C}^{-1}\boldsymbol{H}'\boldsymbol{V}_1(\boldsymbol{H\beta} - \boldsymbol{h})$ and $\boldsymbol{V}_2 = \boldsymbol{C}^{-1}(\boldsymbol{I}_p - \boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{HC}^{-1})$. Similarly, the following estimator is unbiased for $\sigma^2_{\boldsymbol{\epsilon}}$ under $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$,

$$S^{*2} = \frac{1}{n-p+q}\left[(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\right]. \qquad (19)$$

In this part, we calculate the test statistic for testing the linear null hypothesis $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$.

Direct computations using Corollary 1 [21], gives the likelihood ratio test statistic as follows

$$\mathcal{L}_n = \frac{(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})'\boldsymbol{V}_1(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})}{qS^2}. \tag{20}$$

where $\tilde{\boldsymbol{\beta}}$ and $S^2$ are given respectively by Eq. (6), Eq. (8). $\boldsymbol{V}_1 = [\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}']^{-1}$, $\boldsymbol{H}$ is a $q \times p$ constant matrix of row rank $q(q \leq p)$ and $\boldsymbol{C} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$.
Under the hypothesis $H_0$, the pdf of the likelihood ratio test statistic $\mathcal{L}_n$ is given by

$$\boldsymbol{g}^*_{q,m}(\mathcal{L}_n) = \frac{\left(\frac{q}{m}\right)^{\frac{q}{2}} \mathcal{L}_n^{\frac{q}{2}-1}}{B\left(\frac{q}{2}, \frac{m}{2}\right)\left(1 + \frac{q}{m}\mathcal{L}_n\right)^{\frac{1}{2}(q+m)}} \tag{21}$$

where $m = n - p$, $B(\cdot, \cdot)$ is the beta function and $q$ is the rank of $\boldsymbol{H}$ and dimension of vector of constants $\boldsymbol{h}$ in our restriction. Eq. (21) shows that $\mathcal{L}_n$ follows the central F-distribution with $(q, m)$ degrees of freedom (df).
Based on that model, $\boldsymbol{\beta}$ may not belong to the subspace defined by $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$. In such a situation one combines the estimators of $\boldsymbol{\beta}$ and the test-statistic in order to obtain shrinkage estimators as in [7]. Firstly, we consider the PT estimator as a convex combination of the LS estimator $\tilde{\boldsymbol{\beta}}$ (designated as unrestricted estimator - UE) and $\hat{\boldsymbol{\beta}}$ (designated as RE) which are defined by Eq. (18) as follows:

$$\hat{\boldsymbol{\beta}}^{PT} = \tilde{\boldsymbol{\beta}} I(\mathcal{L}_n \geq F_\alpha) + \hat{\boldsymbol{\beta}} I(\mathcal{L}_n < F_\alpha), \tag{22}$$

where $I(A)$ is an indicator function of the set $A$ and $F_\alpha$ is the upper $\alpha^{th}$ percentile of the central F-distribution with $(q, m)$ df. The PT estimator has disadvantage since it is defined with $\alpha$ $(0 < \alpha < 1)$, the significance level. In addition to this, it yields the extreme results. Therefore, $\hat{\boldsymbol{\beta}}$ or $\tilde{\boldsymbol{\beta}}$ depends on the outcome of the test. To overcome this problem, we suggest a Stein-type shrinkage (SS) estimator of $\boldsymbol{\beta}$, as

$$\hat{\boldsymbol{\beta}}^{S} = \hat{\boldsymbol{\beta}} + (1 - d\mathcal{L}_n^{-1})(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) = \tilde{\boldsymbol{\beta}} - d\mathcal{L}_n^{-1}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}), \tag{23}$$

where $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are LS and RE estimators given by Eq. (6) and Eq. (18), respectively, and $\mathcal{L}_n$ as in Eq. (20), and

$$d = \frac{(q-2)m}{q(m+2)} \qquad \text{with} \qquad q \geq 3. \tag{24}$$

However, the SS has some drawbacks for small values of $\mathcal{L}_n$ such as the shrinkage factor $(1 - d\mathcal{L}_n^{-1})$ becomes negative for $\mathcal{L}_n < d$. Another estimator is proposed using the positive-rule shrinkage (PRS) estimator as follows

$$\hat{\boldsymbol{\beta}}^{S+} = \hat{\boldsymbol{\beta}} + (1 - d\mathcal{L}_n^{-1}) I(\mathcal{L}_n > d)(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}), \tag{25}$$

where $I()$ shows the indicator function.

**Bias and Risk for the Estimators**

For an estimator $\boldsymbol{\beta}^*$, the risk function under the BLF in Eq. (5) is given by

$$\boldsymbol{R}^{\boldsymbol{W}}_{\omega,\boldsymbol{\delta}_0}(\boldsymbol{\beta}^*;\boldsymbol{\beta}) = E\left\{E[L^{\boldsymbol{W}}_{\omega,\boldsymbol{\delta}_0}(\boldsymbol{\beta}^*;\boldsymbol{\beta})]|\boldsymbol{\beta}\right\}. \tag{26}$$

Note that the computation of the risk function in Eq. (26) depends on the balanced-type expectation (inner expectation) w.r.t the sampling distribution. Then the expectation (outer expectation) from the prior information is obtained by equation (15).
This section is dedicated to deriving the bias and the risk function in Eq. (26) for the five different estimators when the target estimator is determined by $\boldsymbol{\delta}_0 = \tilde{\boldsymbol{\beta}}$ and the weight matrix is given by $\boldsymbol{W} = \boldsymbol{C}$, where $\boldsymbol{C}$ is defined in Eq. (6). We will simply write $\boldsymbol{R}^{\boldsymbol{W}}_0(\boldsymbol{\beta}^*;\boldsymbol{\beta})$ for $\omega = 0$.

**Computation of bias**

The bias of the estimator $\hat{\boldsymbol{\beta}}^B$ in Eq. (14) and RE estimator $\hat{\boldsymbol{\beta}}$ in Eq. (18) are

$$\boldsymbol{b}_1 = E[\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}] = \boldsymbol{0}, \text{ and } \boldsymbol{b}_2 = E[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] = -\boldsymbol{\Delta}, \tag{27}$$

respectively, where $\boldsymbol{\Delta} = \boldsymbol{C}^{-1}\boldsymbol{H}'\boldsymbol{V}_1(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h})$. Using [22], the bias of the PT is

$$\begin{aligned}\boldsymbol{b}_3 &= E(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta}) = E[\tilde{\boldsymbol{\beta}} - I(\mathcal{L}_n \le F_\alpha)(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}] \\ &= -\boldsymbol{C}\boldsymbol{H}'\mathbf{V}_1^{1/2}\, E[I(\mathcal{L}_n \le F_\alpha)\mathbf{V}_1^{1/2}(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})] = -\boldsymbol{\Delta}G^{(2)}_{q+2,m}\left(F_\alpha;\Delta_*^2\right),\end{aligned} \tag{28}$$

where $\Delta_*^2 = \theta/\sigma_{\boldsymbol{\epsilon}}^2$, $\theta = (\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h})'\boldsymbol{V}_1'(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h})$,

$$G^{(2-j)}_{q+2i,m}(l_\alpha,\Delta_*^2) = \sum_{r=0}^{\infty} K_r^{(j)}(\Delta_*^2) I_{l_\alpha}\left[\frac{q+2i}{2} + r, \frac{m}{2}\right], \text{ and } j = 0, 1, \tag{29}$$

$l_\alpha = \frac{qF_{q,m}(\alpha)}{m+qF_{q,m}(\alpha)}$, $I_x[a,b] = \frac{1}{B(a,b)}\int_0^x u^{a-1}(1-u)^{b-1}du$ is the incomplete function of beta,

$$K_r^{(j)}(\Delta_*^2) = \frac{[-2\psi'(0)]^r}{r!}\left(\frac{\Delta_*^2}{2}\right)^2 \int_0^\infty t^{r-j} e^{t\psi'(0)\Delta_*^2} W(t)dt, \tag{30}$$

and $W(t)$ is a weight function.
Then, the bias of the SS becomes

$$\begin{aligned}\boldsymbol{b}_4 &= E(\hat{\boldsymbol{\beta}}^S - \boldsymbol{\beta}) = E[\tilde{\boldsymbol{\beta}} - d\mathcal{L}_n^{-1}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}] \\ &= -d\boldsymbol{C}^{-1}\boldsymbol{H}'\boldsymbol{V}_1^{1/2}E[\mathcal{L}_n^{-1}\boldsymbol{V}_1^{1/2}(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})] = -dq\boldsymbol{\Delta}E^{(2)}[\chi^{*-2}_{q+2}(\Delta_*^2)],\end{aligned} \tag{31}$$

and here

$$E^{(2-j)}[\chi^{*-2}_{q+s}(\Delta_*^2)] = \sum_{r\ge 0}\frac{1}{r!}K_r^{(j)}(\Delta_*^2)(q+s-2+2r)^{-1}, \text{ for } s = 2, 4. \tag{32}$$

Finally, the bias of the PRSE is obtained as follows:

$$\begin{aligned}\boldsymbol{b}_5 &= E(\hat{\boldsymbol{\beta}}^S - \boldsymbol{\beta}) - E[I(\mathcal{L}_n \le d)(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})] + dE[\mathcal{L}_n^{-1} I(\mathcal{L}_n \le d)(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})] \\ &= -dq\boldsymbol{\Delta} E_N^{(2)}[\chi_{q+2}^{*-4}(\Delta_*^2)] + \boldsymbol{\Delta} G_{q+2,m}^{(2)}\left(d; \Delta_*^2\right) \\ &+ \frac{qd}{q+2}\,\boldsymbol{\Delta} E^{(2)}\left[F_{q+2,m}^{-1}(\Delta_*^2) I\left(F_{q+2,m}(\Delta_*^2) \le \frac{qd}{q+2}\right)\right], \end{aligned} \tag{33}$$

where

$$\begin{aligned} &E^{(2-j)}[F_{q+s,n-p}^{-k}(\Delta_*^2) I(F_{q+s,n-p}(\Delta_*^2) < d_1)] \\ &= \sum_{r=0}^{\infty} K_r^{(j)}(\Delta_*^2)\left(\frac{q+s}{n-p}\right)^k \frac{B(\frac{q+s+2r-2k}{2}, \frac{m+2k}{2})}{B(\frac{q+s+2r}{2}, \frac{m}{2})} I_{x'}\left[\frac{q+s+2r-2k}{2}, \frac{m+2k}{2}\right], \end{aligned} \tag{34}$$

in which $d_1 = \frac{dq}{q+2}$, and $x' = \frac{dq}{m+dq}$.
Note that for the non-centrality parameter $\Delta_*^2 (= \boldsymbol{\Delta}'\boldsymbol{C}\boldsymbol{\Delta}/\sigma_{\boldsymbol{\epsilon}}^2) \to \infty$, the bias of these estimators are obtained by $\boldsymbol{b}_1 = \boldsymbol{b}_3 = \boldsymbol{b}_4 = \boldsymbol{b}_5 = \boldsymbol{0}$, except for $\boldsymbol{b}_2$ which becomes unbounded. However, under $H_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$ hypothesis, because $\boldsymbol{\Delta} = \boldsymbol{0}$, the bias of that estimators is defined as $\boldsymbol{b}_1 = \boldsymbol{b}_2 = \boldsymbol{b}_3 = \boldsymbol{b}_4 = \boldsymbol{b}_5 = \boldsymbol{0}$.

**Computation of risk**

The risk function of the Bayes estimator $\hat{\boldsymbol{\beta}}^B$, $\boldsymbol{R}_{\omega,\tilde{\boldsymbol{\beta}}}^{\boldsymbol{C}}(.; \boldsymbol{\beta})$ in Eq. (26), is computed as

$$\begin{aligned}\boldsymbol{R}_{\omega,\tilde{\boldsymbol{\beta}}}^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}^B; \boldsymbol{\beta}) &= (1-\omega) E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right) E[(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{C}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})]|\boldsymbol{\beta}\right\} \\ &= p\,\sigma_{\boldsymbol{\epsilon}}^2 (1-\omega) E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)\right\} \\ &= p\,\phi^{-1}\sigma_{\boldsymbol{\epsilon}}^2(1-\omega). \end{aligned} \tag{35}$$

where $\omega \in [0, 1]$ (see Eq. (5)), $\sigma_{\boldsymbol{\epsilon}}^2 = -2\psi'(0)\sigma^2$, and

$$\phi = \left(\frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_{\mathbb{R}+} y^{\frac{n}{2}-1} g_n(y) \mathrm{d}y\right)^{-1}. \tag{36}$$

Using the fact that (see [8]) $\boldsymbol{V}_1^{\frac{1}{2}}(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h}) \sim \mathcal{E}_q(\boldsymbol{V}_1^{\frac{1}{2}}(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}), \sigma^2 \boldsymbol{I}_q, \boldsymbol{g})$, the risk of the estimator RE is given by

$$\begin{aligned}\boldsymbol{R}_{\omega,\tilde{\boldsymbol{\beta}}}^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}; \boldsymbol{\beta}) &= \omega E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right) E[(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})'\boldsymbol{V}_1(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})]|\boldsymbol{\beta}\right\} \\ &+ (1-\omega) E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right) E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{C}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]|\boldsymbol{\beta}\right\} \\ &= -q\omega\sigma_{\boldsymbol{\epsilon}}^2 E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)\right\} + (1-\omega) E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)\right\}[\sigma_{\boldsymbol{\epsilon}}^2 \mathrm{tr}(\boldsymbol{V}_2\boldsymbol{C}) + \boldsymbol{\Delta}'\boldsymbol{C}\boldsymbol{\Delta}] \\ &= -q\omega\sigma_{\boldsymbol{\epsilon}}^2\phi^{-1} + \boldsymbol{R}_{\omega,\tilde{\boldsymbol{\beta}}}^{\boldsymbol{C}}(\tilde{\boldsymbol{\beta}}; \boldsymbol{\beta}) + \phi^{-1}(1-\omega)(-\sigma_{\boldsymbol{\epsilon}}^2 \mathrm{tr}[\boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{H}\boldsymbol{C}] + \boldsymbol{\Delta}'\boldsymbol{C}\boldsymbol{\Delta}) \\ &= \boldsymbol{R}_{\omega,\tilde{\boldsymbol{\beta}}}^{\boldsymbol{C}}(\tilde{\boldsymbol{\beta}}; \boldsymbol{\beta}) - q\,\phi^{-1}\sigma_{\boldsymbol{\epsilon}}^2 + (1-\omega)\phi^{-1}\theta, \end{aligned} \tag{37}$$

where $\theta = \boldsymbol{\Delta}'\boldsymbol{C}\boldsymbol{\Delta} = (\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h})'\boldsymbol{V}_1(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h})$. Note that $\boldsymbol{R} = \boldsymbol{C}^{-1/2}\boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{H}\boldsymbol{C}^{-1/2}$ is a symmetric idempotent matrix of rank $q \leq p$ and $\boldsymbol{C} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$. Thus, there exists an orthogonal matrix $\boldsymbol{Q}$ ($\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}_p$) such that $\boldsymbol{Q}\boldsymbol{R}\boldsymbol{Q}' = \begin{bmatrix} \boldsymbol{I}_q & 0 \\ 0 & 0 \end{bmatrix}$. A random vector is defined by $\boldsymbol{w} = \boldsymbol{Q}\boldsymbol{C}^{1/2}\tilde{\boldsymbol{\beta}} - \boldsymbol{Q}\boldsymbol{C}^{-1/2}\boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{h}$, then $\boldsymbol{w} \sim \mathcal{E}_p(\boldsymbol{\eta}, \sigma^2\boldsymbol{I}_p, \boldsymbol{g})$, where $\boldsymbol{\eta} = \boldsymbol{Q}\boldsymbol{C}^{1/2}\boldsymbol{\beta} - \boldsymbol{Q}\boldsymbol{C}^{-1/2}\boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{h}$. Partitioning the vector $\boldsymbol{w} = (\boldsymbol{w}_1', \boldsymbol{w}_2')'$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \boldsymbol{\eta}_2')'$, where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are the subvectors of sequence $q$ and $p-q$ respectively, one may rewrite the test statistic $\mathcal{L}_n$ defined in Eq. (17) as

$$\mathcal{L}_n = \frac{\boldsymbol{w}_1'\boldsymbol{w}_1}{qS^2}, \quad \theta = \boldsymbol{\eta}_1'\boldsymbol{\eta}_1. \tag{38}$$

Consequently, the risk of the PT, $\boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{\beta})$, noting that $\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} = \boldsymbol{C}^{-1}\boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{H}\boldsymbol{C}^{-\frac{1}{2}}\boldsymbol{w}$ as follows:

$$\begin{aligned}
\boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{\beta}) &= \omega E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)E[I(\mathcal{L}_n < F_\alpha)(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\boldsymbol{C}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})]|\boldsymbol{\beta}\right\} \\
&+ (1-\omega)E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)E[(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})'\boldsymbol{C}(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})]|\boldsymbol{\beta}\right\} \\
&= \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}) - (1-2\omega)E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)E[\boldsymbol{w}_1'\boldsymbol{w}_1 I(\mathcal{L}_n \leq F_\alpha)]|\boldsymbol{\beta}\right\} \\
&+ 2(1-\omega)E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)\boldsymbol{\eta}_1'E[\boldsymbol{w}_1 I(\mathcal{L}_n \leq F_\alpha)]|\boldsymbol{\beta}\right\} \\
&= \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}) - (1-2\omega)q\sigma^2_{\boldsymbol{\epsilon}}\phi^{-1}G^{(1)}_{q+2,m}\left(F_\alpha;\Delta^2_*\right) \\
&+ 2\theta(1-\omega)\phi^{-1}\left[2G^{(2)}_{q+2,m}\left(F_\alpha;\Delta^2_*\right) - G^{(2)}_{q+4,m}\left(F_\alpha;\Delta^2_*\right)\right],
\end{aligned} \tag{39}$$

where we obtain this result by simplifying the expressions.
Similarly, after simplifying the expressions, the risk of the SE $\boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S};\boldsymbol{\beta})$ is obtained by

$$\begin{aligned}
\boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S};\boldsymbol{\beta}) &= \omega d^2 E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)E[\mathcal{L}_n^{-1}\boldsymbol{w}_1'\boldsymbol{w}_1]|\boldsymbol{\beta}\right\} + \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}) \\
&- 2d(1-\omega)E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)E[\mathcal{L}_n^{-1}(\boldsymbol{w}_1'\boldsymbol{w}_1 - \boldsymbol{\eta}_1'\boldsymbol{w}_1)]|\boldsymbol{\beta}\right\} \\
&+ d^2(1-\omega)E_{\boldsymbol{\beta}}\left\{r\left(\|\boldsymbol{\beta}\|^2\right)E[\mathcal{L}_n^{-2}\boldsymbol{w}_1'\boldsymbol{w}_1]|\boldsymbol{\beta}\right\} \\
&= \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}) + q\ \phi^{-1}\Bigg\{\left[d^2\omega - 2d(1-\omega)\right]E^{(1)}[\chi^{*-2}_{q+2}(\Delta^2_*)] \\
&+ d^2(1-\omega)E^{(1)}[\chi^{*-4}_{q+2}(\Delta^2_*)]\Bigg\} + \theta\ \phi^{-1} \\
&\times\Bigg\{\left[d^2\omega - 2d(1-\omega)\right]E^{(2)}[\chi^{*-2}_{q+4}(\Delta^2_*)] \\
&- 2d(1-\omega)E^{(2)}[\chi^{*-2}_{q+2}(\Delta^2_*)] + d^2(1-\omega)E^{(2)}[\chi^{*-4}_{q+4}(\Delta^2_*)]\Bigg\},
\end{aligned} \tag{40}$$

where

$$E^{(2-j)}[\chi^{*-4}_{q+s}(\Delta^2_*)] = \sum_{r\geq 0}\frac{1}{r!}K^{(j)}_r(\Delta^2_*)(q+s-2+2r)^{-1}(q+s-4+2r)^{-1}, \tag{41}$$

$\omega \in [0,1]$ is the weight used in the BLF, and (24)

$$d = \frac{(q-2)m}{q(m+2)} \qquad \text{with} \qquad q \geq 3, \tag{42}$$

and $\boldsymbol{w}_1$ is the first sub-partition of $\boldsymbol{w}$. Finally, the risk function of the PRSE, $\boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S+};\boldsymbol{\beta})$ is found as follows

$$\begin{aligned}
\boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S+};\boldsymbol{\beta}) &= \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S};\boldsymbol{\beta}) - E_{\boldsymbol{\beta}}\Big\{ r\left(\|\boldsymbol{\beta}\|^2\right) E[(1-d\mathcal{L}_n^{-1})^2 I(\mathcal{L}_n \leq d) \\
&\times (\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})'\boldsymbol{C}(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})]|\boldsymbol{\beta}\Big\} - 2E_{\boldsymbol{\beta}}\Big\{ r\left(\|\boldsymbol{\beta}\|^2\right) E\{[(1-d\mathcal{L}_n^{-1}) I(\mathcal{L}_n \leq d) \\
&\times (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{C}(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})]\}|\boldsymbol{\beta}\Big\} \\
&= \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S};\boldsymbol{\beta}) - E_{\boldsymbol{\beta}}\left\{ r\left(\|\boldsymbol{\beta}\|^2\right) E[(1-d\mathcal{L}_n^{-1})^2 I(\mathcal{L}_n \leq d)\boldsymbol{w}_1'\boldsymbol{w}_1]|\boldsymbol{\beta}\right\} \\
&- 2E_{\boldsymbol{\beta}}\left\{ r\left(\|\boldsymbol{\beta}\|^2\right) E[(1-d\mathcal{L}_n^{-1}) I(\mathcal{L}_n \leq d)(\boldsymbol{w}_1'\boldsymbol{w}_1 - \boldsymbol{\eta}_1'\boldsymbol{w}_1)]|\boldsymbol{\beta}\right\} \\
&= \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S};\boldsymbol{\beta}) \\
&- \phi^{-1}\sigma^2_{\boldsymbol{\epsilon}}\Big\{ qE^{(1)}\left[\left(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(\Delta^2_*)\right)^2 I\left(F_{q+2,m}(\Delta^2_*) \leq \frac{qd}{q+2}\right)\right] \\
&+ \frac{\theta}{\sigma^2_{\boldsymbol{\epsilon}}} E^{(2)}\left[\left(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(\Delta^2_*)\right)^2 I\left(F_{q+2,m}(\Delta^2_*) \leq \frac{qd}{q+2}\right)\right]\Big\} \\
&- 2\phi^{-1}\theta E^{(2)}\left[\left(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(\Delta^2_*)\right) I\left(F_{q+2,m}(\Delta^2_*) \leq \frac{qd}{q+2}\right)\right].
\end{aligned} \tag{43}$$

**Comparison of Risks**

Recall that the LS estimator $\tilde{\boldsymbol{\beta}}$ is designated as UE and $\hat{\boldsymbol{\beta}}$ is termed as RE. Then, the risk difference of the estimator RE in Eq. (18) and the estimator UE in Eq. (6) is given by

$$\mathcal{D}_{21} = \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}};\boldsymbol{\beta}) - \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}) = \phi^{-1}\left[(1-\omega)\theta - q\sigma^2_{\boldsymbol{\epsilon}}\right]. \tag{44}$$

It is shown that $\hat{\boldsymbol{\beta}}$ outperforms to $\tilde{\boldsymbol{\beta}}$ ($\hat{\boldsymbol{\beta}} \succeq \tilde{\boldsymbol{\beta}}$) - in other words $\hat{\boldsymbol{\beta}}$ dominates $\tilde{\boldsymbol{\beta}}$ - provided $0 \leq \theta \leq \frac{q\sigma^2_{\boldsymbol{\epsilon}}}{1-\omega}$ for $\omega \neq 1$ since $\phi > 0$.
First, let us start the comparison of risk with $\hat{\boldsymbol{\beta}}^{PT}$ versus $\tilde{\boldsymbol{\beta}}$. We use the risk difference defined by

$$\begin{aligned}
\mathcal{D}_{13} &= \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}) - \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{\beta}) = (1-2\omega)q\,\phi^{-1}\sigma^2_{\boldsymbol{\epsilon}} G^{(1)}_{q+2,m}\left(F_\alpha;\Delta^2_*\right) \\
&- 2\theta\,\phi^{-1}(1-\omega)[2G^{(2)}_{q+2,m}\left(F_\alpha;\Delta^2_*\right) - G^{(2)}_{q+4,m}\left(F_\alpha;\Delta^2_*\right)].
\end{aligned} \tag{45}$$

where $G$ is defined in Eq. (29), $\omega \in [0,1]$ is the weight under the BLF, $q$ is the rank of matrix $\boldsymbol{H}$,

$$\phi^{-1} = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_{\mathbb{R}^+} y^{\frac{n}{2}-1} g_n(y)\mathrm{d}y \tag{46}$$

$\sigma_{\boldsymbol{\epsilon}}^2 = -2\psi'(0)\sigma^2$, $F_\alpha$ is the upper $\alpha^{th}$ percentile of the central F-distribution with $(q, m)$ df, and $\Delta_*^2 = \theta/\sigma_{\boldsymbol{\epsilon}}^2$, $\theta = (\boldsymbol{H\beta} - \boldsymbol{h})'\boldsymbol{V}_1'(\boldsymbol{H\beta} - \boldsymbol{h})$. For $\omega \neq 1$, the of Eq. (45) is not negative $\hat{\boldsymbol{\beta}}^{PT} \succeq \tilde{\boldsymbol{\beta}}$ whenever

$$\theta \leq \frac{(1-2\omega)q\sigma_{\boldsymbol{\epsilon}}^2 G_{q+2,m}^{(1)}\left(F_\alpha; \Delta_*^2\right)}{2(1-\omega)\left[2G_{q+2,m}^{(2)}\left(F_\alpha; \Delta_*^2\right) - G_{q+4,m}^{(2)}\left(F_\alpha; \Delta_*^2\right)\right]}. \tag{47}$$

Under the hypothesis $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$ when $\theta = 0$, $\hat{\boldsymbol{\beta}}^{PT} \succeq \tilde{\boldsymbol{\beta}}$ for $\omega$ with $\omega \leq \frac{1}{2}$.
Secondly, the comparison of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{PT}$ using the risk difference is defined as follows

$$\begin{aligned}\mathcal{D}_{23} &= \boldsymbol{R}_{\omega,\tilde{\boldsymbol{\beta}}}^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}; \boldsymbol{\beta}) - \boldsymbol{R}_{\omega,\tilde{\boldsymbol{\beta}}}^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}^{PT}; \boldsymbol{\beta}) \\ &= -q\,\phi^{-1}\sigma_{\boldsymbol{\epsilon}}^2[1 - (1-2\omega)G_{q+2,m}^{(1)}\left(F_\alpha; \Delta_*^2\right)] + \theta\,\phi^{-1}(1-\omega)[1 - 2G_{q+2,m}^{(2)}\left(F_\alpha; \Delta_*^2\right) \\ &\quad + G_{q+4,m}^{(2)}\left(F_\alpha; \Delta_*^2\right)],\end{aligned} \tag{48}$$

where $G$ is defined in Eq. (29), $\omega \in [0, 1]$ is the weight used in the BLF, $q$ is the rank of matrix $\boldsymbol{H}$,

$$\phi^{-1} = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_{\mathbb{R}^+} y^{\frac{n}{2}-1} g_n(y)\mathrm{d}y, \tag{49}$$

$\sigma_{\boldsymbol{\epsilon}}^2 = -2\psi'(0)\sigma^2$, $F_\alpha$ is the upper $\alpha^{th}$ percentile of the central F-distribution with $(q, m)$ df, and $\Delta_*^2 = \theta/\sigma_{\boldsymbol{\epsilon}}^2$, $\theta = (\boldsymbol{H\beta} - \boldsymbol{h})'\boldsymbol{V}_1'(\boldsymbol{H\beta} - \boldsymbol{h})$.
Thus $\hat{\boldsymbol{\beta}}^{PT} \succeq \hat{\boldsymbol{\beta}}$ whenever

$$\theta \geq \frac{q\sigma_{\boldsymbol{\epsilon}}^2\left[1 - (1-2\omega)G_{q+2,m}^{(1)}\left(F_\alpha; \Delta_*^2\right)\right]}{(1-\omega)\left[1 - 2G_{q+2,m}^{(2)}\left(F_\alpha; \Delta_*^2\right) + G_{q+4,m}^{(2)}\left(F_\alpha; \Delta_*^2\right)\right]}, \tag{50}$$

and vice versa. Under $H_0$, the superiority order of $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{PT}$ is defined by

$$\hat{\boldsymbol{\beta}} \succeq \hat{\boldsymbol{\beta}}^{PT} \succeq \tilde{\boldsymbol{\beta}}, \quad \text{or} \quad \hat{\boldsymbol{\beta}}^{PT} \succeq \hat{\boldsymbol{\beta}} \succeq \tilde{\boldsymbol{\beta}}, \tag{51}$$

depending on the value $\alpha$ satisfying (5.5).
To demonstrate the superiority of $\hat{\boldsymbol{\beta}}^S$ to $\tilde{\boldsymbol{\beta}}$, we give the following results.

**Theorem 1.** Let us take into account the model in Eq. (1) with the error vector based on the ECD, $\mathcal{E}_n(\mathbf{0}, \sigma^2\mathbf{V}, \boldsymbol{g})$. Then the SS estimator of $\boldsymbol{\beta}$ is defined by

$$\hat{\boldsymbol{\beta}}^S = \tilde{\boldsymbol{\beta}} - d^*\mathcal{L}_n^{-1}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \tag{52}$$

where $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ are the UE and RE, respectively. The SS estimator uniformly controls the Bayes estimator $\tilde{\boldsymbol{\beta}}$ with respect to the BLF $L_0^{\boldsymbol{C}}(\boldsymbol{\delta}; \boldsymbol{\beta})$ and it is minimax iff $0 < d^* \leq \frac{2m}{m+2}$. The biggest risk decrease is obtained for $d^* = \frac{m}{m+2}$.

**Proof**: Let $\left(\frac{mS^2}{\sigma^2}\right)\Big|\tau \sim \tau^{-1}\chi_m^2$ and $\tilde{\boldsymbol{\beta}}'\boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{H}\tilde{\boldsymbol{\beta}} \mid \tau \sim \tau^{-2}\sigma^4\chi_q^2(\dot{\boldsymbol{\Delta}})$. Defining $\dot{\boldsymbol{z}} = \boldsymbol{H}'\boldsymbol{V}_1(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})$, the SS estimator is as follows

$$\begin{aligned}\hat{\boldsymbol{\beta}}^S &= \tilde{\boldsymbol{\beta}} - qd^* S^2 \left[(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})'\boldsymbol{V}_1(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h})\right]^{-1} \boldsymbol{C}^{-1}\boldsymbol{H}'\boldsymbol{V}_1(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h}) \\ &= \tilde{\boldsymbol{\beta}} - qd^* S^2 \left(\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}\right)^{-1} \boldsymbol{C}^{-1}\dot{\boldsymbol{z}}. \end{aligned} \quad (53)$$

Using the result of [23], the weight function $W(t)$ in elliptical models satisfies $\int_0^\infty t^{-1}W^+(dt) < \infty$ and $\int_0^\infty t^{-1}W^-(dt) < \infty$, where $W^+ - W^-$ is the Jordan decomposition of $W$ in positive and negative part.

Therefore, the risk difference between the SS estimator and the Bayes estimator under BLF is then given by

$$\begin{aligned}\mathcal{D}_{41} &= E_{\boldsymbol{\beta}} \left\{ E(\hat{\boldsymbol{\beta}}^S - \boldsymbol{\beta})'\boldsymbol{C}(\hat{\boldsymbol{\beta}}^S - \boldsymbol{\beta}) - E(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{C}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})|\boldsymbol{\beta}\right\} \\ &= E_{\boldsymbol{\beta}} \left\{ (d^*)^2 E\left[q^2 S^4 \left(\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}\right)^{-1}\right] - 2d^* E\left[qS^2 \left(\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}\right)^{-1} (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\dot{\boldsymbol{z}}\right] |\boldsymbol{\beta}\right\} \\ &= E_{\boldsymbol{\beta}} \Bigg\{ (d^*)^2 E_\tau \left\{ E_N \left[ q^2 S^4 \left(\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}\right)^{-1} \middle| \tau \right]\right\} \\ &\quad - 2d^* E_\tau \left\{ E_N \left[ qS^2 \left(\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}\right)^{-1} (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{H}'\boldsymbol{V}_1(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h}) \middle| \tau \right]\right\} |\boldsymbol{\beta} \Bigg\} \\ &= \phi^{-1} \left\{ \frac{q^2(m+2)}{m} (d^*)^2 E_\tau \left( \frac{\tau^{-2}}{\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}} \right) - 2q^2 d^* E_\tau \left( \frac{\tau^{-2}}{\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}} \right) \right\}, \end{aligned} \quad (54)$$

where $\dot{\boldsymbol{\Delta}} = \boldsymbol{\beta}'\boldsymbol{H}'\boldsymbol{V}_1\boldsymbol{H}\boldsymbol{\beta}$ and $E_N$ means getting expectation with respect to multivariate normal with covariance $\tau^{-1}\sigma^2\boldsymbol{V}$ and $E_\tau$ means getting expectation with respect to measure $dW(.)$.

Therefore, $\mathcal{D}_{41} \le 0$ iff $0 < d^* \le \frac{2m}{m+2}$ since $\int_0^\infty \frac{\tau^{-2}}{\dot{\boldsymbol{z}}'\boldsymbol{C}^{-1}\dot{\boldsymbol{z}}}\, dW(\tau) > 0$.

**Remark 1.** Consider the shrinkage coefficient $d$ given by Eq. (24). For $q \ge 3$, we get $0 < d = \frac{(q-2)m}{q(m+2)} < \frac{2m}{m+2}$ and thus using Theorem 1, the $\hat{\boldsymbol{\beta}}^S$ uniformly controls the $\tilde{\boldsymbol{\beta}}$ on the entire parameter space under BLF.

**Lemma 2.** (Arashi, 2012)

(i) The estimator $\boldsymbol{\delta}_0(\boldsymbol{X})+(1-\omega)\boldsymbol{h}_1(\boldsymbol{X})$ controls $\boldsymbol{\delta}_0(\boldsymbol{X})+(1-\omega)\boldsymbol{h}_2(\boldsymbol{X})$ under the BLF $L^{\boldsymbol{W}}_{\omega,\boldsymbol{\delta}_0}(\boldsymbol{\delta};\boldsymbol{\beta})$ iff $\boldsymbol{\delta}_0(\boldsymbol{X})+\boldsymbol{h}_1(\boldsymbol{X})$ dominates $\boldsymbol{\delta}_0(\boldsymbol{X})+\boldsymbol{h}_2(\boldsymbol{X})$ with respect to the quadratic loss function $L^{\boldsymbol{W}}_0(\boldsymbol{\delta};\boldsymbol{\beta})$.

(ii) Suppose that the estimator $\boldsymbol{\delta}_0(\boldsymbol{X})$ has the constant risk $\gamma$ with respect to the quadratic loss function $L^{\boldsymbol{W}}_0(\boldsymbol{\delta};\boldsymbol{\beta})$. Then $\boldsymbol{\delta}_0(\boldsymbol{X})$ is minimax under the BLF $L^{\boldsymbol{W}}_{\omega,\boldsymbol{\delta}_0}(\boldsymbol{\delta};\boldsymbol{\beta})$ with constant risk $(1-\omega)\gamma$ iff $\boldsymbol{\delta}_0(\boldsymbol{X})$ is minimax under the quadratic loss function $L^{\boldsymbol{W}}_0(\boldsymbol{\delta};\boldsymbol{\beta})$ with constant risk $\gamma$.

**Theorem 2.** Suppose $\boldsymbol{\epsilon} \sim \mathcal{E}_n(\boldsymbol{0}, \sigma^2\boldsymbol{V}, \boldsymbol{g})$ in the model (1). Then the SS estimator

$$\hat{\boldsymbol{\beta}}^S_* = \tilde{\boldsymbol{\beta}} - d(1-\omega)\mathcal{L}_n^{-1}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \quad (55)$$

uniformly dominates the $\tilde{\boldsymbol{\beta}}$ under the BLF $L^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta})$.

With Lemma 2(i) and Theorem 1, the theorem is proven easily.

**Corollary 1.** Let us take into consideration the Eq. (1), $\boldsymbol{\epsilon} \sim \mathcal{E}_n(\boldsymbol{0}, \sigma^2\boldsymbol{V}, \boldsymbol{g})$. In this case, $\hat{\boldsymbol{\beta}}^S \succeq \tilde{\boldsymbol{\beta}}$ under the BLF $L^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta})$.

Using Theorem 5.2 with $\omega = 0$, this corollary is proven, easily.

**Lemma 3.** Suppose $\boldsymbol{\epsilon} \sim \mathcal{E}_n(\mathbf{0}, \sigma^2\boldsymbol{V}, \boldsymbol{g})$ in the model (1). Then the estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is minimax under the BLF $L^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta})$ given in Eq. (2).

Theorem is proven using the information that $\tilde{\boldsymbol{\beta}}$ is minimax under quadratic loss function and applying Lemma 2 (ii).

**Remark 2.** Using Corollary 1 and Lemma 2, the SS estimator $\hat{\boldsymbol{\beta}}^S$ of $\boldsymbol{\beta}$ is minimax.

To compare $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^S$, it is easy to show that

$$\boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}}^S;\boldsymbol{\beta}) = \boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}};\boldsymbol{\beta}) + \phi^{-1}\Bigg(q\sigma^2_{\boldsymbol{\epsilon}} - \theta - dq^2\sigma^2_{\boldsymbol{\epsilon}}\Bigg\{(q-2)E[\chi^{*-4}_{q+2}(\Delta^2_*)] + \left[1 - \frac{(q+2)\theta}{2q\sigma^2_{\boldsymbol{\epsilon}}\Delta^2_*}\right](2\Delta^2_*)E[\chi^{*-4}_{q+4}(\Delta^2_*)]\Bigg\}\Bigg). \tag{56}$$

Under the hypothesis $H_0$, Eq. (56) becomes

$$\boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}}^S;\boldsymbol{\beta}) = \boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}};\boldsymbol{\beta}) + qd_n^{-1}\sigma^2_{\boldsymbol{\epsilon}}(1-d) \geq \boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}};\boldsymbol{\beta}), \tag{57}$$

where note that

$$\boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}};\boldsymbol{\beta}) = \boldsymbol{R}^{\boldsymbol{C}}_0(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}) - q\phi^{-1}\sigma^2_{\boldsymbol{\epsilon}} \leq \boldsymbol{R}^{\boldsymbol{C}}_0(\tilde{\boldsymbol{\beta}};\boldsymbol{\beta}). \tag{58}$$

Therefore, under the hypothesis $H_0$, $\hat{\boldsymbol{\beta}} \succeq \hat{\boldsymbol{\beta}}^S$ with the BLF $L^{\boldsymbol{C}}_0(\boldsymbol{\beta}^*,\boldsymbol{\beta})$. Under the hypothesis $H_0$, using Lemma 1(i), $\hat{\boldsymbol{\beta}} \succeq \hat{\boldsymbol{\beta}}^S$ with the BLF $L^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}^*;\boldsymbol{\beta})$. As $\boldsymbol{\eta}_1$ moves away from 0, $\theta$ increases and the risk of $\hat{\boldsymbol{\beta}}$ becomes unbounded while the risk of $\tilde{\boldsymbol{\beta}}^S$ remains below the risk of $\tilde{\boldsymbol{\beta}}$. $\tilde{\boldsymbol{\beta}}^S$ dominates $\hat{\boldsymbol{\beta}}$ outside an interval around the origin under the BLF $L^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}^*;\boldsymbol{\beta})$. The situation is repeated while comparing $\hat{\boldsymbol{\beta}}^S$ and $\hat{\boldsymbol{\beta}}^{PT}$. Let us consider the risk function under the hypothesis $H_0$ as follows

$$\boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}}^S;\boldsymbol{\beta}) = \boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{\beta}) + q\phi^{-1}\sigma^2_{\boldsymbol{\epsilon}}[1-\alpha-d] \geq \boldsymbol{R}^{\boldsymbol{C}}_0(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{\beta}), \tag{59}$$

for all $\alpha$ such that $F^{-1}_{q+2,m}(d,0) \leq \frac{qF_\alpha}{q+2}$, which means the estimator $\hat{\boldsymbol{\beta}}^S$ does not dominate $\hat{\boldsymbol{\beta}}^{PT}$ under $H_0$. Under hypothesis $H_0$ with $\alpha$ holding $F^{-1}_{q+2,m}(d,0) \leq \frac{qF_\alpha}{q+2}$ with the balanced loss function, $\hat{\boldsymbol{\beta}} \succeq \hat{\boldsymbol{\beta}}^{PT} \succeq \hat{\boldsymbol{\beta}}^S \succeq \tilde{\boldsymbol{\beta}}$ holds.

Afterwards, we compare the risks of $\hat{\boldsymbol{\beta}}^{S+}$ and $\hat{\boldsymbol{\beta}}^S$ with the risk difference given by

$$\begin{aligned}\mathcal{D}_{54} = \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^{S+};\boldsymbol{\beta}) - \boldsymbol{R}^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\hat{\boldsymbol{\beta}}^S;\boldsymbol{\beta}) = \\ -\phi^{-1}\sigma^2_{\boldsymbol{\epsilon}}\Bigg\{qE^{(1)}\left[\left(1 - \frac{qd}{q+2}F^{-1}_{q+2,m}(\Delta^2_*)\right)^2 I\left(F_{q+2,m}(\Delta^2_*) \leq \frac{qd}{q+2}\right)\right] \\ + \frac{\theta}{\sigma^2_{\boldsymbol{\epsilon}}}E^{(2)}\left[\left(1 - \frac{qd}{q+2}F^{-1}_{q+2,m}(\Delta^2_*)\right)^2 I\left(F_{q+2,m}(\Delta^2_*) \leq \frac{qd}{q+2}\right)\right]\Bigg\} \\ -2\phi^{-1}\theta E^{(2)}\left[\left(1 - \frac{qd}{q+2}F^{-1}_{q+2,m}(\Delta^2_*)\right) I\left(F_{q+2,m}(\Delta^2_*) \leq \frac{qd}{q+2}\right)\right].\end{aligned} \tag{60}$$

The r.h.s. of the above equality is negative for $F_{q+2,m}(\Delta_*^2) \leq \frac{qd}{q+2}$, $(\frac{qd}{q+2}F_{q+2,m}(\Delta_*^2) - 1) \geq 0$ and the expectation of a positive random variable is also positive. Thus, $\hat{\boldsymbol{\beta}}^{S+} \succeq \hat{\boldsymbol{\beta}}^{S}$.

**Remark 3.** The positive-rule shrinkage estimator $\hat{\boldsymbol{\beta}}^{S+}$ of $\boldsymbol{\beta}$ is minimax.

Let us extend the comparisons under $L_0^{\boldsymbol{C}}(\boldsymbol{\beta}^*;\boldsymbol{\beta})$. The same results can be obtained for the BLF $L^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}^*;\boldsymbol{\beta})$. For comparison of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{S+}$, let us take the case under the hypothesis $H_0$ i.e., $\boldsymbol{\eta}_1 = 0$. For this

$$\boldsymbol{R}_0^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}^{S+};\boldsymbol{\beta}) = \boldsymbol{R}_0^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}};\boldsymbol{\beta}) + q\phi^{-1}\sigma_{\boldsymbol{\epsilon}}^2\Bigg\{(1-d) - E\Bigg[(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(0))^2 \times I(F_{q+2,m}(0) \leq \frac{qd}{q+2})\Bigg]\Bigg\} \geq \boldsymbol{R}_0^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}};\boldsymbol{\beta}), \tag{61}$$

since $E\left[(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(0))^2 I(F_{q+2,m}(0) \leq \frac{qd}{q+2})\right] \leq E\left[(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(0))^2\right] = 1-d$.
Under the hypothesis $H_0$, $\hat{\boldsymbol{\beta}} \succeq \hat{\boldsymbol{\beta}}^{S+}$. As $\boldsymbol{\eta}_1$ moves away from 0, $\theta$ increases and the risk of $\hat{\boldsymbol{\beta}}$ becomes unbounded when the risk of $\tilde{\boldsymbol{\beta}}^{S+}$ stays below the risk of $\tilde{\boldsymbol{\beta}}$. This means that $\tilde{\boldsymbol{\beta}}^{S+}$ predominates $\hat{\boldsymbol{\beta}}$ outside an interval around the origin.

Then, we compare $\hat{\boldsymbol{\beta}}^{S+}$ and $\hat{\boldsymbol{\beta}}^{PT}$. When hypothesis $H_0$ holds, $G^*_{q+2,m}(F_\alpha, 0) = 1-\alpha$. The risk is given by

$$\boldsymbol{R}_0^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}^{S+};\boldsymbol{\beta}) = \boldsymbol{R}_0^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{\beta}) + q\phi^{-1}\sigma_{\boldsymbol{\epsilon}}^2\Bigg\{1-\alpha-d - E\Bigg[(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(0))^2 \times I(F_{q+2,m}(0) \leq \frac{qd}{q+2})\Bigg]\Bigg\} \geq \boldsymbol{R}_0^{\boldsymbol{C}}(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{\beta}) \tag{62}$$

for $\forall\ \alpha$ $E\left[(1-\frac{qd}{q+2}F^{-1}_{q+2,m}(0))^2 I(F_{q+2,m}(0) \leq \frac{qd}{q+2})\right] \leq 1-\alpha-d$.
As a result, $\hat{\boldsymbol{\beta}}^{S+}$ does not always dominate $\hat{\boldsymbol{\beta}}^{PT}$ when the null-hypothesis $H_0$ holds. Under the null hypothesis, the dominance order of five estimators under the BLF $L^{\boldsymbol{C}}_{\omega,\tilde{\boldsymbol{\beta}}}(\boldsymbol{\beta}^*;\boldsymbol{\beta})$ can be determined under following two categories

$$1.\quad \hat{\boldsymbol{\beta}} \succeq \hat{\boldsymbol{\beta}}^{PT} \succeq \hat{\boldsymbol{\beta}}^{S+} \succeq \hat{\boldsymbol{\beta}}^{S} \succeq \tilde{\boldsymbol{\beta}} \text{ and } 2.\quad \hat{\boldsymbol{\beta}} \succeq \hat{\boldsymbol{\beta}}^{\mathrm{S+}} \succeq \hat{\boldsymbol{\beta}}^{\mathrm{S}} \succeq \hat{\boldsymbol{\beta}}^{\mathrm{PT}} \succeq \tilde{\boldsymbol{\beta}}. \tag{63}$$

To look closely at the dominance relationships above, we examine the risk function of the estimators graphically in Figure 1. Let us suppose the error term in Eq. (1) has the multivariate Student's t (MT) distribution with $\boldsymbol{\epsilon} \sim \mathcal{T}_n(\mathbf{0}, \boldsymbol{I}_n, \nu)$. Figure 1 shows the results of the data sets with $n = 30$, $p = 5$, $q = 3$ for different df values $\nu = 5, 10$ and for different values of $\omega \in \{0, 0.5, 0.9\}$ to fulfill possible situations. The equations required for risk functions are considered as in [24], [25], and [26]. For more clarity, equations can be simply obtained from equations (35), (37), (39), (40), (43) and the fact that

$$K_{(r)}^{(\Delta_*^2)} = \frac{\Gamma\left(\frac{\nu}{2}+r\right)\left(\frac{\Delta_*^2}{\nu-2}\right)^r}{\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{\Delta_*^2}{\nu-2}\right)^{\frac{\nu}{2}+r}},$$

and

$$\begin{aligned}
G_{p,m}^{(j)}(l_\gamma;\Delta_*^2) &= \sum_{r\geq 0} \frac{\Gamma\left(\frac{\nu}{2}+r\right)\left(\frac{\Delta_*^2}{\nu-2}\right)^r I_x\left[\frac{1}{2}(p+2r), \frac{m}{2}\right]}{\Gamma(r+1)\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{\Delta_*^2}{\nu-2}\right)^{\frac{\nu}{2}+r}},\\
E^{(j)}[\chi_{p+s}^{*-2}(\Delta_*^2)] &= \sum_{r\geq 0} \frac{\Gamma\left(\frac{\nu}{2}+r+j-2\right)}{\Gamma(r+1)\Gamma\left(\frac{\nu}{2}\right)(p+s-2+2r)}\\
&\quad\times \frac{\left(\frac{\Delta_*^2}{\nu-2}\right)^r}{\left(1+\frac{\Delta_*^2}{\nu-2}\right)^{\frac{\nu}{2}+r+j-2}},\\
E^{(j)}[\chi_{p+s}^{*-4}(\Delta_*^2)] &= \sum_{r\geq 0} \frac{\Gamma\left(\frac{\nu}{2}+r+j-2\right)\left(\frac{\Delta_*^2}{\nu-2}\right)^r}{\Gamma(r+1)\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{\Delta_*^2}{\nu-2}\right)^{\frac{\nu}{2}+r+j-2}}\\
&\quad\times \frac{1}{(q+s-2+2r)(p+s-4+2r)}\\
E^{(j)}[F_{p+s,m}^{-1}(\Delta_*^2)I(F_{p+s,m}(\Delta_*^2)<c_i)] &= \sum_{r\geq 0} \frac{\Gamma\left(\frac{\nu}{2}+r+j-2\right)\left(\frac{\Delta_*^2}{\nu-2}\right)^r}{\Gamma(r+1)\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{\Delta_*^2}{\nu-2}\right)^{\frac{\nu}{2}+r+j-2}}\\
&\quad\times \frac{(p+s)I_{x'}\left[\frac{p+s-2+2r}{2}, \frac{m+2}{2}\right]}{(p+s-2+2r)},\\
E^{(j)}[F_{p+s,m}^{-2}(\Delta_*^2)I(F_{p+s,m}(\Delta_*^2)<c_i)] &= \sum_{r\geq 0} \frac{\Gamma\left(\frac{\nu}{2}+r+j-2\right)\left(\frac{\Delta_*^2}{\nu-2}\right)^r}{\Gamma(r+1)\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{\Delta_*^2}{\nu-2}\right)^{\frac{\nu}{2}+r+j-2}}\\
&\quad\times \frac{(q+s)^2 I_{x'}\left[\frac{p+s-4+2r}{2}, \frac{m+4}{2}\right]}{m(p+s-2+2r)(p+s-4+2r)}.
\end{aligned}$$

The value of $\phi^{-1}$ as in Eq. (2) is explored for arbitrary selection the arguments using Eq. (17) as

$$\phi^{-1} = \frac{(\pi\nu)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{p+\nu}{2}\right)}.$$

**Application**

Table 1 presents specific values of $\phi^{-1} = (\pi\nu)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)/\Gamma\left(\frac{p+\nu}{2}\right)$ for various parameter combinations $(p,\nu)$. It is important to note that when $\phi^{-1}$ (for the t-distribution kernel) becomes smaller than 1, it leads to a reduction in risk values. We observe that as the number of parameters $p$ increases while $\nu$ remains relatively small, the risk values decrease compared to situations with small $p$ and large $\nu$.

Therefore, as the dimension of the parameter space grows, we recommend using the MT distribution rather than the multivariate normal distribution for the error terms. Note that we eliminated the effect of $\phi^{-1}$ in the graphs, since it gives the same amount of $54.16$ and $679.36$ for $\nu = 5$ and $\nu = 10$, respectively, according to each risk value as seen in Table 1.

The dimension $p$ is constrained by the sample size for the results to remain valid, as the theoretical framework assumes $n > p$. Hence, if the model includes a large number of regressors (i.e., a high value of $p$), the sample size must also be sufficiently large to ensure the validity of the model.

***Table 1.*** *Values of $\phi^{-1}$ for different parameter values $(p, \nu)$*

| $(p, \nu)$ | 1 | 2 | 5 | 10 | 100 |
|---|---|---|---|---|---|
| 1 | 3.14 | 2.82 | 2.63 | 2.56 | 2.51 |
| 2 | $2\pi$ | $2\pi$ | $2\pi$ | $2\pi$ | $2\pi$ |
| 5 | 15.56 | 29.77 | 54.16 | 70.95 | 95.36 |
| 10 | 10.36 | 81.60 | 679.36 | 2023.93 | 8063.73 |
| 100 | $< 1$ | $< 1$ | $< 1$ | 74883.62 | $\infty$ |

Figure 1 demonstrates the risk function results of the suggested estimators for different numbers of df comparing the risks of the PT, SE, and PRSE for the chosen number of df (here which is 5) with different $\omega$. Two upper frames of Figure 1 illustrate the superiority relationship given in equation (63). The two lower frames of Figure 1 clearly demonstrate that as the $w$ and $\alpha$ decrease, the risk increases. Furthermore, as we deviate from the origin, or the null hypothesis, the risk values increase. From this perspective, Figure 1 supports the results of the analytical comparison covered earlier here. The risk values decrease when $\omega$ increases, which means, if the model fit is good under the structure of BLF, then the risk values decrease naturally.

Figure 2 shows how rapidly $\phi^{-1}$ increases as both the arguments $(p, \nu)$ increase. Moreover, the contour plot shows the skewed performance relative to the df, i.e., as the number of df gets larger, the coefficient $\phi^{-1}$ increases rapidly with two exceptions for the cases $p = 1, 2$.

## Conclusions and Remarks

This paper addresses the multiple regression model with elliptically contoured errors, utilizing a Bayesian approach that integrates a prior-based balanced loss function (BLF) applied to the parameter space. This methodology inherently influences the statistical properties of the resulting estimators. Specifically, from a Bayesian standpoint, we introduce a set of shrinkage estimators designed for regression analysis. In this novel framework, we derive key statistical criteria, such as bias and risk, and re-assess the properties of these estimators across various scenarios.

Of particular interest is the fact that the proposed balanced loss shrinkage factor demonstrates robustness to outliers in the data. By adjusting for extreme values, this approach provides a more stable estimation process compared to traditional methods. Furthermore, we examine the conditions under which a broad class of shrinkage estimators can achieve minimax properties—ensuring optimal performance across a range of possible parameter values. This investigation reveals the circumstances under which these estimators minimize both bias and variance simultaneously.

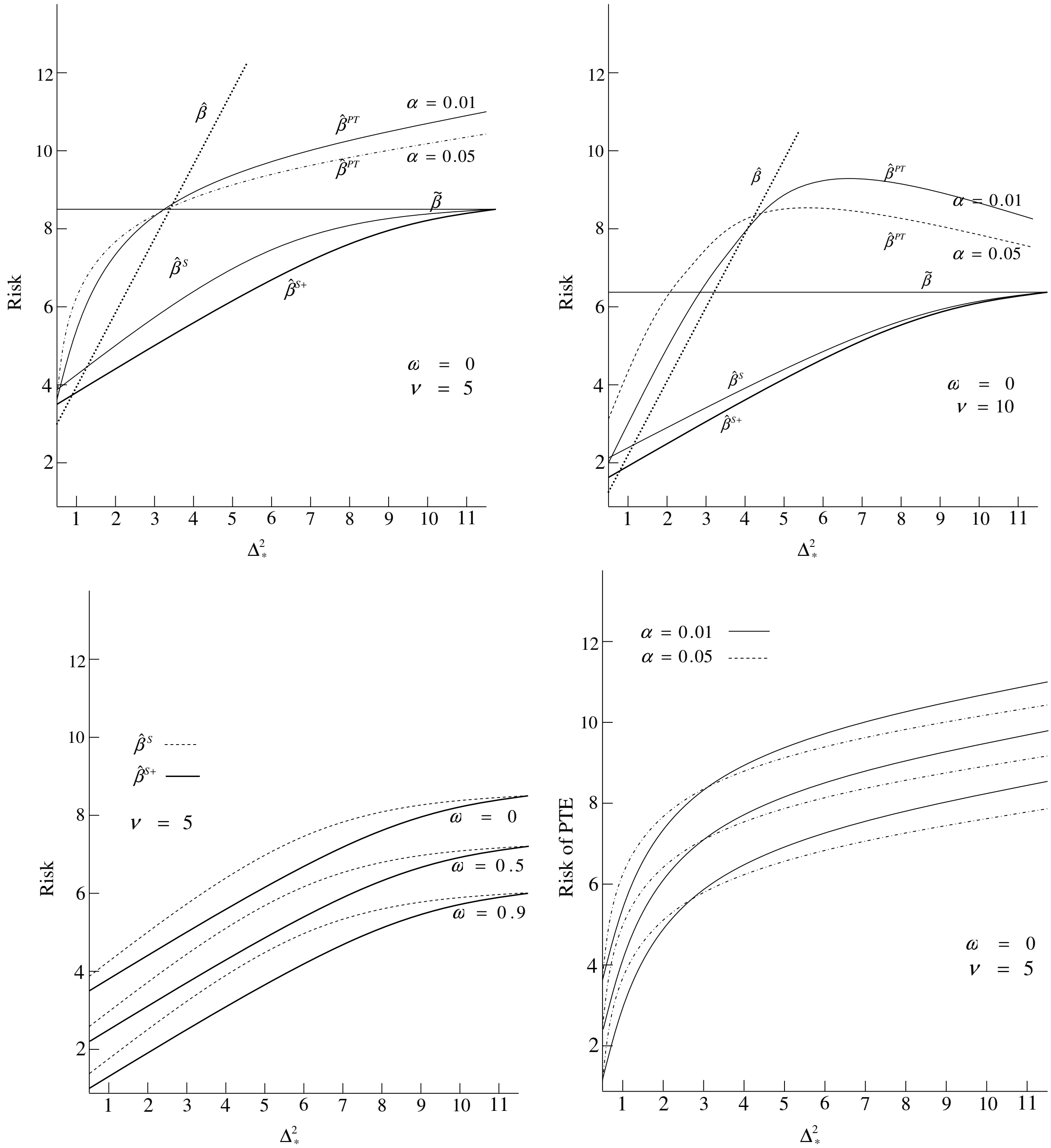


***Figure 1.*** *Risk Performance*

Additionally, we explore the risk performance of the proposed shrinkage estimators, providing a comparison with conventional methods based on the closeness to the null hypothesis. This comparison allows us to highlight the strengths and potential limitations of the shrinkage-based approach in practical applications, offering insights into how the estimators behave under different settings. Ultimately, this paper contributes to the understanding of shrinkage estimators in Bayesian regression, emphasizing their utility in improving model robustness, especially when facing data with irregularities or outliers.

***Acknowledgments*** We would like to thank two anonymous reviewers for their constructive comments that led us to add many details and improve the presentation.

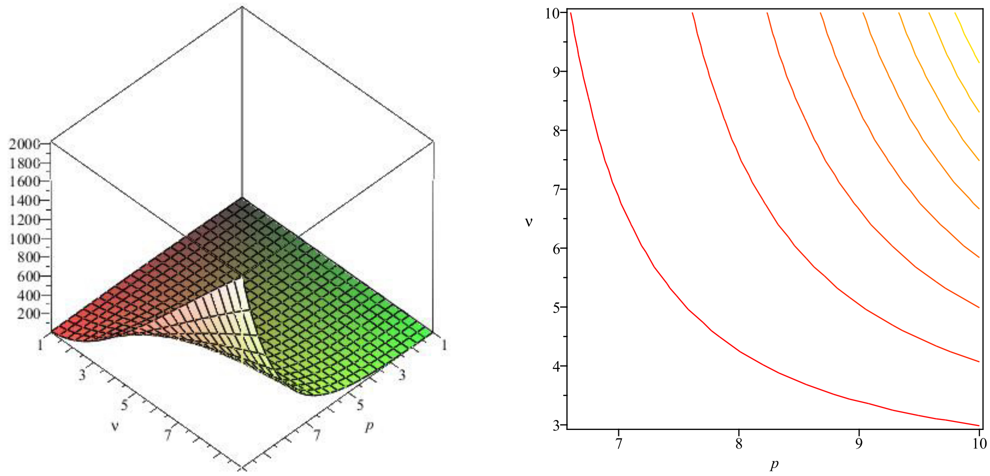


***Figure 2.*** *Visualization of the performance of* $\phi^{-1}$

***Funding/Financial Disclosure*** The authors have no received any financial support for the research, authorship, or publication of this study.

***Ethics Committee Approval and Permissions*** The work does not require ethics committee approval and any private permission.

***Conflict of Interests*** The authors stated that there are no conflict of interest in this article.

***Authors Contribution*** Authors contributed equally to the study.

## References

[1] Fang, K. T., & Zhang, Y. T. (1990). Generalized Multivariate Analysis. Springer, New York.

[2] Jozani, M. J., Marchand, E., & Parsian, A. (2006). On estimation with weighted balanced-type loss function. *Statistics and Probability Letters*, *76*, 773–780. https://doi.org/10.1016/j.spl.2005.10.026

[3] Zellner, A.(1994). Bayesian and non-Bayesian estimation using balanced loss functions. In: J.O. Berger and S.S. Gupta, (Eds.), Statistical Decision Theory and Methods V, Springer, New York, pp. 337–390.

[4] Gomez-Deniz, E. (2008). A generalization of the credibility theory obtained by using the weighted balanced loss function. *Insurance: Mathematics and Economics*, *42*, 850-854. https://doi.org/10.1016/j.insmatheco.2007.09.002

[5] Dey, D., Ghosh, M., & Strawderman, W. E. (1999). On estimation with balanced loss function. *Statistics and Probability Letters*, *45*, 97–101. https://doi.org/10.1016/S0167-7152(99)00047-4

[6] Evans, M., & Jang, G. H. (2011). Inference from prior-based loss function. Technical report No. 1104, University of Toronto.

[7] Saleh, A. K. Md. E. (2006). Theory of Preliminary Test and Stein-type Estimation with Applications. John Wiley, New York.

[8] Arashi, M., Saleh, A. K., Ehsanes, A. K. Md., & Tabatabaey, S. M. M. (2013). Regression model with elliptically contoured errors. *Statistics*, *47*(6), 1266–1284. https://doi.org/10.1080/02331888.2012.694442

[9] Cellier, D., Fourdrinier, D., and Strawderman, W. E. (1995). Shrinkage positive rule estimators for spherically symmetrical distributions. *Journal of multivariate analysis*, *53*2, 194–209. https://doi.org/10.1006/jmva.1995.1032

[10] Khan, S. & Saleh, A. K. Md. E. (1997). Shrinkage Pre-Test Estimator of the Intercept Parameter for a Regression Model with Multivariate Student-t Errors. *Biometrical Journal*, *39*2, 131–147. https://doi.org/10.1002/bimj.4710390202

[11] Khan, S. & Saleh, A. K. Md. E. (1998). Comparison of estimators of means based on p-samples from multivariate student-t population. *Communications in Statistics-Theory and Methods*, *27*1 193–210. https://doi.org/10.1080/03610929808832660

[12] Khan, S. & Saleh, A. K. Md. E. (2001). On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. *Statistical Papers*, *42*(4), 451–473.

[13] Khan, S., Hoque, Z., & Saleh, A. K. Md. E. (2002). Estimation of the slope parameter for linear regression model with uncertain prior information. *Journal of Statistical Research*, *36*, 55–73.

[14] Khan, S. (2005). Estimation of parameters fo the simple multivariate linear models with Student-t errors. *Journal of Statistical Research*, *39*(2), 71–86.

[15] Khan, S. (2008). Shrinkage Estimators of Intercept Parameters of Two Simple Regression Models with Suspected Equal Slopes. *Communications in Statistics—Theory and Methods*, *37*, 247–260. https://doi.org/10.1080/03610920701648961

[16] Saleh, A. K. Md. E & Kibria, B. M. G. (2011). On Some Ridge Regression Estimators: A Nonparametric Approach. *Journal of Nonparametric Statistics*, *23*3, 819–851. https://doi.org/10.1080/10485252.2011.567335

[17] Saleh, A. K. Md. E & Sen, P. K. (1978). Nonparametric Estimation of Location Parameters after a Preliminary Test on Regression. *The Annals of Statistics*, *6*, 154–168.

[18] Stein, C. (1956). The admissibility of Hotelling's T2-test. *The Annals of Mathematical Statistics*, *27*, 616–623.

[19] Tabatabaey, S. M. M., Saleh, A. K. Md. E.,& Kibria, B. M. G. (2004). Simultaneous estimation of regression parameters with spherically symmetric errors under possible stochastic constrains. *International Journal of Statistical Sciences*, *3*, 1–20.

[20] Jeffreys, H. (1961). Theory of Probability, Oxford: Clarendon, 1961.

[21] Anderson, T. W., Fang, K. T., & Hsu, H. (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. *The Canadian Journal of Statistics*, *14*, 55–59. https://doi.org/10.2307/3315036

*E-ISSN: 2564-7873*

[22] Arashi, M. (2012). Preliminary test and Stein estimators in simultaneous linear equations. *Linear Algebra and its Applications*, *436*(5), 1195–1211. https://doi.org/10.1016/j.laa.2011.07.036

[23] Srivastava, M., & Bilodeau, M. (1989). Stein estimation under elliptical distribution. *Journal of Multivariate Analysis*, *28*, 247–259. https://doi.org/10.1016/0047-259X(89)90108-5

[24] Khan, S. (2000). Improved estimation of the mean vector for student-t model. *Communications in Statistics-Theory and Methods*, *29*3, 507–527. https://doi.org/10.1080/03610920008832499

[25] Arashi, M., Saleh, A. K. Md. E., & Tabatabaey, S. M. M. (2010). Estimation of parameters of parallelism model with elliptically distributed errors. *Metrika*, *71*, 79-100. https://doi.org/10.1007/s00184-008-0203-6

[26] Saleh, A. K. Md. E., Arashi, M., & Tabatabaey, S. M. M. (2014). Statistical Inference for Models with Multivariate t-Distributed Errors, John Wiley, New Jersey.