Preliminary test and Stein estimations in simultaneous linear equations

M. Arashi

Faculty of Mathematics, Shahrood University of Technology, P.O. Box 316-361995161, Shahrood, Iran

ARTICLE INFO

Article history:
Received 23 February 2011
Accepted 25 July 2011
Available online 13 October 2011

Submitted by R.A. Brualdi

AMS classification:
Primary 62H12
Secondary 62F30

Keywords:
Balanced loss function
Elliptically contoured distribution
Preliminary test estimator
Stein-type shrinkage estimator
Positive-rule shrinkage estimator
Seemingly unrelated regression
Stochastic constraints

ABSTRACT

In modeling of an economic system, there may occur some stochastic constraints, that can cause some changes in the estimators and their respective behaviors. In this approach we formulate the simultaneous equation models into the problem of estimating the regression parameters of a multiple regression model, under elliptical errors. We define five different sorts of estimators for the vector-parameter. Their exact risk expressions are also derived under the balanced loss function. Comparisons are then made to clarify the performance of the proposed estimators. It is shown that the shrinkage factor of the Stein estimator is robust with respect to departures from normality assumption.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction and the methodology

A seemingly unrelated regression (SUR) system comprises several individual relationships that are linked by the fact that their disturbances are correlated. Such models have found many applications. For example, demand functions can be estimated for different households (or household types) for a given commodity. The correlation among the equation disturbances could come from several sources such as correlated shocks to household income. Alternatively, one could model the demand of a household for different commodities, but adding-up constraints leads to restrictions on the parameters of different equations in this case. On the other hand, equations explaining some phenomenon in different cities,
states, countries, firms or industries provide a natural application as these various entities are likely
to be subject to spillovers from economy-wide or worldwide shocks. There are two main motivations
for use of SUR. The first one is to gain efficiency in estimation by combining information on different
equations. The second motivation is to impose and/or test restrictions that involve parameters in
different equations. The second motivation is to impose and/or test restrictions that involve parameters in
different equations.

As a prelude to defining a SUR system, let

$$y_i^* = X_i^* B_i + \epsilon_i^*$$  \hspace{1cm} (1.1)

be the ith equation of an M equation regression system with $y_i^*$ a $T \times 1$ vector of observations on
the ith "dependent" variable, $X_i^*$ a $T \times p_i$ matrix with rank $l_i$, of observations on $l_i$ "independent"
non-stochastic variables, $B_i$ a $p_i \times 1$ vector of regression coefficients and $\epsilon_i^*$, a $T \times 1$ vector of random
error terms, each with mean zero. The system of which (1.1) is an equation may be written as:

$$\begin{bmatrix}
  y_1^* \\
  y_2^* \\
  \vdots \\
  y_M^*
\end{bmatrix} =
\begin{bmatrix}
  X_1^* & 0 & \cdots & 0 \\
  0 & X_2^* & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & X_M^*
\end{bmatrix}
\begin{bmatrix}
  B_1 \\
  B_2 \\
  \vdots \\
  B_M
\end{bmatrix} +
\begin{bmatrix}
  \epsilon_1^* \\
  \epsilon_2^* \\
  \vdots \\
  \epsilon_M^*
\end{bmatrix}$$  \hspace{1cm} (1.2)

The disturbance vector in (1.2) is assumed to have the following variance–covariance matrix:

$$V^* = \begin{bmatrix}
  \sigma_{11} I & \sigma_{12} I & \cdots & \sigma_{1M} I \\
  \sigma_{21} I & \sigma_{22} I & \cdots & \sigma_{2M} I \\
  \vdots & \vdots & \ddots & \vdots \\
  \sigma_{M1} I & \sigma_{M2} I & \cdots & \sigma_{MM} I
\end{bmatrix} = \Sigma^* \otimes I_T$$  \hspace{1cm} (1.3)

for $\Sigma^* \in S(M)$ where $S(M)$ denotes the set of all positive definite matrices of order $(M \times M)$ and $I_T$ is
the identity matrix of order $T \times T$ and $\sigma_{ii} = E(\epsilon_i^* \epsilon_i^*)$ for $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, M$. Zellner
[30,29] formulated a SUR model to solve a set of simultaneous linear equations technically.

Working with the model (1.2), when $M \geq 3$ leads to face feasible type least squares (LS) estimators
in small sample problems. Otherwise we should consider the asymptotic performance of estimators.
This takes part because of the unknown $\Sigma^*$. In other words, it is not possible to survey on small
sample properties of the LS estimators. Now we take an alternative point of view into consideration
to somehow solve this problem.

One important set of hypotheses are checking aggregation bias, i.e., testing the null hypothesis
$B_1 = B_2 = \cdots = B_M$. Now consider a situation of a reduced form say, a system of $M$ seemingly
uncorrelated equations in which $B_1 = B_2 = \cdots = B_M = B, p_1 = \cdots = p_M = p$, and the error
vector $\epsilon = (\epsilon_1, \ldots, \epsilon_M)'$ has the following covariance structure

$$\text{Cov}(\epsilon) = \sigma^2 \mathbf{V}, \quad \mathbf{V} =
\begin{bmatrix}
  \Sigma_1 & 0 & \cdots & 0 \\
  0 & \Sigma_2 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & \Sigma_M
\end{bmatrix},$$  \hspace{1cm} (1.4)

where in general $\Sigma_i, i = 1, \ldots, M$ is known and $\sigma^2$ is unknown. For practical use such as econometric
studies we may also assume that

$$E(\epsilon_1) = 0, \quad E(\epsilon_j) = \mu_j, \quad \text{for } j = 2, \ldots, M.$$  \hspace{1cm} (1.5)
In this regard, under a restricted situation, one may desire to check whether the following null hypothesis occurs or not.

\[ H_0 : E(\epsilon_1) = E(\epsilon_2) = \cdots = E(\epsilon_M) \text{, i.e., } H_0 : \mu_2 = \mu_3 = \cdots = \mu_M = 0. \]  

(1.6)

Then one can combine the models to get the following relation

\[
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_M
\end{bmatrix} = 
\begin{bmatrix}
  X_1 \\
  X_2 \\
  \vdots \\
  X_M
\end{bmatrix} B + 
\begin{bmatrix}
  \epsilon_1 \\
  \epsilon_2 \\
  \vdots \\
  \epsilon_M
\end{bmatrix},
\]

(1.7)

where \( y_i \) a \( n_i \times 1 \) vector of observations on the \( i \)th "dependent" variable, \( X_i \) a \( n_i \times p \) matrix with rank \( p \), of observations on \( p \) "independent" non-stochastic variables, \( B \) a \( p \times 1 \) vector of regression coefficients and \( \epsilon_i \), a \( n_i \times 1 \) vector of random error terms each with mean according to (1.5). Based on the representation (1.7), the full model can be rewritten as

\[ y = X\beta + \epsilon \]

(1.8)

where for \( n = \sum_{i=1}^{M} n_i \), \( y \) is a \( (n \times 1) \) vector of responses and \( X \) is an \( n \times (p + q) \), \( q = \sum_{i=2}^{M} n_i \), non-stochastic matrix represented as

\[
X = 
\begin{bmatrix}
  X_1 & 0 & \cdots & 0 \\
  0 & I_{n_2 \times n_2} & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & I_{n_M \times n_M}
\end{bmatrix}, \quad \beta_{(p+q) \times 1} = 
\begin{bmatrix}
  B \\
  X_2 B + \mu_2 \\
  X_3 B + \mu_3 \\
  \vdots \\
  X_M B + \mu_M
\end{bmatrix},
\]

and

\[
\epsilon = 
\begin{bmatrix}
  \epsilon_1 \\
  \epsilon_2 - \mu_2 \\
  \epsilon_3 - \mu_3 \\
  \vdots \\
  \epsilon_M - \mu_M
\end{bmatrix}.
\]

(1.9)

Thus the error term has the mean equal to zero, and the covariance structure given by (1.4). More important, the underlying restriction is subjected to

\[ H_0 : H\beta = \mu = 
\begin{bmatrix}
  \mu_2 \\
  \mu_3 \\
  \vdots \\
  \mu_M
\end{bmatrix} = 0. \]

(1.10)

where
This can be interpreted as a SUR like model with a minor difference in which all \( \Sigma_i, i = 1, \ldots, M \) are known and a common face \( \sigma^2 \) still is unknown. Using such models as defined by (1.8), we can accomplish an examination and analysis of one’s own thoughts and feelings (prior information via introspection). Also one may have prior information from a previous sample, which usually makes some relations through stochastic sub-space restrictions. Furthermore, combining stochastic restrictions with Zellner’s [30] seemingly unrelated estimators, we can demonstrate good performance of estimators using MSE criterion comparatively. Eventually, considering stochastic sub-space restrictions, we can apply the statistical models to the broad range of microeconomic models. For complete review on why we deal with stochastic constraints under the case \( M = 2 \), and the applications see Theil and Goldberger [27].

We organize our paper as follows: Some preliminary utilities are included in Section 2. Section 3, contains the estimation and the test of hypothesis along with proposed estimators of \( \beta \). Section 4 deals with the bias, risk expressions of the proposed estimators while the analysis of the risks and comparison results are presented in Section 5. Some remarks that make the ending are included in Section 6 and formal proofs are contained in Section 7.

2. Preliminaries and some notations

Importance of any estimation problem is boosted if we can furnish our driven estimators with good performance in the sense of having smaller risk. In this case, the loss function under study plays determinant role. However, selecting objective or subjective points of view changes the results, it is utterly important to take reasonable and practical losses into account.

Let \( \beta^* \) denote any estimator of \( \beta \); then the quadratic loss function which reflects the goodness of fit of the model is \( (X \beta^* - y)'(X \beta^* - y) \). Similarly, the precision of estimation of \( \beta^* \) is measured by the weighted loss function \( (\beta^* - \beta)'X'X(\beta^* - \beta) \). Generally, both of the previous criteria are used to judge the performance of any estimator. Throughout this paper, we shall consider the estimation problem through the following loss function

\[
L_{\omega, \delta_0}^W(\delta; \beta) = \omega r(\|\beta\|^2) (\delta - \delta_0)'W(\delta - \delta_0) + (1 - \omega)r(\|\beta\|^2) (\delta - \beta)'W(\delta - \beta),
\]

where \( \omega \in [0, 1], r(.) \) is a positive weight function, \( W \) is a weight matrix, and \( \delta_0 \) is a target estimator. This loss is pioneered by Jozani et al. [12] inspiring by Zellner’s [28] balanced loss function. This loss function takes both goodness of fit and error of estimation into account. The \( \omega r(\|\beta\|^2) (\delta - \delta_0)'W(\delta - \delta_0) \) part of the loss is analogous to a penalty term for lack of smoothness in non-parametric regression. The weight \( \omega \) in (1.2) calibrates the relative importance of these two criteria. For the case \( \omega = 0 \), we will simply write \( L_0^W(\delta; \beta) \) as the quadratic loss function. Of course, duty of the weight function \( r(.) \) is clearly apparent in the Bayesian viewpoint. In this paper, we take it into consideration for the sake of generality. As it can be seen later, the structure of \( r(.) \) does not alter the whole superiority conclusions.

**Lemma 2.1.** Assume \( h_i : \mathbb{R}^n \to \mathbb{R}^{p+q}, i = 1, 2 \) are measurable functions.

(i) The estimator \( \delta_0(X) + (1 - \omega) h_1(X) \) dominates \( \delta_0(X) + (1 - \omega) h_2(X) \) under the balanced loss function \( L_{\omega, \delta_0}^W(\delta; \beta) \) if and only if \( \delta_0(X) + h_1(X) \) dominates \( \delta_0(X) + h_2(X) \) under the quadratic loss function \( L_0^W(\delta; \beta) \).
(ii) Suppose the estimator $\delta_0 (X)$ has constant risk $\gamma$ under the quadratic loss function $L_0^{W} (\delta; \beta)$. Then $\delta_0 (X)$ is minimax under the balanced loss function $L_0^{W, \omega, \delta_0} (\delta; \beta)$ with constant (and minimax) risk $(1 - \omega) \gamma$ if and only if $\delta_0 (X)$ is minimax under the quadratic loss function $L_0^{W} (\delta; \beta)$ with constant (and minimax) risk $\gamma$.

In a precise setup, we assume $\epsilon$ is distributed according to the law belonging to the class of elliptically contoured distributions (ECDs), $\mathcal{E}_n (0, \sigma^2 \mathbf{V}, \psi)$, where $\mathbf{V}$ is defined by (1.4) with the following characteristic function

$$
\phi_\epsilon (t) = \psi \left( \sigma^2 t' \mathbf{V} t \right)
$$

(2.2)

for some functions $\psi : [0, \infty) \rightarrow \mathbb{R}$ say characteristic generator [9]. If $\epsilon$ possesses a density, then it can be represented as [7]

$$
f (\epsilon) = d_n |\mathbf{V}|^{-\frac{n}{2}} \mathbf{g}_n \left[ \frac{1}{\sigma^2} \epsilon' \mathbf{V}^{-1} \epsilon \right] = \int_0^\infty W(\tau) \mathcal{N}_n \left( 0, \tau^{-1} \sigma^2 \mathbf{V} \right) d\tau
$$

$$
= \int_0^\infty W(\tau) \left( \frac{1}{\sqrt{2\pi \sigma^2 \tau^{-1}}} \right)^n |\mathbf{V}|^{-\frac{n}{2}} e^{\frac{1}{2} \sigma^2 \tau^{-1} \epsilon' \mathbf{V}^{-1} \epsilon} d\tau
$$

(2.3)

where $d_n$ is the normalizing constant and

$$
W(\tau) = (2\pi)^{\frac{n}{2}} |\sigma^2 \mathbf{V}|^{\frac{1}{2}} \tau^{-\frac{n}{2}} \mathcal{L}^{-1} \{ f(s) \},
$$

$\mathcal{L}^{-1} \{ f(s) \}$ denotes the inverse Laplace transform of $f(s)$ for $s = \epsilon' \mathbf{V}^{-1} \epsilon / 2\sigma^2$. For details on the properties of Laplace transform and its inverse.

On integrating $f(.)$ over $\mathbb{R}^n$, $W(\tau)$ integrates to 1. Thus for nonnegative function $W(\tau)$, it is a density and can be interpreted as a scale mixture of normal distribution [19]. Then we will use the notation $\epsilon \sim \mathcal{E}_n (0, \sigma^2 \mathbf{V}, \mathbf{g}_n)$. The condition

$$
\int_0^\infty \tau^{\frac{n}{2} - 1} \mathbf{g}_n (\tau) d\tau < \infty
$$

guarantees $\mathbf{g}_n (\tau)$ is a density generator. Also $\mathbf{g}_n$ and $\psi$ determine each other for each specific member of this family. In addition, if $\mathbf{g}_n$ does not depend on $n$, we use the notation $\mathbf{g}$ instead.

The mean of $\epsilon$ is the zero-vector and the covariance-matrix of $\epsilon$ is

$$
E(\epsilon' \epsilon) = -2\psi' (0) \mathbf{V} = \sigma^2 \mathbf{V}, \quad \text{where} \quad \sigma^2 = -2\psi' (0) \sigma^2.
$$

(2.4)

Some of the well known members of the class of ECDs are the multivariate normal, Kotz Type, Pearson Type VII, multivariate Student’s $t$, multivariate Cauchy, and generalized slash distributions. Dating back to Kelker [13], there are many known results concerning ECDs including particular mathematical properties and statistical inference. Among others, Cambanis et al. [6], Muirhead [19] Fang et al. [9], Fang and Zhang [8] and Gupta and Varga [10] are studied ECDs comprehensively.

The gist of this paper is the estimation of the regression parameters, $\beta$ when it is suspected that $\beta$ generally may belong to the sub-space defined by $\mathbf{H} \beta = \mathbf{h}$ (non sample information) where $\mathbf{H}$ is a $q \times (p + q)$ matrix of constants and $\mathbf{h}$ is a $q$-vector of known constants with focus on the Stein-type estimators of $\beta$ in addition to preliminary test estimator (PTE). For the sake of simplicity, throughout we assume $\mathbf{h} = \mathbf{0}$, as formulated in previous section.

Recent book of Saleh [20] presents an overview on the topic under normal as well as non-parametric theory covering many standard models. No systematic work has been done so far when error-distribution is elliptically contoured, $\mathcal{E}_n (0, \sigma^2 \mathbf{V}, \mathbf{g})$ in SUR models under a balanced loss function. Arashi [3] considered the problem under study in multiple regression models.
3. Estimation and test of hypothesis

Using standard conditions, it is well-known that the LSE of $\beta$ is

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y = C^{-1}X'V^{-1}y, \quad C = X'V^{-1}X$$

Its distribution is $\mathcal{N}_p(\beta, \sigma^2C^{-1}, g)$. Thus the 1st moment is the zero vector and the 2nd central moment is $E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) = \sigma^2I$. Similarly the estimate of the $\sigma^2$ is

$$\sigma^2 = \frac{1}{n}(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})$$

Then we conclude that $S^2 = \frac{n\sigma^2}{n-p}$ is an unbiased estimator of $\sigma^2$. For test of $H\beta = 0$, we first consider the restricted estimator given by

$$\hat{\beta} = \hat{\beta} - C^{-1}H'V_1H\hat{\beta}, \quad V_1 = [HC^{-1}H']^{-1}. \quad (3.3)$$

Consequently $\hat{\beta} \sim \mathcal{N}_q(\beta - \delta, \sigma^2V_2, g)$, where $\delta = C^{-1}H'V_1H\beta$ and $V_2 = C^{-1}(I_p - H'V_1HC^{-1})$. Under $H_0$, the following estimator is unbiased for $\sigma^2$.

$$S_0^2 = \frac{(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})}{n - p + q}, \quad (3.4)$$

from least square theory. Now we consider the linear hypothesis $H\beta = 0$ and obtain the test statistic for the null hypothesis $H_0 : H\beta = 0$.

**Theorem 3.1.** Let $\Omega = \{ (\beta, \sigma, V) : \beta \in \mathbb{R}^p, \sigma \in \mathbb{R}^+, V \in S(n) \}$ and $\omega = \{ (\beta, \sigma, V) : \beta \in \mathbb{R}^q, H\beta = 0, \sigma \in \mathbb{R}^+, V \in S(n) \}$. Suppose that in the model (1.8), $\epsilon \sim \mathcal{N}(0, \sigma^2V, g)$. Moreover, $y^2g(y)$ has a finite positive maximum $y_g$. Then the likelihood ratio test statistic for testing $H_0 : \mu = 0$ is given by

$$L_n = \frac{\hat{\beta}'H'V_1H\hat{\beta}}{qS^2}. \quad (3.5)$$

having the following generalized non-central F-distribution

$$g_{q,m}(L_n) = \sum_{r \geq 0} \frac{\left( \frac{q}{m} \right)^{\frac{r}{2} + \frac{1}{2}} L_n^{\frac{r}{2} + \frac{1}{2}} K_{(r)}^{(\Delta^2)} K_{(r)}^{(\Delta^2)}}{r! B \left( \frac{q + 2r}{2}, \frac{m}{2} \right)} \left( 1 + \frac{q}{m} L_n \right)^{\frac{1}{2} \left( q + m + 2r \right)}$$

where $m = n - p, \Delta^2 = \frac{\theta^2}{\sigma^2}$ for $\theta = \beta'V_1H\beta$, and the distribution of $K_{(r)}^{(\Delta^2)}$ becomes

$$K_{(r)}^{(\Delta^2)} = \int_0^\infty W(\tau) \left( -\psi'(0)\tau \Delta^2 \right)^r e^{\psi'(0)\tau \Delta^2} d\tau. \quad (3.6)$$

1 The test statistics is the same as under normality assumption. Hence it is robust in this sense.
Corollary 3.1. Under $H_0$, the pdf of $L_n$ is given by

$$g^*_{q,m}(L_n) = \frac{\binom{q}{2}}{B\left(\frac{q}{2}, \frac{m}{2}\right)} \left(1 + \frac{q}{m}L_n\right)^{\frac{1}{2}(q+m)}$$

which is the central $F$-distribution with $(q, m)$ degrees of freedom.

Now, consider the calculations of the probability of that $L_n \leq F_\alpha$, which gives the power function of the test as

$$G^*_{q,m}(F_\alpha; \Delta_\gamma^2) = \sum_{r \geq 0} \frac{1}{r!} K^{(\gamma^2)}(r) I_x \left[\frac{1}{2}(q + 2r), \frac{m}{2} \right]$$

where $x = \frac{qF_\alpha}{(m + qF_\alpha)}$ and $K(x, a)$ is the incomplete Beta-function,

$$K(x, a, b) = \frac{1}{B(a, b)} \int_0^x u^{a-1} (1 - u)^{b-1} du.$$  

The function (3.7) stands for the power function at $\alpha$-level of significance and may be called the generalized non-central $F$-distribution cdf of the statistic $L_n$. Similarly, the cdf of a generalized non-central Chi-square distribution with $\gamma$ d.f. may be written as

$$H^\gamma_{\gamma + 2r}(x; \delta^2) = \sum_{r \geq 0} \frac{1}{r!} K^{(\gamma^2)}(r) H_{\gamma + 2r}(x; 0)$$

where $H_{\gamma + 2r}(x; 0)$ is the cdf of Chi-square distribution with $\gamma + 2r$ d.f.

In many practical situations, along with the model one may suspect that $\beta$ belongs to the sub-space defined by $\mu = H\beta = 0$. In such situation one combines the estimate of $\beta$ and the test-statistic to obtain 3 or more estimators as in Saleh [20], in addition to the unrestricted and the restricted estimators of $\beta$. First we consider the preliminary test estimator (PTE) of $\beta$ which is a convex combination of $\tilde{\beta}$ and $\hat{\beta}$:

$$\hat{\beta}^{PT} = \tilde{\beta} I(L_n \geq F_\alpha) + \hat{\beta} I(L_n < F_\alpha)$$

where $I(A)$ is the indicator function of the set $A$ and $F_\alpha$ is the upper $\alpha$th percentile of the central $F$-distribution with $(q, m)$ d.f. PTE was initiated by Bancroft [4,5] and extended by Han and Bancroft [11] and Saleh and Sen [21] in parametric and non-parametric setups respectively. The PTE has the disadvantage that it depends on $\alpha (0 < \alpha < 1)$, the level of significance and also it yields the extreme results, namely $\hat{\beta}$ and $\tilde{\beta}$ depending on the outcome of the test. Therefore we define Stein-type shrinkage estimator (SE) of $\beta$, as

$$\hat{\beta}^S = \tilde{\beta} + (1 - dL_n^{-1})(\tilde{\beta} - \hat{\beta})$$

$$= \tilde{\beta} - dL_n^{-1} (\tilde{\beta} - \hat{\beta}),$$

where

$$d = \frac{(q - 2)m}{q(m + 2)}$$

and

$$q \geq 3.$$  

The SE has the disadvantage that it has strange behavior for small values of $L_n$. Also, the shrinkage factor $(1 - dL_n^{-1})$ becomes negative for $L_n < d$. Hence we define a better estimator by positive-rule shrinkage estimator (PRSE) of $\beta$ as
4. Bias and risk of the estimators

The risk function for any estimator $\hat{\beta}^*$ of $\beta$ associated with (2.1) is defined as

$$ R_W^{\omega, \delta_0} (\beta^*; \beta) = E[L_{\omega, \delta_0} (\beta^*; \beta)]. \tag{4.1} $$

In this section, we determine the bias, and using the risk function (4.1) when $\delta_0 = \tilde{\beta}$, as the target estimator, and $W = C$, given by (3.1), evaluate the risks of the five different estimators understudy. For the case $\omega = 0$, we will simply write $R_W^{\omega} (\beta^*; \beta)$. First we consider bias expressions of the estimators. Directly

$$ b_1 = E[\tilde{\beta} - \beta] = 0, \quad b_2 = E[\hat{\beta} - \beta] = -\delta. $$

For the bias of PTE we have $2$  

$$ b_3 = E\left(\tilde{\beta}^{PT} - \beta\right) = -\delta G_{q+2,m}^{(2)} (F_\alpha; \Delta_\ast^2). $$

where $G_{q,m}^{(i)} (l; \Delta_\ast^2) = \sum_{r \geq 0} \frac{1}{r!} K_{(r+j-2)}^{(\Delta_\ast^2)} \left[ \frac{1}{2} (q + 2r), m \right]$. Also,

$$ b_4 = E\left(\tilde{\beta}^S - \beta\right) = -dq \delta E_{q+2, m}^{(2)} (X_{q+2}^{-1} (\Delta_\ast^2)). $$

where $E^{(j)} [X_{p+s}^{-1} (\Delta_\ast^2)] = \sum_{r \geq 0} \frac{1}{r!} K_{(r+j-2)}^{(\Delta_\ast^2)} (p + s - 2 + 2r)^{-1}$. For the final bias expression we have

$$ b_5 = E\left(\tilde{\beta}^S - \beta\right) - E\left[I(\mathcal{L}_n \leq d) (\tilde{\beta} - \beta)\right] + dE\left[\mathcal{L}_n^{-1} I(\mathcal{L}_n \leq d) (\tilde{\beta} - \beta)\right] $$

$$ = -dq \delta E_{q+2, m}^{(2)} (X_{q+2}^{-1} (\Delta_\ast^2)) + dq \delta E_{q+2, m}^{(2)} (d; \Delta_\ast^2) $$

$$ + \frac{qd}{q + 2} \delta E_{q+2, m}^{(2)} \left[ F_{q+2, m}^{-1} (\Delta_\ast^2) I \left( F_{q+2, m} (\Delta_\ast^2) \leq \frac{qd}{q + 2} \right) \right] $$

where

$$ E^{(j)} [F_{q+s,m}^{-1} (\Delta_\ast^2) I (F_{q+s,m} (\Delta_\ast^2) < c_l)] $$

$$ = \sum_{r \geq 0} \frac{1}{r!} K_{(r+j-2)}^{(\Delta_\ast^2)} (q + s)(q + s - 2 + 2r)^{-1} I \left[ \frac{q + s - 2 + 2r}{2}, \frac{m}{2} \right] $$

and $x' = \frac{qd}{m + d q}$.  

Note that as the non-centrality parameter $\Delta_\ast^2 \rightarrow \infty$, $b_1 = b_3 = b_4 = b_5 = 0$ while $b_2$ becomes unbounded. However, under $H_0 : \mu = 0$, all estimators are unbiased since $\delta = 0$.  

For the risks of the estimators, taking $R_{\omega, \tilde{\beta}}^{\omega} (\beta; \beta)$ given by (4.1), we have

$$ R_{\omega, \tilde{\beta}}^{\omega} (\hat{\beta}; \beta) = p \sigma^2_{\epsilon} (1 - \omega) r (\| \beta \|^2). \tag{4.2} $$

$^2$ To save the space, we just bring the final simplified statements. Detailed computations are ready in request.
Afterward, using the fact that $V_1^\frac{1}{2}H\hat{\beta} \sim \mathcal{E}_q(V_1^\frac{1}{2}H\beta, \sigma^2I_q, g)$, it can be concluded that

$$R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta) = -q\sigma^2 \epsilon^2 r(\|\beta\|^2) + R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta) + (1 - \omega) r(\|\beta\|^2) \theta,$$

(4.3)

where $\theta = \delta^T C \delta = \beta^T H^T V_1 H \beta$. Note that $R = C_1^{-\frac{1}{2}} H^T V_1 H C_1^{-\frac{1}{2}}$ is a symmetric idempotent matrix of rank $q \leq p$. Thus, there exists an orthogonal matrix $Q$ such that $Q R Q^T = \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}$. Now we define random variable $w = Q C_1^{-\frac{1}{2}} \tilde{\beta}$, then $w \sim \mathcal{E}_p(\eta, \sigma^2 I_p, g)$, where $\eta = Q C_1 \beta$. Partitioning the vectors $w = (w_1', w_2')'$ and $\eta = (\eta_1', \eta_2')'$ where $w_1$ and $w_2$ are sub-vectors of order $q$ and $p - q$ respectively, we can represent the test statistic $L_n$ given by (3.5) as

$$L_n = \frac{w_1^T w_1}{q S^2}, \quad \theta = \eta_1^T \eta_1.$$

Consequently, for the risk of PTE, noting that $\tilde{\beta} = C^{-1} H^T V_1 H C^{-\frac{1}{2}} w$, simplifying the equations we can write

$$R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta) = R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta) - (1 - 2\omega) r(\|\beta\|^2) E[w_1^T w_1(\mathcal{L}_n \leq F_{\alpha})]$$

$$+ 2(1 - \omega) r(\|\beta\|^2) \eta_1^T E[w_1^T w_1(\mathcal{L}_n \leq F_{\alpha})]$$

$$= R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta) - (1 - 2\omega) q \sigma^2 r(\|\beta\|^2) C_{q+2, m}^{(1)}(F_{\alpha}; \Delta^2)$$

$$+ 2\theta(1 - \omega) r(\|\beta\|^2) [2 G_{q+2, m}^{(2)}(F_{\alpha}; \Delta^2) - G_{q+4, m}^{(2)}(F_{\alpha}; \Delta^2)].$$

(4.4)

Similarly, for the risk of SE, after simplifying the equations we have

$$R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta) = \omega r(\|\beta\|^2) d^2 E[\mathcal{L}_n^{-1} w_1^T w_1] + R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta)$$

$$- 2d(1 - \omega) r(\|\beta\|^2) E[\mathcal{L}_n^{-1} (w_1^T w_1 - \eta_1^T w_1)]$$

$$+ d^2 (1 - \omega) r(\|\beta\|^2) E[\mathcal{L}_n^{-2} w_1^T w_1]$$

$$= R_{\omega, \hat{\beta}}^C(\hat{\beta}; \beta) + q r(\|\beta\|^2) \left\{ d^2 \omega - 2d(1 - \omega) \right\} E^{(1)}[\mathcal{X}_q^{s-2}(\Delta^2)]$$

$$+ d^2 (1 - \omega) E^{(1)}[\mathcal{X}_q^{s-4}(\Delta^2)] + \theta r(\|\beta\|^2)$$

$$\times \left\{ d^2 \omega - 2d(1 - \omega) \right\} E^{(2)}[\mathcal{X}_q^{s-2}(\Delta^2)]$$

$$- 2d(1 - \omega) E^{(2)}[\mathcal{X}_q^{s-2}(\Delta^2)] + d^2 (1 - \omega) E^{(2)}[\mathcal{X}_q^{s-4}(\Delta^2)],$$

(4.5)

where $E^{(i)}[\mathcal{X}_q^{s-4}(\Delta^2)] = \sum_{r \geq 0} \frac{1}{r!} K_{(r+j-2)}(q + s - 2 + 2r)^{-1}(q + s - 4 + 2r)^{-1}$. Finally, for the risk of PRSE we can obtain
\[ R_{\omega, \hat{\beta}}^C (\hat{\beta}^{S+}; \beta) = R_{\omega, \tilde{\beta}}^C (\hat{\beta}^S; \beta) - r \left( \|\beta\|^2 \right) E[(1 - d\mathcal{L}_n^{-1})^2 I(\mathcal{L}_n \leq d) w_i' w_1] \]
\[ \geq 2r \left( \|\beta\|^2 \right) E[(1 - d\mathcal{L}_n^{-1}) I(\mathcal{L}_n \leq d) (w_i' w_1 - \eta_1' w_1)] \]
\[ = R_{\omega, \tilde{\beta}}^C (\hat{\beta}^S; \beta) \]
\[ - \sigma^2 \epsilon \left[ qE^{(1)} \left( \frac{q}{q+2} F_{q+2,m}^{-1}(\Delta^2_\epsilon) \right)^2 I \left( F_{q+2,m}(\Delta^2_\epsilon) \leq \frac{q}{q+2} \right) \right] \]
\[ + \frac{\theta}{\sigma^2 \epsilon} E^{(2)} \left( \frac{q}{q+2} F_{q+2,m}^{-1}(\Delta^2_\epsilon) \right)^2 I \left( F_{q+2,m}(\Delta^2_\epsilon) \leq \frac{q}{q+2} \right) \]
\[ - 2\theta E^{(2)} \left( \frac{q}{q+2} F_{q+2,m}^{-1}(\Delta^2_\epsilon) \right) I \left( F_{q+2,m}(\Delta^2_\epsilon) \leq \frac{q}{q+2} \right) \],
\text{(4.6)}

where
\[ E^{(j)}[F_{q+s,m}^{-2}(\Delta^2_\epsilon) I(\mathcal{L}_n \leq \epsilon)] = \sum_{r \geq 0} \frac{(q+s)^2}{r!m} K^{(\Delta^2_\epsilon)}[(q + s - 2 + 2r)(q + s - 4 + 2r)]^{-1} \times I_k \left[ \frac{q + s - 4 + 2r}{2}, m + 4 \right]. \]

5. Risk analysis

Providing risk analysis of the underlying estimators with the weight matrix \( C \), by making use of the equations (4.2) and (4.3) the risk difference is given by
\[ D_1 = R_{\omega, \hat{\beta}}^C (\hat{\beta}; \beta) - R_{\omega, \tilde{\beta}}^C (\hat{\beta}; \beta) = r \left( \|\beta\|^2 \right) \left[ (1 - \omega)\theta - q \sigma^2 \epsilon \right]. \]

Then it can be directly considered that \( \hat{\beta} \) performs better than \( \tilde{\beta} \) say, \( \hat{\beta} \) dominates \( \tilde{\beta} (\hat{\beta} \geq \tilde{\beta}) \) provided that \( 0 \leq \theta \leq \frac{q \sigma^2 \epsilon}{1 - \omega} \), for \( \omega \neq 1 \). Also taking \( r \left( \|\beta\|^2 \right) = \beta' H' V_1 H \beta = \theta \) into account, gives the same result.

Comparing \( \hat{\beta}^{PT} \) versus \( \tilde{\beta} \), using risk difference,
\[ D_2 = R_{\omega, \hat{\beta}}^C (\hat{\beta}; \beta) - R_{\omega, \tilde{\beta}}^C (\hat{\beta}; \beta) \]
\[ = (1 - 2\omega) q \sigma^2 \epsilon r \left( \|\beta\|^2 \right) G_{q+2,m}^{(1)}(F_\alpha; \Delta^2_\epsilon) \]
\[ - 2\theta (1 - \omega) \left[ 2G_{q+2,m}^{(2)}(F_\alpha; \Delta^2_\epsilon) - G_{q+4,m}^{(2)}(F_\alpha; \Delta^2_\epsilon) \right]. \]
\text{(5.2)}

It can be followed that right hand side of (5.2) is nonnegative, i.e., \( \hat{\beta}^{PT} \geq \tilde{\beta} \) for \( \omega \neq 1 \) whenever
\[ \theta \leq \frac{(1 - 2\omega) q \sigma^2 \epsilon r \left( \|\beta\|^2 \right) G_{q+2,m}^{(1)}(F_\alpha; \Delta^2_\epsilon)}{2(1 - \omega) \left[ 2G_{q+2,m}^{(2)}(F_\alpha; \Delta^2_\epsilon) - G_{q+4,m}^{(2)}(F_\alpha; \Delta^2_\epsilon) \right]}. \]
\text{(5.3)}

Moreover, under \( H_0 : \mu = 0 \), because of \( \theta = 0 \), \( \hat{\beta}^{PT} \geq \tilde{\beta} \) for such values \( \omega \) that \( \omega \leq \frac{1}{2} \). Now we compare \( \hat{\beta} \) and \( \tilde{\beta}^{PT} \) by the risk difference as follows
The proof directly follows from Theorem 5.2 for the special case ω and thus using Theorem 5.1, \( \hat{\beta} \) uniformly dominates the unrestricted estimator \( \tilde{\beta} \) under the balanced loss function \( L_{\omega, \tilde{\beta}}(\hat{\beta}; \beta) \) for any \( \beta \in \mathbb{R}^p \).

Thus \( \hat{\beta}^{PT} \geq \hat{\beta} \) whenever

\[
\theta \geq \frac{q\sigma^2 \left[ 1 - (1 - 2\omega)G_{q+2,m}^{(1)}(F_{\alpha}; \Delta^2_\alpha) \right]}{(1 - \omega) \left[ 1 - 2G_{q+2,m}^{(2)}(F_{\alpha}; \Delta^2_\alpha) + G_{q+4,m}^{(2)}(F_{\alpha}; \Delta^2_\alpha) \right]}. \tag{5.5}
\]

and vice versa.

In order to determine the superiority of \( \hat{\beta}^S \) to \( \tilde{\beta} \), we give the following results. In fact we show that the shrinkage factor \( d \) of the Stein-type estimator is robust with respect to \( \beta \) and the unknown mixing distribution.

**Theorem 5.1.** Consider the model (1.8) where the error-vector belongs to the ECD, \( \varepsilon_n(0, \sigma^2 V, g) \). Then the Stein-type shrinkage estimator, \( \hat{\beta}^S \) of \( \beta \) given by

\[
\hat{\beta}^S = \hat{\beta} - d^* \mathcal{L}^{-1}_n(\hat{\beta} - \tilde{\beta})
\]

uniformly dominates the unrestricted estimator \( \tilde{\beta} \) with respect to the quadratic loss function \( L_{\omega, \tilde{\beta}}^S(\delta; \beta) \) and is minimax if and only if \( 0 < d^* \leq \frac{2m}{m+2} \). The largest reduction of the risk is attained when \( d^* = \frac{m}{m+2} \).

**Remark 5.1.** Consider the coefficient \( d \) given by (3.11). From \( q \geq 3 \), we get \( 0 < d = \frac{(q-2)m}{q(m+2)} < \frac{2m}{m+2} \) and thus using Theorem 5.1, \( \hat{\beta}^S \) in equation (3.10) uniformly dominates \( \tilde{\beta} \) on the whole parameter space under quadratic loss function.

**Theorem 5.2.** Suppose in the model (1.8), \( \varepsilon \sim \varepsilon_n(0, \sigma^2 V, g) \). Then the Stein-type shrinkage estimator

\[
\hat{\beta}_*^S = \hat{\beta} - d(1 - \omega) \mathcal{L}^{-1}_n(\hat{\beta} - \tilde{\beta}) \tag{5.6}
\]

uniformly dominates \( \tilde{\beta} \) under the balanced loss function \( L_{\omega, \tilde{\beta}}^C(\hat{\beta}; \beta) \).

The proof directly follows using Lemma 2.1 (i) and Theorem 5.1.

**Corollary 5.1.** Suppose in the model (1.8), \( \varepsilon \sim \varepsilon_n(0, \sigma^2 V, g) \). Then \( \hat{\beta}^S \geq \tilde{\beta} \) under the balanced loss function \( L_{\omega, \tilde{\beta}}^C(\hat{\beta}; \beta) \).

The proof directly follows from Theorem 5.2 for the special case \( \omega = 0 \).

**Lemma 5.1.** Suppose in the model (1.8), \( \varepsilon \sim \varepsilon_n(0, \sigma^2 V, g) \). Then the estimator \( \tilde{\beta} \) of \( \beta \) is minimax under the balanced loss function \( L_{\omega, \tilde{\beta}}^C(\hat{\beta}; \beta) \) given by (2.1).

The proof directly follows by knowing that \( \tilde{\beta} \) is minimax under quadratic loss function and applying Lemma 2.1 (ii).
Remark 5.2. Using Corollary 4.1 and Lemma 5.1, the Stein-type shrinkage estimator $\hat{\beta}^S$ of $\beta$ is minimax.

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

$$R_0^C(\hat{\beta}^S; \beta) = R_0^C(\hat{\beta}; \beta) + q\sigma^2 - \theta - dq^2\sigma^2 \left( (q - 2)E[\chi_{q+2}^{-4}(\Delta^2_\alpha)] \right)$$

$$+ \left[ 1 - \frac{(q + 2)\theta}{2q\sigma^2\Delta^2_\alpha} \right] (2\Delta^2_\alpha)E[\chi_{q+4}^{-4}(\Delta^2_\alpha)].$$

(5.7)

Under $H_0$, this becomes

$$R_0^C(\hat{\beta}^S; \beta) = R_0^C(\hat{\beta}; \beta) + q\sigma^2(1 - d) \geq R_0^C(\hat{\beta}; \beta),$$

while

$$R_0^C(\hat{\beta}; \beta) = R_0^C(\tilde{\beta}; \beta) - q\sigma^2 \leq R_0^C(\tilde{\beta}; \beta).$$

Therefore, $\hat{\beta} \geq \hat{\beta}^S$ under $H_0$ with the quadratic loss $L_0^C(\beta^*, \beta)$. Therefore using Lemma 2.1 (i), under $H_0$, $\hat{\beta} \geq \hat{\beta}^S$ with the balanced loss $L_{\omega, \beta}^C(\beta^*; \beta)$. However, as $\eta_1$ moves away from 0, $\theta$ increases and the risk of $\hat{\beta}$ becomes unbounded while the risk of $\hat{\beta}^S$ remains below the risk of $\hat{\beta}$; thus for similar reasons, $\hat{\beta}^S$ dominates $\hat{\beta}$ outside an interval around the origin under the balanced loss $L_{\omega, \beta}^C(\beta^*; \beta)$.

This scenario repeats when we compare $\hat{\beta}^S$ and $\hat{\beta}^{PT}$. Consider under $H_0$

$$R_0^C(\hat{\beta}^S; \beta) = R_0^C(\hat{\beta}^{PT}; \beta) + q\sigma^2[1 - \alpha - d] \geq R_0^C(\hat{\beta}^{PT}; \beta),$$

for all $\alpha$ such that $F_{q+2, m}(d, 0) \leq \frac{qd}{q+2}$. This means the estimator $\hat{\beta}^S$ does not always dominates $\hat{\beta}^{PT}$ under $H_0$. Thus, under $H_0$ with $\alpha$ satisfying $F_{q+2, m}(d, 0) \leq \frac{qd}{q+2}$ taking the balanced loss function we have $\hat{\beta} \geq \hat{\beta}^{PT} \geq \hat{\beta}^S \geq \hat{\beta}$. Then we compare, the risks of $\hat{\beta}^{S+}$ and $\hat{\beta}^S$. Subsequently, the risk difference is given by

$$D_5 = R_{\omega, \beta}^C(\hat{\beta}^{S+}; \beta) - R_{\omega, \beta}^C(\hat{\beta}; \beta)$$

$$= -\sigma^2 \left( qE(1) \left[ \left( 1 - \frac{qd}{q+2} F_{q+2, m}(\Delta^2_\alpha) \right)^2 I \left( F_{q+2, m}(\Delta^2_\alpha) \leq \frac{qd}{q+2} \right) \right] 

+ \frac{\theta}{\sigma^2} E(2) \left[ \left( 1 - \frac{qd}{q+2} F_{q+2, m}(\Delta^2_\alpha) \right)^2 I \left( F_{q+2, m}(\Delta^2_\alpha) \leq \frac{qd}{q+2} \right) \right] 

- 2\theta E(2) \left[ \left( 1 - \frac{qd}{q+2} F_{q+2, m}(\Delta^2_\alpha) \right) I \left( F_{q+2, m}(\Delta^2_\alpha) \leq \frac{qd}{q+2} \right) \right] \right) \right) \right).$$

The right hand side of the above equality is negative since for $F_{q+2, m}(\Delta^2_\alpha) \leq \frac{qd}{q+2}$, $(\frac{qd}{q+2} F_{q+2, m}(\Delta^2_\alpha) - 1) \geq 0$ and also the expectation of a positive random variable is positive. That for all $\beta$, $\hat{\beta}^{S+} \geq \hat{\beta}^S$.

Remark 5.3. The positive-rule shrinkage estimator $\hat{\beta}^{S+}$ of $\beta$ is minimax.
In the rest we continue the comparisons under $L_0^C(\beta^*; \beta)$. The results are the same for the balanced loss $L_{\omega, \beta}^C(\beta^*; \beta)$. To compare $\hat{\beta}$ and $\hat{\beta}_{S^+}$, first consider the case under $H_0$, i.e., $\eta_1 = 0$. In this case

$$R_0^C(\hat{\beta}_{S^+}; \beta) = R_0^C(\hat{\beta}; \beta) + q\sigma_\epsilon^2 \left\{ (1 - d) - E \left[ \left( 1 - \frac{qd}{q + 2} F_{q+2,m}(0) \right)^2 \right] \right\} \times 1 \left( F_{q+2,m}(0) \leq \frac{qd}{q + 2} \right) \geq R_0^C(\hat{\beta}; \beta),$$

since $E \left[ \left( 1 - \frac{qd}{q + 2} F_{q+2,m}(0) \right)^2 I(F_{q+2,m}(0) \leq \frac{qd}{q + 2}) \right] \leq E \left[ \left( 1 - \frac{qd}{q + 2} F_{q+2,m}(0) \right)^2 \right] = 1 - d$. Thus under $H_0$, $\hat{\beta} \geq \hat{\beta}_{S^+}$. However, as $\eta_1$ moves away from 0, $\theta$ increases and the risk of $\hat{\beta}$ becomes unbounded while the risk of $\hat{\beta}_{S^+}$ remains below the risk of $\hat{\beta}$; thus $\hat{\beta}_{S^+}$ dominates $\hat{\beta}$ outside an interval around the origin.

Now, we compare $\hat{\beta}_{S^+}$ and $\hat{\beta}^\text{PT}$. When $H_0$ holds, because $G_{q+2,m}(F_{\alpha}, 0) = 1 - \alpha$

$$R_0^C(\hat{\beta}_{S^+}; \beta) = R_0^C(\hat{\beta}^\text{PT}; \beta) + q\sigma_\epsilon^2 \left\{ 1 - \alpha - d - E \left[ \left( 1 - \frac{qd}{q + 2} F_{q+2,m}(0) \right)^2 \right] \right\} \times 1 \left( F_{q+2,m}(0) \leq \frac{qd}{q + 2} \right) \geq R_0^C(\hat{\beta}^\text{PT}; \beta),$$

for all $\alpha$ satisfying $E \left[ \left( 1 - \frac{qd}{q + 2} F_{q+2,m}(0) \right)^2 I(F_{q+2,m}(0) \leq \frac{qd}{q + 2}) \right] \leq 1 - \alpha - d$. Thus, $\hat{\beta}_{S^+}$ does not always dominates $\hat{\beta}^\text{PT}$ when the null-hypothesis $H_0$ holds.

Now consider the class of local alternatives $\{K_{(n)}\}$ defined by

$$K_{(n)} : H \beta = n^{-\frac{1}{2}} \xi.$$

Furthermore, following Saleh [20], consider the following regularity conditions hold

(i) $\max_{1 \leq i \leq n} x_i'(X'V^{-1}X)^{-1}x_i \to 0$ as $n \to \infty$, where $x_i'$ is the $i$th row of $X$;
(ii) $\lim_{n \to \infty} \{n^{-1} (X'V^{-1}X) \} = C$ for finite $C \in S(p)$.

Then using Theorem 7.8.3 from Saleh [20] in addition to Theorem 3.1 we obtain the following important result for the test statistic

$$\lim_{n \to \infty} P(C_n \leq \xi) = \mathcal{H}_q(\alpha; \beta^*, \xi).$$

Based on the above results, one can easily obtain the asymptotic distributional bias, risk and MSE matrix of each estimator under study using the following definition

$$G_p(\alpha) = \lim_{n \to \infty} P_{K_{(n)}} \\{ \sqrt{n}S^{-2}(\beta^* - \beta) \leq \alpha \}.$$

Then $b(\beta^*) = \int x dG_p(\alpha), M_{\omega, \beta}^W(\beta^*) = \int xx'dG_p(\alpha),$ and $K_{\omega, \beta}^W(\beta^*; \beta) = tr[WM(\beta^*)^W \omega, \beta]$; which have similar notations to those are given in this paper.

To end this section we display some graphical results for the risk of the proposed estimators. In this regard, we suppose that the error term in (1.8) follows the multivariate Student’s $t$ (MT) distribution denoted by $\epsilon \sim t_\nu(0, I_n, \nu)$. The graphs are displayed for $n = 30, p = 5, q = 3$ and different degrees of freedom $\nu = 5, 10$ and $\omega \in \{0, 0.5, 0.9\}$ to cover all possible situations. The corresponding necessary equations to compute the risk functions are given in Khan [16].
Fig. 1 shows the risk behavior of the proposed estimators for varying number of degrees of freedom and compares the risks of the PTE, SE and PRSE for selected number of degree of freedom ($\nu = 5$) and varying values $\omega$. The graphs in Fig. 1 reaffirms the analytical comparison covered in this section. More important as $\omega$ increases the risk values decrease. In other words, based on the structure of BLF, it confirms that if the model fit is good then the risk values are decreased as a natural consequence.

6. Concluding remarks

In this approach we considered an uncorrelated alternative model to a SUR system to study the performance of small sample properties of some estimators. However, the model has less application rather than the SUR model, it contains some special properties for practical goals. For example when we describe $M$ models with a minor similarity $\sigma^2$, the complement of $H_0$ may represent $M - i, i =
1, 2, . . . , M − 1 instability in the common performance. Under a sub-class of elliptical models for the error term, we proposed three sorts of improved estimators by combining the unrestricted and restricted estimators. We studied the performance of the proposed estimators in details. Beside its theoretical nature, we hope this work can bring new insights in defining one similar kind to study the small sample properties of estimators. To complete our purpose, from the presented results we may pay more attention to the following remarks.

- Since most of practical economic and financial studies contain outliers and extreme values, the proposed elliptical model covers all possible situations, even litter and heavier alternatives to the normal model.
- The proposed model is particularly formulated for considering stochastic constraints when the number of linear constraints is greater than two. It is fully discussed in Section 1.
- Involving the inverse Laplace transform of the density of error term, it exists if the following conditions are satisfied. (i) $f(t)$ is differentiable when $t$ is sufficiently large. (ii) $f(t) = o(t^{-m})$ as $t \to \infty, m > 1$. However, it is rather difficult to calculate the inverse Laplace transform of some functions, we can handle it for many density generators of elliptical densities. See Debnath and Bhatta (2007) for more details.
- From constructing improved estimators point of view, in the sense of having smaller risk, the general form of shrinkage and its positive part are considered and the performance are studied in details.
- Theorem 3.1 plays deterministic role in decision theoretic in elliptical models. It can be easily derived for heteroscedastic model as well as SUR model based on the results given in Zellner [30]. From Anderson et al. [2] and Anderson [1] it can be realized that the test statistic in elliptical models is the same as in normal models. However, its non-null distribution has not been discussed so far as the way in this theorem.
- As one of the important results, it is shown that the shrinkage factor of the Stein estimator is robust with respect to departures from normality assumption (Theorem 5.1). Thus practitioners can apply the shrinkage coefficient in (3.11) for all non-normal studies as an optimal value.
- The behavior of Stein-type estimators are restricted by the condition $q \geq 3$. However, The PTE requires the size of testing $H_0 : \mu = 0$. For $W(.)$ as dirac delta function, the maximal savings in risk (SIR) for the shrinkage estimator is $\frac{m(q-2)}{p(m+2)}$, while for $W(.)$ as inverse-gamma function, it is equal to $\frac{m(q-2)}{p(m+2)} \frac{v-2}{\nu}$, where $\nu$ is d.f. of multivariate Student-t distribution. This amount of saving can be evaluated for other elliptical distribution based on the structure of $W(.)$ which is equal to

$$SIR = \int_0^\infty t^{-1}W(t)dt.$$ 

- For any further research, this basic study opens many insights in statistical decision theory under non-normality assumption. Interested readers may extend the presented results for practical goals in ridge regression as well as Bayesian point of view. Furthermore, it is nicely recommend to investigate on the performance of the Stein-type shrinkage estimator given by

$$\hat{\beta}^S = \tilde{\beta} - d^*h(\mathbb{L}_n)(\tilde{\beta} - \hat{\beta}),$$

based on the Borel measurable function $h(.)$. In this regard, the minimaxity and admissibility conditions are welcome.

**Acknowledgements**

The author would like to thank the anonymous referee for his/her comments and Prof. Brualdi for handling and processing the paper.
A. Appendix

We shall provide systematic proofs for the proposed main theorems. Those are essentially important to remain valid under the properties of elliptical models.

The proof of Lemma 2.1 is a direct consequence and extension of Corollary 1 (b) and Theorem 1 of Jozani et al. [12] under multivariate case.

Proof of Theorem 3.1. For the LRT we have

\[
\mathcal{L}_R = \frac{\max_{\omega \Omega} f(y)}{\max_{\omega} f(y)} = \frac{d_n |\tilde{\sigma}^2 V|^{-\frac{1}{2}} \max_{\omega \Omega} g \left[ \frac{yV^{-1}y}{2\sigma^2} \right]}{d_n |\tilde{\sigma}^2 V|^{-\frac{1}{2}} \max_{\omega} g \left[ \frac{yV^{-1}y}{2\sigma^2} \right]}
\]

\[
= \left( \frac{\tilde{\sigma}}{\sigma} \right)^n \frac{g(y_k)}{g(y_{\hat{k}})} = \left( \frac{m \tilde{e} V^{-1} \tilde{e}}{(n-1) \left[ \tilde{e} V^{-1} \tilde{e} + \tilde{\beta}' H' V_1 H \tilde{\beta} \right]} \right)^n
\]

which is decreasing with respect to \( \mathcal{L}_n \). Hence, \( \mathcal{L}_n \) is the LRT for testing the null hypothesis. For its non-null distribution consider that

\[
(n-p)S^2 |t = (y - \hat{X}\tilde{\beta})' V^{-1} (y - \hat{X}\tilde{\beta}) | t
\]

\[
= y' \left[ (V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}) \right] y | t \sim \chi^2_{n-p}
\]

Also \( (HC^{-1}H')^{-1/2} H \tilde{\beta} | t \sim \mathcal{N}_q((HC^{-1}H')^{-1/2} H \beta, t^{-1} \sigma^2 I_q) \). Then

\[
\tilde{\beta}' H' (HC^{-1}H')^{-1} H \tilde{\beta} | t \sim \chi^2_{q, \Delta^2_t}
\]

where \( \Delta^2_t = \frac{it}{\sigma^2} \). Also using the fact that \( (y - \hat{X}\tilde{\beta})' V^{-1} (y - \hat{X}\tilde{\beta}) | t \) and \( \tilde{\beta}' H' (HC^{-1}H')^{-1} H \tilde{\beta} | t \) are independent, we get

\[
\mathcal{L}_n | t = \frac{\tilde{\beta}' H' (HC^{-1}H')^{-1} H \tilde{\beta} | t}{qS^2 | t} \sim F_{q, n-p, \Delta^2_t}
\]

Hence

\[
\mathcal{G}^*_q, m(\mathcal{L}_n) = \int_0^\infty W(t) F_{q, m, \Delta^2_t}(\mathcal{L}_n | t) dt
\]

\[
= \sum_{r=0}^\infty \left( \frac{q}{m} \right)^{q+r} L_n^{q+r-1} K_r^{(0)}(\Delta^2_t) \frac{L_n^{q+r-1} K_r^{(0)}(\Delta^2_t)}{B(\frac{q}{2} + r, \frac{m}{2}) (1 + \frac{q}{m} L_n)^{\frac{q+m+r}{2}}}
\]

Proof of Theorem 5.1. By making use of \( \tilde{z} = H' V_1 H \tilde{\beta} \), the SE can be rewritten as

\[
\beta^* = \tilde{\beta} - qa^* S^2 \frac{C^{-1} H' V_1 H \tilde{\beta}}{\tilde{\beta}' H' V_1 H \tilde{\beta}}
\]

\[
= \tilde{\beta} - qa^* S^2 \left( \tilde{z}' C^{-1} \tilde{z} \right)^{-1} C^{-1} \tilde{z}.
\]

Then, the risk difference of the SE and the UE under quadratic loss function, is given by
\[
D_4 = E \left( \hat{\beta}^2 - \beta \right) C \left( \hat{\beta}^2 - \beta \right) - E \left( \hat{\beta} - \beta \right) C \left( \hat{\beta} - \beta \right)
\]

\[
= (d^*)^2 E \left[ q^2 S^4 \left( \hat{z}'C^{-1}\hat{z} \right)^{-1} \right] - 2d^* E \left[ q^2 S^2 \left( \hat{z}'C^{-1}\hat{z} \right)^{-1} \left( \hat{\beta} - \beta \right)' \hat{z} \right]
\]

\[
= (d^*)^2 E_{\tau} \left[ E_N \left[ q^2 S^4 \left( \hat{z}'C^{-1}\hat{z} \right)^{-1} \mid \tau \right] \right] - 2d^* E_{\tau} \left[ E_N \left[ q^2 S^2 \left( \hat{z}'C^{-1}\hat{z} \right)^{-1} \left( \hat{\beta} - \beta \right)'H'V_1 \left( H\hat{\beta} - h \right) \mid \tau \right] \right]
\]

\[
= \frac{q^2(m+2)}{m} (d^*)^2 E_{\tau} \left( \frac{\tau^{-2}}{\hat{z}'C^{-1}\hat{z}} \right) - 2q^2 d^* E_{\tau} \left( \frac{\tau^{-2}}{\hat{z}'C^{-1}\hat{z}} \right),
\]

since \( \frac{mS^2}{\sigma^2} \) \( \tau \sim \tau^{-1} \chi_m^2 \) and \( \hat{\beta}'H'V_1H\hat{\beta} \mid \tau \sim \tau^{-2} \sigma^2 \chi_q^2(d) \), where \( d = \beta'H'V_1H\beta \).

Therefore, \( D_4 \leq 0 \) if and only if \( 0 < d^* \leq \frac{2m}{m+2} \) since \( \int_{0}^{\infty} \frac{\tau^{-2}}{\chi_m^2} \) \( dW(\tau) > 0 \). (See [24]).

References