JIRSS (2018)
Vol. 17, No. 1, pp 49-61
DOI: 10.29252/jirss.17.1.49

Shrinkage Estimation in Restricted Elliptical Regression Models

Reza Fallah [1], **Mohammad Arashi** [2], **Seyed Mohammad M. Tabatabaey** [3]

[1] Department of Statistics, Ferdowsi University of Mashhad, International Campus, Iran.
[2] Department of Statistics, Shahrood University of Technology, Shahrood, Iran.
[3] Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran.

Received: 21/10/2016, Revision received: 03/07/2017, Published online: 19/05/2018

Abstract. In the restricted elliptical linear model, an approximation for the risk of a general shrinkage estimator of the regression vector-parameter is given. Superiority condition of the shrinkage estimator over the restricted estimator is investigated under the elliptical assumption. It is evident from numerical results that the shrinkage estimator performs better than the unrestricted one in the multivariate t-regression model.

Keywords. Approximate risk; Elliptically contoured distribution; Linear regression model; Restricted estimator; Shrinkage estimator.

MSC: Primary: 62F10, 62F03; Secondary: 62H05, 62H12.

1 Introduction

Conventional estimators, such as the maximum likelihood and least squares estimators, are only based on the sample responses. An improved estimator, on the other hand,

Reza Falah (rezafallah62@gmail.com)
Corresponding Author: Mohammad Arashi (m_arashi_stat@yahoo.com)
Seyed Mohammad M. Tabatabaey (tabatabaey@yahoo.com)

incorporates both the sample and non-sample (prior) information in the definition of the estimator. Among the popular improved estimators, shrinkage estimators perform better than the conventional ones under certain conditions. If improvement in the risk or prediction error is more desirable than unbiasedness, shrinkage estimators are preferred.

On the other hand, if the validity of the prior assumption is not tested, neither the restricted nor unrestricted (conventional) estimator makes use of the available information in an optimal way. The Stein-type shrinkage estimator incorporates this uncertain prior information, and combines the restricted and unrestricted estimators in a superior manner.

Employing shrinkage estimation in the analysis started after the seminal work of Stein (1956). Since then, considerable researches focusing on shrinkage estimation of location parameters have been conducted. For instance, we refer to Sen and Saleh (1985), Saleh and Sen (1987), Saleh and Kibria (1993), Kibria and Saleh (2003), Nkurunziza (2011, 2012), and Chen and Nkurunziza (2016).

Stein (1981) derived an expression for the risk of general shrinkage estimators in normal models. Ullah et al. (1983) discussed the properties of shrinkage estimators in regression models when disturbances are not normal. Green and Strawderman (1991) showed how to combine biased and unbiased estimators using the Stein shrinkage concept. Verma and Singh (2002) studied a general class of shrinkage estimators in multivariate t-regression models. Saleh (2006) provided a comprehensive discussion of Stein-type estimators under different statistical models. Arashi and Tabatabaey (2008) evaluated the performance of Stein-type shrinkage estimators in multivariate t-regression models under stochastic restrictions. Tsukuma (2009) studied shrinkage estimation in restricted elliptical models. Saleh and Kibria (2010) developed shrinkage estimation in elliptical regression models, and applied shrinkage estimation in ridge and logistic regression models in 2011 and 2013, respectively. The contributions of Nkurunziza (2011, 2012) and Nkurunziza and Ahmed (2010, 2011) to the field of matrix shrinkage estimation should be also acknowledged.

For practical sake, it is of most importance to develop shrinkage estimators in regression models with restricted parameter spaces. The rest of this paper is organized as follows. Section 2 is devoted to the formulation of the general form of shrinkage estimator in the elliptical regression model with restricted parameter space. In Section 3 we find the risk function of the shrinkage estimator under some level of approximation, while some numerical studies through a Monte Carlo simulation and a real data example are provided in Section 4. Proof of the main result is given in the Appendix.

2 Model and Estimators

Consider the linear model

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}, \tag{2.1}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of responses, X is an $n \times p$ matrix of n observations on p explanatory variables (of full column rank p), $\boldsymbol{\beta}$ is a $p\times1$ vector of regression coefficients and $\boldsymbol{u}$ is the $n \times 1$ vector of random errors distributed according to the law belonging to the class of elliptically contoured distributions (ECDs) by the characteristic function $\varphi_{\boldsymbol{u}}(\boldsymbol{t}) = \psi(\sigma^2 \boldsymbol{t}^T V \boldsymbol{t})$, where $V \in S(n)$ is known, $S(n)$ denotes the set of all positive definite matrices of order$(n \times n)$, $\sigma \in \mathbb{R}^+$ is unknown and $\psi : [0, \infty] \longrightarrow R$ is the characteristic generator.

This paper deals with a class of ECDs preserving a specific property for the density function. We further suppose that $\boldsymbol{u}$ has a density function of the form

$$f(\boldsymbol{u}) = |\sigma^2 V|^{-1/2} h\left(\frac{\boldsymbol{u}^T V^{-1} \boldsymbol{u}}{2\sigma^2}\right) = \int_{R^+} W(t) g(\boldsymbol{u}|t) dt, \tag{2.2}$$

where $h(.)$ is a non-negative function over $\mathbb{R}^+$ such that $f(.)$ is a density function with respect to (w.r.t.) a σ-finite measure μ on $\mathbb{R}^p$, $g(.|t)$ is the density function of $N_n\left(0, t^{-1}\sigma^2 V\right)$ and $W(.)$ is a weighting function. We refer to Provost and Cheong (2002), Nkurunziza (2013) and Nkurunziza and Chen (2013) for more details and applications.

Under our assumptions, the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$ is of the form $\hat{\boldsymbol{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \boldsymbol{y} = C^{-1} X^T V^{-1} \boldsymbol{y}$, $C = X^T V^{-1} X$, with the variance-covariance matrix $\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma_\psi^2 C^{-1}$, $\sigma_\psi^2 = -2\psi'(0)\sigma^2$, where $\psi'(0)$ is the first derivative of ψ at zero. An unbiased estimator of σ_ψ^2 is given by $S^2 = \hat{\boldsymbol{u}}^T V^{-1} \hat{\boldsymbol{u}}/(n-p)$, $\hat{\boldsymbol{u}} = \boldsymbol{y} - X\hat{\boldsymbol{\beta}}$.

Let the exact linear restriction on $\boldsymbol{\beta}$ in model (2.1) is of the form $R\boldsymbol{\beta} = \boldsymbol{r}$, where R is a known $J \times p$ matrix of full row rank ($J \leq p$) and $\boldsymbol{r}$ is a known $J \times 1$ vector. Under the linear restriction $R\boldsymbol{\beta} = \boldsymbol{r}$, the restricted GLSE is given by

$$\hat{\boldsymbol{\beta}}^R = \hat{\boldsymbol{\beta}} + C^{-1} R^T V_1 (r - R\hat{\boldsymbol{\beta}}), \qquad V_1 = (RC^{-1}R^T)^{-1}, \tag{2.3}$$

with the variance-covariance matrix $\text{cov}(\hat{\boldsymbol{\beta}}^R) = \sigma^2 \Omega C^{-1}$, $\Omega = \left[I - C^{-1} R^T V_1 R\right]$.

Saleh and Kibria (2010) considered restricted estimators of the form given in (2.3) and respective shrinkage estimators in their study. Let $s = \hat{\boldsymbol{u}}^T V^{-1} \hat{\boldsymbol{u}} / \hat{\boldsymbol{v}}^T \hat{\boldsymbol{v}}$, $\hat{\boldsymbol{v}} = C^{\frac{1}{2}} \hat{\boldsymbol{\beta}}$. Srivastava and Chandra (1991) suggested to replace $\hat{\boldsymbol{\beta}}$ in the restricted estimator $\hat{\boldsymbol{\beta}}^R$ by $(I - ksD)\hat{\boldsymbol{\beta}}$, where $k \in \mathbb{R}$ is a fixed scalar and $D \in S(p)$.

Replacing $\hat{\boldsymbol{\beta}}$ by the term $(I - ksD)\hat{\boldsymbol{\beta}}$, gives a new estimator given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{SC} &= (I - ksD)\hat{\boldsymbol{\beta}} + C^{-1}R^T V_1(r - R(I - ksD)\hat{\boldsymbol{\beta}}) \\ &= \hat{\boldsymbol{\beta}}^R - ks\Omega D\hat{\boldsymbol{\beta}}.\end{aligned}$$

Following Verma and Singh (2002), we consider a more general class of shrinkage estimators in the elliptical regression model of the form

$$\hat{\boldsymbol{\beta}}^{GR} = \hat{\boldsymbol{\beta}}^R - g(s)\Omega D\hat{\boldsymbol{\beta}}, \tag{2.4}$$

where the function $g(s)$, having validity conditions of Taylor series expansion, satisfies some mild regularity conditions.

Indeed, the coefficient ks in Srivastava and Chandra (1991) is replaced by the general function $g(s)$.

One must also note that $\hat{\boldsymbol{\beta}}^{GR}$ is not a restricted estimator since $R\hat{\boldsymbol{\beta}}^{GR} = R\hat{\boldsymbol{\beta}}^R - g(s)R\Omega D\hat{\boldsymbol{\beta}} = \boldsymbol{r} - g(s)R\Omega D\hat{\boldsymbol{\beta}} \neq \boldsymbol{r}$. In fact, the general function $g(s)$ will result in the estimator $\hat{\boldsymbol{\beta}}^{GR}$ to not satisfy the restriction, even if $D = \Omega^{-1}$.

3 Approximate Risk Function

In this section, we provide the approximate risk function of the shrinkage estimator $\hat{\boldsymbol{\beta}}^{GR}$ of $\boldsymbol{\beta}$ in the elliptical regression model. For any estimator $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$, the risk function is evaluated according to $\text{Risk}(\boldsymbol{\beta}^*) = \mathbb{E}(\boldsymbol{\beta}^* - \boldsymbol{\beta})^T Q(\boldsymbol{\beta}^* - \boldsymbol{\beta})$, where Q is a positive definite weight matrix, respectively. For our purpose, we need the following regularity conditions to hold.

(A1) Let first three derivatives of $g(s)$ w.r.t. s be bounded such that $g(s = 0) = 0$.

(A2) $g(s) = O(\theta^{-1})$, as $\theta \to \infty$, with $\theta = \boldsymbol{v}^T\boldsymbol{v}$, $\boldsymbol{v} = C^{\frac{1}{2}}\boldsymbol{\beta}$.

(A3) Let $\kappa^{(h)} = E(t^{-h}) = \int \left(\frac{1}{t}\right)^h W_0(t)dt$. For $i = 1, 2, 3$, $\kappa^{(i)}$ exists.

Theorem 3.1. *Assume (A1)-(A3). Then, ignoring the terms of order $O(\theta^{-3})$, the risk function of $\hat{\boldsymbol{\beta}}^{GR}$ is given by*

$$\text{Risk}(\hat{\boldsymbol{\beta}}^{GR}) = \kappa^{(1)} trQ^* - 2(n-p)g'(0)\sigma^2 \left[\kappa^{(2)}\left(\frac{trQ^{**}}{\theta} - 2\frac{\boldsymbol{\beta}^T {C^{\frac{1}{2}}}^T Q^{**} C^{\frac{1}{2}}\boldsymbol{\beta}}{\theta^2}\right.\right.$$

$$
\begin{aligned}
&-\frac{n-p+2}{2\theta^2}g'(0)\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{***} C^{\frac{1}{2}}\boldsymbol{\beta}\Bigg) - \kappa^{(3)}\Bigg(\frac{(p+6)\sigma^2 trQ^{**}}{\theta^2}\\
&-\frac{(n-p+2)\sigma^2}{2\theta^2}\frac{g''(0)}{g'(0)}trQ^{**} - \frac{(n-p+2)\sigma^4}{2\theta^2}g'(0)trQ^{***}\\
&-\frac{16\sigma^2}{\theta^3}\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{**} C^{\frac{1}{2}}\boldsymbol{\beta} + \frac{4\sigma^2}{\theta^3}(p+2)\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{**} C^{\frac{1}{2}}\boldsymbol{\beta}\\
&-\frac{2\sigma^2}{\theta^3}\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{**} C^{\frac{1}{2}}\boldsymbol{\beta}\frac{g''(0)}{g'(0)}\\
&+\frac{2\sigma^4}{\theta^3}g'(0)(n-p+2)\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{***} C^{\frac{1}{2}}\boldsymbol{\beta}\Bigg)\Bigg],
\end{aligned}
$$

where $Q^* = C^{-\frac{1}{2}}\Omega Q\Omega C^{-\frac{1}{2}}$, $Q^{**} = C^{-\frac{1}{2}}\Omega Q\Omega D C^{-\frac{1}{2}}$, *and* $Q^{***} = C^{-\frac{1}{2}}D^T\Omega Q\Omega D C^{-\frac{1}{2}}$.

For the proof, see Appendix.

3.1 Superiority Conditions

In this section, we derive the necessary conditions under which the shrinkage estimator $\hat{\boldsymbol{\beta}}^{GR}$ is superior to $\hat{\boldsymbol{\beta}}^{R}$ in risk sense, in the elliptical regression model. The following result is a direct conclusion from Theorem 3.1.

Proposition 3.1. *Under the assumption of Theorem 3.1 and ignoring the terms of order* $O(\frac{tr(Q^{**})}{\theta^2})$, $O(\frac{tr(Q^{***})}{\theta^2})$, $O(\frac{\beta^T C^{\frac{1}{2}^T} Q^{**} C^{\frac{1}{2}}\beta}{\theta^3})$, $O(\frac{\beta^T C^{\frac{1}{2}^T} Q^{***} C^{\frac{1}{2}}\beta}{\theta^3})$, *we have*

$$
\begin{aligned}
\text{Risk}(\hat{\boldsymbol{\beta}}^{GR}) &= \text{Risk}(\hat{\boldsymbol{\beta}}^{R}) + 2(n-p)g'(0)\kappa^{(2)}\sigma^2\Bigg(-\frac{trQ^{**}}{\theta} + 2\frac{\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{**} C^{\frac{1}{2}}\boldsymbol{\beta}}{\theta^2}\\
&\quad +\frac{n-p+2}{2\theta^2}g'(0)\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{***} C^{\frac{1}{2}}\boldsymbol{\beta}\Bigg)
\end{aligned}
$$

The following result, provides the necessary conditions for superiority of the shrinkage estimator over the restricted GLSE, in the elliptical regression model.

Proposition 3.2. *Under the assumptions of Proposition 3.1, the restricted GLSE is inadmissible if the function* $g'(0)$ *satisfies the inequality*

$$
0 < g'(0) < \frac{2}{n-p+2}\left(d - \frac{2\boldsymbol{\beta}^T\Omega Q\Omega D\boldsymbol{\beta}}{\boldsymbol{\beta}^T D^T\Omega Q\Omega D\boldsymbol{\beta}}\right),\ d = \frac{\theta trQ^{**}}{\boldsymbol{\beta}^T C^{\frac{1}{2}^T} Q^{***} C^{\frac{1}{2}}\boldsymbol{\beta}}.
$$

4 Illustrations

In this section, we provide some numerical results to support our findings. For different choices of D and $g(s)$ satisfying the regularity conditions of $\hat{\boldsymbol{\beta}}^{GR}$, several particular estimators may be obtained. Here, we study the numerical properties of $\hat{\boldsymbol{\beta}}^{GR}$ in the elliptical regression model, in the special case where $g(s)$ is equal to $ks/(s+1)$, for some positive values k. Hence, the specific form of $\hat{\boldsymbol{\beta}}^{GR}$ is given by

$$\hat{\boldsymbol{\beta}}^{GR} = \hat{\boldsymbol{\beta}}^{R} - \frac{ks}{s+1}\Omega D\hat{\boldsymbol{\beta}}. \tag{4.1}$$

One must note that any candidate for the $g(s)$ function in (2.4) should satisfy the proposed regularity conditions (A1)-(A2). Boundedness of polynomial and fractional functions can be tested relatively easily as well as the condition $g(0) = 0$. For checking condition (A2), consider that $g(s) = g(\hat{\boldsymbol{u}}^T\hat{\boldsymbol{u}}/\theta)$. Hence, for any function $g(x) = p(x)/q(x)$ for which the degree of $q(\cdot)$ is larger than or equal to the degree of $p(\cdot)$, the condition satisfies. Again, for fractional functions, one may simply check whether $\theta g(\hat{\boldsymbol{u}}^T\hat{\boldsymbol{u}}/\theta)$ is bounded. For these reasons, we started by taking $g(s) = ks/(s+1)$ to study a simple example satisfying our conditions.

In this section, we consider the multivariate t-distribution as the distribution of the error term in (2.1).

4.1 Simulation

In this section, we carry out a Monte Carlo simulation experiment to investigate the (quadratic) risk performance of the proposed estimators. For this purpose, this simulation is based on a multivariate t-regression model with $\gamma = 9$ degrees of freedom. In our sampling experiment, the sample size is n with p as the number of predictors. We set $\boldsymbol{\beta} = (0, 0, 0, \mathbf{1}^T_{(p-3)\times 1})^T$ in our scheme. For simulation purposes, we consider the particular restriction $R\boldsymbol{\beta} = \mathbf{0}$, with $R = (I_3 \vdots 0_{3\times(p-3)})$. Hence, we generate responses using $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$, where $\boldsymbol{u}$ is generated according to the t-distribution $\mathcal{T}_n(\mathbf{0}, I_n, \gamma)$, and the ith covariate generated from $\mathcal{N}_n(\mathbf{0}, I_n)$. The whole process is repeated $N = 1000$ times to evaluate the risk of the estimators. The performance of an estimator $\boldsymbol{\beta}$, say $\boldsymbol{\beta}^*$, will be measured in terms of the empirical risk evaluated by $\text{Risk}(\hat{\boldsymbol{\beta}}^*) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}^*_i - \boldsymbol{\beta})^T(\hat{\boldsymbol{\beta}}^*_i - \boldsymbol{\beta})$, where $\hat{\boldsymbol{\beta}}^*_i$ is the estimator value in ith replication. Table 1 summarizes the relative risk

Table 1: Simulated (empirical) relative risk values of restricted estimators with respect to $\hat{\boldsymbol{\beta}}$ for different parameter values n and p.

	n			
p	10	30	50	100
3	1.4898	1.0029	1.0013	1.0001
5	1.0046	1.0032	1.0006	1.0001
8	1.45	1.0018	1.0005	1.00004

Table 2: Simulated (empirical) relative risk values of shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ for different parameters values n , p and k.

		k						
n	p	0.01	0.05	0.1	0.15	0.2	1	2
10	3	1.1166	1.2954	1.5682	3.0521	3.9878	5.346	0.7372
	5	1.1380	1.3080	1.5755	1.9253	2.3894	3.0701	0.1559
	8	3.9852	4.0633	4.0187	4.0120	3.997	0.0325	0.011
30	3	1.0128	1.0128	1.0125	1.9019	2.2920	4.2558	0.5671
	5	1.0513	1.1504	1.472	1.8252	1.7991	1.5601	0.9147
	8	1.4568	1.2864	1.6273	2.1341	3.0599	0.3037	0.4077
50	3	1.0226	1.1145	1.0013	1.4	1.5803	5.5785	0.7677
	5	1.0374	1.2050	1.472	1.8252	2.2964	0.8185	0.1342
	8	1.0467	1.2634	1.6273	2.1341	2.8279	0.36621	0.0483
100	3	1.0249	1.3316	1.2929	1.4869	1.7247	4.8288	0.4478
	5	1.0721	1.195	1.4489	1.7861	2.2424	1.1763	0.1431
	8	1.1018	1.2840	1.6948	2.3055	3.2270	0.3687	0.04969

values evaluated according to $\frac{\text{Risk}(\hat{\boldsymbol{\beta}})}{\text{Risk}(\hat{\boldsymbol{\beta}}^R)}$ and Table 2 summarizes the relative risk values evaluated according to $\frac{\text{Risk}(\hat{\boldsymbol{\beta}})}{\text{Risk}(\hat{\boldsymbol{\beta}}^{GR})}$. From Tables 1 and 2, the larger k is, the worse is the performance of the shrinkage estimator. It is also realized that for $k \leq 1$, the shrinkage estimator still performs better than the restricted estimator. Changing the values p and n does not lead in substantial changes or any specific trend. However, as one may expect, when n gets larger, the performance of $\hat{\boldsymbol{\beta}}$ becomes better.

Table 3: Simulated (empirical) relative MSPE values of estimators with respect to $\hat{\boldsymbol{y}}$.

	k								
	0.01	0.05	0.1	0.5	1	5	10	20	100
$\frac{\text{MSPE}(\hat{\boldsymbol{\beta}})}{\text{MSPE}(\hat{\boldsymbol{\beta}}^{GR})}$	1.0030	1.0039	1.0049	1.0127	1.0224	1.0977	1.1804	1.2716	0.285

4.2 Real Example

The data we use here is from a clinical trial on 34 male patients with 3 covariates, body weight (WT) in kg, serum creatinine (SC) concentration in mg/deciliter, and age in years, and one outcome variable, endogenous creatinine(CR) clearance. Shih and Weisberg (1986) fitted a multivariate t-regression model to this data. Here, we calculate the empirical mean squared prediction error to investigate the (quadratic) risk performance of the estimators. To this end, we consider the particular restriction $\boldsymbol{R}\boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{R} = (1\ -1\ 0)$. With $N = 1000$ replications, the empirical mean square prediction error of the estimators is evaluated using the formula $\text{MSPE}(\hat{\boldsymbol{y}}^*) = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{y_i} - \hat{\boldsymbol{y}}_i^*\right)^2$, where $\hat{\boldsymbol{y}}_i^*$ is the prediction value in the ith replication of $\hat{\boldsymbol{y}}^*$, and the prediction obtains by using $\hat{\boldsymbol{\beta}}^*$ as the estimator of $\boldsymbol{\beta}$. Table 3 summarizes the relative MSPE values according to $\frac{\text{MSPE}(\hat{\boldsymbol{\beta}})}{\text{MSPE}(\hat{\boldsymbol{\beta}}^{GR})}$. We also found that $\frac{\text{MSPE}(\hat{\boldsymbol{\beta}})}{\text{MSPE}(\hat{\boldsymbol{\beta}}^{R})} = 1.003$. From Table 3, For larger values of k, the shrinkage estimator is still superior. However, as the k factor goes to infinity, the shrinkage estimator is dominated by the GLS estimator. This scenario is the same for the comparison between the shrinkage and restricted GLS.

5 Conclusions

In this paper, following the general class of estimators proposed by Srivastava and Chandra (1991), we defined a class of shrinkage estimators for the regression vector-parameter in the elliptical regression model. The approximate risk function was obtained using the Taylor series expansion of some order. Further, superiority condition of any estimator in this class was investigated and compared to the restricted estimator in the elliptical regression model. Some numerical studies were conducted to show the superiority of a specific member in the class over the restricted estimator in the risk sense. In this respect, we illustrated the findings by a Monte Carlo simulation and a real data example in the context of clinical trials.

There are some plausible extensions that can be considered for future researches.

The result of this paper can be extended for stochastic restrictions in which the form of elliptical distribution is of importance. The proposed framework is general and can be used in the context of ridge/Liu estimator for multicollinear situations. To this end, the GLS estimator must be replaced by the ridge/Liu estimator. On the other hand, there are some limitations. One of them is the specific choice taken for $g(s)$ in the numerical studies. Hence, the superior performance did not check numerically for all members of the class.

Acknowledgements

The authors would like to thank two anonymous reviewers for constructive comments which led to adding more details in the paper and a great deal of improvement in the presentation.

References

Arashi, M. and Tabatabaey, S. M. M. (2008), *Stein-type improvement under stochastic constraints: Use of multivariate Student-t model in regression*. Statist. Prob. Lett., **78**, 2142-2153.

Chen, F. and Nkurunziza, S. (2016), *A class of Stein-rules in multivariate regression model with structural changes*. Scandinavian J. Statist., **43**(1), 83-102.

Green, E. J. and Strawderman, W. E. (1991), *A James-Stein type estimator for combining unbiased and possibly biased estimators*. J. Amer. Statist. Assoc., **86**(416), 1001-1006.

Kibria, B. M. G. and Saleh, A. K. Md. Ehsanes (2003), *Preliminary test ridge regression estimators with students t errors and conflicting test-statistics*. Metrika, **59**(2), 105124.

Nkurunziza, S. (2013), *The bias and risk functions of some Stein-rules in elliptically contoured distributions*. Math. Meth. Statist., **22**(1), 70-82.

Nkurunziza, S. (2012), *Shrinkage strategies in some multiple multi-factor dynamical systems*. ESAIM: Prob. Statist., **16**, 139-150.

Nkurunziza, S. (2011), *Shrinkage strategy in stratified random sample subject to measurement error*. Statist. Prob. Lett., **81**(2), 317-325.

Nkurunziza, S. and Ahmed, S. E. (2011), *Estimation strategies for the regression coefficient parameter matrix in multivariate multiple regression*. Statist. Neerlandica, **65**(4), 387-406.

Nkurunziza, S. and Ahmed, S. E. (2010), *Shrinkage drift parameter estimation for multi-factor Ornstein-Uhlenbeck processes*. Appl. Stoch. Mod. Bus. Ind., **26**(2), 103-124.

Nkurunziza, S., and Chen, F. (2013), *On extension of some identities for the bias and risk functions in elliptically contoured distributions*. J. Mult. Anal., **122**, 190-201.

Provost, S. B. and Cheong, Y-H. (2002), *The distribution of Hermitian quadratic forms in elliptically contoured random vectors*. J. Statist Plann. Inf., **102**, 303-316.

Saleh, A. K. Md. Ehsanes and Kibria, B. M. G. (2013), *Improved ridge regression estimators for the logistic regression model*. Comp. Statist., **28**(6), 2519-2558.

Saleh, A. K. Md. Ehsanes and Kibria, B. M. G. (2011), *On some ridge regression estimators: a nonparametric approach*. J. Nonparametric Statist., **23**(3), 819-851.

Saleh, A. K. Md. Ehsanes and Kibria, B. M. G. (2010), *Estimation of the mean vector of a multivariate elliptically contoured distribution*. Calcutta Statist. Assoc. Bull., **62**, 247-268.

Saleh, A. K. Md. Ehsanes and Kibria, B. M. G. (1993), *Performance of some new preliminary test ridge regression estiamtors and their properties*. Comm. Statist. Theo. Meth., **22**, 2747-2764.

Saleh, A. K. Md. Ehsnaes and Sen, P. K. (1987), *Relative performance of Stein-rule and preliminary test estimators in linear models: Least square theory*. Comm Statist. Theo. Meth., **16**(2), 461-476.

Sen, P. K. and Saleh, A. K. Md. Ehsanes (1985), *On some shrinkage estimators of multivariate location*. Ann. Statist., **13**, 272-281.

Singh, R. K., Pandey, S. K. and Srivastava, V. K. (1994), *A generalized class of shrinkage estimators in linear regression when disturbances are not normal*. Comm. Statist. Theo. Meth. **23**(7), 2029-2064.

Shih, W. J. and Weisberg, S. (1986), *Assessing influence in multiple linear regression with incomplete data*. Technometrics, **28**, 231-239.

Srivastava, A. K. and Chandra, R. (1991), *Improved estimation of restricted regression model when disturbances are not necessarily normal*. Sankhya, Ser. B, **53**, 119-133.

Stein, C. (1956), *Inadmissibility of the usual estimator for the mean of a multivariate distribution*. Proc. 3rd Berkeley Symp. Math. Statist. Prob, **1**, 197-206.

Stein, C. (1981), *Estimation of the mean of a multivariate normal distribution*. Ann. Statist, **9**(6), 1135-1151.

Tsukuma, H. (2009), *Shrinkage estimation in elliptically contoured distribution with restricted parameter space*. Statist. Dec, **27**, 25-35.

Ullah, A., Srivastva, V. K. and Chandra, R. (1983), *Propertes of shrinkage estimators in linear regression when disturbances are not normal*. J. Econometrics, **21**, 389-402.

Verma S. and Singh R. K. (2002), *Estiamtion in restricted regression model with multivariate t-distributed error*. Metron, **1**, 67-82.

Appendix

In this section, we provide the sketch of proof of Theorem 3.1.

Let, for any scalar $t \in \mathbb{R}^+$, $\boldsymbol{z} = \sigma^{-1} t^{\frac{1}{2}} C^{\frac{1}{2}} \hat{\boldsymbol{\beta}}$, $\boldsymbol{\delta} = \sigma^{-1} t^{\frac{1}{2}} C^{\frac{1}{2}} \boldsymbol{\beta}$, and $v = \sigma^{-2} t \hat{\boldsymbol{u}}^T V^{-1} \hat{\boldsymbol{u}}$. According to our assumption (see Eq.(2.2)), $\boldsymbol{z}|t \sim N(\boldsymbol{\delta}, I)$ and $v|t \sim \chi^2_{(n-p)}$ are independent. Expanding $g(s)$ in third order Taylor series about point $s = 0$, we have

$$\hat{\boldsymbol{\beta}}^{GR} - \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{R} - \boldsymbol{\beta} - \left\{ g(0) + s g'(0) + \frac{s^2}{2!} g''(0) + \frac{s^3}{3!} g'''(s_0) \right\} \Omega D \hat{\boldsymbol{\beta}}, \tag{5.1}$$

where $s_0 = \omega s$, $0 < \omega < 1$.

Since for $\boldsymbol{\varepsilon} = \boldsymbol{z} - \boldsymbol{\delta}$, $s = \frac{\hat{\boldsymbol{u}}^T V^{-1} \hat{\boldsymbol{u}}}{\hat{\boldsymbol{v}}^T \hat{\boldsymbol{v}}} = \frac{\sigma^2 t^{-1} v}{\sigma^2 t^{-1} (z^T z)} = v(\boldsymbol{\delta}^T \boldsymbol{\delta})^{-1} \left[1 + \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T \boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T \boldsymbol{\delta}} \right]^{-1}$, utilizing s in (5.1) and using the fact that $g(0) = 0$, we get

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{GR} - \boldsymbol{\beta} &= \hat{\boldsymbol{\beta}}^{R} - \boldsymbol{\beta} - t^{-\frac{1}{2}} \Bigg[v(\boldsymbol{\delta}^T \boldsymbol{\delta})^{-1} \left(1 + \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T \boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T \boldsymbol{\delta}} \right)^{-1} g'(0) \\ &\quad + \frac{1}{2} v^2 (\boldsymbol{\delta}^T \boldsymbol{\delta})^{-2} \left(1 + \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T \boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T \boldsymbol{\delta}} \right)^{-2} g''(0) \\ &\quad + \frac{v^3 (\boldsymbol{\delta}^T \boldsymbol{\delta})^{-3}}{6} \left(1 + \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T \boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T \boldsymbol{\delta}} \right)^{-3} g'''(s_0) \Bigg] \Omega D C^{-\frac{1}{2}} (\boldsymbol{\varepsilon} + \boldsymbol{\delta}). \end{aligned}$$

Further, note that $\hat{\boldsymbol{\beta}}^{GR} - \boldsymbol{\beta} = t^{-\frac{1}{2}}\Omega C^{-\frac{1}{2}}(\boldsymbol{z} - \boldsymbol{\delta}) - g(s)\Omega D t^{-\frac{1}{2}} C^{-\frac{1}{2}}\boldsymbol{z}$. Then, the risk expression is of the form

$$\begin{aligned}\text{Risk}(\hat{\boldsymbol{\beta}}^{GR}) &= \mathbb{E}(\hat{\boldsymbol{\beta}}^{GR} - \boldsymbol{\beta})^T Q(\hat{\boldsymbol{\beta}}^{GR} - \boldsymbol{\beta}) = \mathbb{E}\mathbb{E}_t t^{-1}(\boldsymbol{\varepsilon}^T Q^* \boldsymbol{\varepsilon}) \\ &\quad -2\mathbb{E}t^{-1} g(s)\left\{\boldsymbol{\varepsilon}^T Q^{**}(\boldsymbol{\varepsilon} + \delta)\right\} + \mathbb{E}t^{-1} g(s)^2\left\{(\boldsymbol{\varepsilon} + \boldsymbol{\delta})^T Q^{***}(\boldsymbol{\varepsilon} + \boldsymbol{\delta})\right\},\end{aligned}$$

where $Q^* = C^{-\frac{1}{2}}\Omega Q \Omega C^{-\frac{1}{2}}$, $Q^{**} = C^{-\frac{1}{2}}\Omega Q \Omega D C^{-\frac{1}{2}}$, and $Q^{***} = C^{-\frac{1}{2}} D^T \Omega Q \Omega D C^{-\frac{1}{2}}$. Expanding $g(s)$ in the third order Taylor series about the point $s = 0$ and using the fact that $g(s = 0) = 0$, we obtain $\text{Risk}(\hat{\boldsymbol{\beta}}^{GR}) = \mathbb{E}\mathbb{E}_t t^{-1}(\boldsymbol{\varepsilon}^T Q^* \boldsymbol{\varepsilon}) - \mathbb{E}\mathbb{E}_t\left[t^{-1}(B_t + C_t)\right]$, where

$$\begin{aligned}B_t = -2v(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}&\left\{\left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-1} g'(0) + \frac{1}{2}v(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}\left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-2} g''(0) + \right. \\ &\left. \frac{v^2}{6}(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-2}\left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-3} g'''(s_0)\right\}\boldsymbol{\varepsilon}^T Q^{**}(\boldsymbol{\varepsilon} + \boldsymbol{\delta})\end{aligned}$$

and

$$\begin{aligned}C_t = &\left\{v(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}\left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-1} g'(0) + \frac{1}{2}v^2(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-2}\left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-2} g''(0) + \right. \\ &\left. \frac{v^3}{6}(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-3}\left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-3} g'''(s_0)\right\}^2 (\boldsymbol{\varepsilon} + \boldsymbol{\delta})^T Q^{***}(\boldsymbol{\varepsilon} + \boldsymbol{\delta}).\end{aligned}$$

Consider the approximations

$$\begin{aligned}\left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-1} &\simeq 1 - (\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}) \\ &\quad +4(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-2}(\boldsymbol{\delta}^T\boldsymbol{\varepsilon})(\boldsymbol{\delta}^T\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}) - 8(\boldsymbol{\delta}^T\boldsymbol{\varepsilon})^3(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-3} \\ \left(1 + \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{\boldsymbol{\delta}^T\boldsymbol{\delta}}\right)^{-2} &\simeq 1 - 4(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}\boldsymbol{\delta}^T\boldsymbol{\varepsilon},\end{aligned}$$

then, B_t and C_t can be approximated by the expressions

$$\begin{aligned}B_t(\theta) &= -2v(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}\left\{\left[1 - (\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon}) + 4(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-2}(\boldsymbol{\delta}^T\boldsymbol{\varepsilon})(\boldsymbol{\delta}^T\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon})\right.\right. \\ &\quad \left.\left. -8(\boldsymbol{\delta}^T\boldsymbol{\varepsilon})^3(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-3}\right] g'(0) + \frac{1}{2}v(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}\left(1 - 4(\boldsymbol{\delta}^T\boldsymbol{\delta})^{-1}\boldsymbol{\delta}^T\boldsymbol{\varepsilon}\right) g''(0)\right\}\boldsymbol{\varepsilon}^T Q^{**}(\boldsymbol{\varepsilon} + \boldsymbol{\delta})\end{aligned}$$

$$= \left\{-2\frac{v}{t\theta}g'(0) + 2\frac{v(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} + 2\boldsymbol{\delta}^T\boldsymbol{\varepsilon})}{t^2\theta^2}g'(0) - 8v\frac{(\boldsymbol{\delta}^T\boldsymbol{\varepsilon})(\boldsymbol{\delta}^T\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon})}{t^3\theta^3}g'(0)\right.$$
$$\left. +16v\frac{(\boldsymbol{\delta}^T\boldsymbol{\varepsilon})^3}{t^4\theta^4}g'(0) - \frac{v^2}{t^2\theta^2}g''(0) + 4v^2\frac{\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}{t^3\theta^3}g''(0)\right\}\boldsymbol{\varepsilon}^T Q^{**}(\boldsymbol{\varepsilon} + \boldsymbol{\delta}),$$

and $C_t(\theta) = \frac{1}{t^2\theta^2}(g'(0))^2\left(v^2 - 4v^2(t\theta)^{-1}\boldsymbol{\delta}^T\boldsymbol{\varepsilon}\right)(\boldsymbol{\varepsilon} + \boldsymbol{\delta})^T Q^{***}(\boldsymbol{\varepsilon} + \boldsymbol{\delta})$. Similar to Singh et al. (1994), it can be shown that, as $\theta \to \infty$, $\theta^{\frac{5}{2}}| - 2t^{-1}g(s)\boldsymbol{\varepsilon}^T Q^{**}(\boldsymbol{\varepsilon} + \boldsymbol{\delta}) - B_t(\theta)| \xrightarrow{p} 0$, and $\theta^{\frac{5}{2}}|t^{-1}g(s)^2(\boldsymbol{\varepsilon} + \boldsymbol{\delta})^T Q^{***}(\boldsymbol{\varepsilon} + \boldsymbol{\delta}) - C_t(\theta)| \xrightarrow{p} 0$. Also $|\mathbb{E}_t\left\{-2t^{-1}g(s)\boldsymbol{\varepsilon}^T Q^{**}(\boldsymbol{\varepsilon} + \boldsymbol{\delta})\right\} - \mathbb{E}_t(B_t(\theta))| = O\left(\theta^{-3}\right)$ and $|\mathbb{E}_t\left\{t^{-1}g(s)^2(\boldsymbol{\varepsilon} + \boldsymbol{\delta})^T Q^{***}(\boldsymbol{\varepsilon} + \boldsymbol{\delta})\right\} - \mathbb{E}_t(C_t(\theta))| = O\left(\theta^{-3}\right)$. Then, the approximate risk expression is of the form $\text{Risk}(\hat{\boldsymbol{\beta}}^{GR}) = \mathbb{E}\mathbb{E}_t t^{-1}(\boldsymbol{\varepsilon}^T Q^{*}\boldsymbol{\varepsilon}) - \mathbb{E}\mathbb{E}_t\left[t^{-1}\left(B_t(\theta) + C_t(\theta)\right)\right] + O(\theta^{-3})$. The required result follows by taking the expectation over the measure $W(\cdot)$.