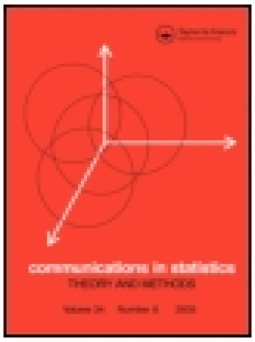

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

Shrinkage estimator of regression model under asymmetric loss

Zahirul Hoque, Jacek Wesolowski & Shahadut Hossain

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2017, VOL. 0, NO. 0, 1–11
https://doi.org/10.1080/03610926.2017.1397169

Taylor & Francis
Taylor & Francis Group

Check for updates

Shrinkage estimator of regression model under asymmetric loss

Zahirul Hoque[a], Jacek Wesolowski[b], and Shahadut Hossain[a]

[a]Department of Statistics, United Arab Emirates University, Al Ain, UAE; [b]Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

ABSTRACT
This article investigates the performance of the shrinkage estimator (SE) of the parameters of a simple linear regression model under the LINEX loss criterion. The risk function of the estimator under the asymmetric LINEX loss is derived and analyzed. The moment-generating functions and the first two moments of the estimators are also obtained. The risks of the SE have been compared numerically with that of pre-test and least-square estimators (LSEs) under the LINEX loss criterion. The numerical comparison reveals that under certain conditions the LSE is inadmissible, and the SE is the best among the three estimators.

ARTICLE HISTORY
Received 20 February 2017
Accepted 19 October 2017

KEYWORDS
Linear regression; Normal error; Shrinkage estimator; Pre-test estimator; LINEX loss and risk function.

MATHEMATICS SUBJECT CLASSIFICATION
Primary 62C12, 62C15; Secondary 62F10

1. Introduction

The main purpose of regression is prediction. Given the data on a response variable Y and its predictor X, our aim is to predict the value of Y as accurately as possible. On the other hand, the accuracy of the prediction of Y depends on the estimation of the model parameters. Aiming at the improvement of sample information-based maximum likelihood estimator (MLE) or the unrestricted estimator (UE), Bancroft (1944) pioneered the idea of combining the sample information-based MLE with the restricted estimator (RE). The RE is in the form of a known value of the parameter obtained from previous studies or expert knowledge. The resulting pre-test estimator (PTE) dominates the UE under the squared error loss criterion. Due to the construction of the PTE, there is no smooth transition between the UE and RE in it.

Stein (1956) and James and Stein (1961) showed that for three- or higher-dimensional population, there is estimator that dominates the usual UE under the squared error loss criterion. Their proposed estimator is widely known as the Stein-type shrinkage estimator (SE). One of the important reasons for the Stein-type estimator being so revolutionary is that many different aspects of something being measured are used to describe a single measurement (cf., Gruber 1998, p. 1). For example, many batting averages of different cricketers can be used to estimate the batting average of a single cricketer. The incidence of a disease in many different regions can be used to estimate the incidence of the disease in a small region. One can do this and produce a better estimator. The seminal discovery of the SE generated a flurry of studies by many researchers in search of improved estimators. Arashi and Tabatabaey (2010a) derived some conditions under which the Stein-type SE outperforms the UE under the squared error loss criterion. Recently, Fuqi and Nkurunziza (2016) has investigated the performance of the SE of the matrix of the regression coefficients in multivariate regression

CONTACT Zahirul Hoque Zahirul.Hoque@uaeu.ac.ae United Arab Emirates University, Al Ain, UAE.

models with unknown change points under the quadratic loss. Saleh and Sen (e.g., 1985, 1986) published a series of papers on the topic both in the parametric and in the non parametric context. For a comprehensive review on the area, readers may see Saleh (2005) and Saleh, Arashi, and Tabatabaey (2014).

The popularity of the symmetric squared error loss is due to its mathematical and interpretational convenience. Saleh and Sen (1986, 1985) and Saleh and Han (1990) are some excellent reviews of the use of the squared error loss. However, there was a growing criticism against the appropriateness of this loss function, particularly where overestimation of a parameter leads to more severe consequences than underestimation and vice versa. Due to the symmetric nature of the squared error loss function, it fails to differentiate between overestimation and underestimation of any parameter.

In a real estate study, Varian (1975) introduced an asymmetric loss function known as the LINEX loss function which has both linear and exponential components in its mathematical expression. This loss function can assign unequal weights to the underestimation and overestimation by taking an appropriate value of the shape parameter. For small values of the shape parameter, the LINEX loss function is approximately symmetric and not too far from the quadratic loss function (cf., Zellner 1986). Therefore, the latter can be considered as a special case of the former.

Zellner (1986) studied properties of estimation and prediction procedures under the LINEX loss function. He showed that some usual estimators that are admissible relative to the squared error loss function are inadmissible relative to the LINEX loss function. For example, Zellner (1986) proved that the MLE $\bar{X}$ of the univariate normal mean is inadmissible relative to the LINEX loss function, as the risk of the estimator $\bar{X} - a\sigma^2/2n$ is less than that of the MLE, where a is the shape parameter of the LINEX loss function, σ^2 is the population variance, and n is the size of the sample. In the case of unknown σ^2, he suggested using $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$. Later, Rojo (1987) generalized Zellner's result and showed that under the LINEX loss function any estimator of the form $c\bar{X} + d$, of θ is admissible if either $0 \le c < 1$, or $c = 1$ and $d = -a\sigma^2/2n$. Otherwise, $c\bar{X} + d$ is inadmissible.

More recent studies devoted to properties of estimators under the LINEX loss include, e.g., Christoffersen and Diebold (1997), Zou (1997), Misra, Iyer, and Singh (2004), Zieliński (2005), Hoque, Khan, and Wesolowski (2009), Arashi and Tabatabaey (2010), Ohyauchi (2013), and Ma and Liu (2013). In particular, Hoque, Khan, and Wesolowski (2009) studied the performance of the PTE of the slope parameter of simple linear regression model under the LINEX loss function and reaffirmed the superiority of PTE over the least-square estimator (LSE). Later, Hoque and Hossain (2012) compared the PTE of the intercept of simple linear regression model with the LSE and revealed similar result.

The main purpose of this article is to investigate the properties of the SE of the parameter of simple linear regression model under the LINEX loss function. As the first component of the risk function under LINEX loss is the moment-generating function (MGF), the MGF of the estimator is obtained in the derivation of the risk function. The risk of the estimator under the squared error loss is derived from the risk under the LINEX loss which conforms to that in the existing literature. The first two moments of the estimators are also obtained. Moreover, we solve the issue of the optimal shrinkage constant. Similar to the form of the LINEX loss function, the form of the risk function is also asymmetric. However, for very small value of the shape parameter of the LINEX loss function, the form of the risk function is almost symmetric. The comparative performance of SE, PTE, and LSE under the LINEX loss is shown graphically and numerically. It reveals that if the non sample prior information about the value of the parameter is not too far from its true value, both the SE and the PTE outperform the LSE. Among the three estimators, SE performs the best. The layout of the article is as follows.

The description of the model, the estimator, and the LINEX loss function is presented in Section 2. The derivation of the risk function of the SE along with some properties is presented in Section 3. In particular, in Subsection 3.1 we consider the issue of finding the optimal shrinkage constant. The numerical comparison of the risk functions of the SE, PTE, and LSE is laid in Section 4. Finally, some concluding remarks are presented in Section 5.

2. Preliminaries

Consider the model

$$Y_i = \beta x_i + \sigma \epsilon_i, \quad i = 1, \dots, n,$$

where x_i, $i = 1, \dots, n$, are given real numbers, and ϵ_i, $i = 1, \dots, n$, are iid standard normal rv's. We want to estimate $\beta \in \mathbb{R}$ while $\sigma > 0$ is not known.

Alternatively, we can write

$$\mathbf{Y} = \beta \mathbf{x} + \sigma \epsilon, \tag{1}$$

where $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{x} = (x_1, \dots, x_n)^T$, and $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$. The traditional LSE of β is

$$\tilde{\beta} = \frac{\mathbf{x}^T \mathbf{Y}}{\mathbf{x}^T \mathbf{x}}. \tag{2}$$

Based on the pioneering work of Bancroft (1944), an alternative (to the LSE of β) estimator is the PTE which is defined as

$$\hat{\beta}_1^{\text{PTE}} = \tilde{\beta} - (\tilde{\beta} - \beta_0)\, I(F < F_{1,\nu}(\alpha)) \tag{3}$$

where β_0 is the value of β usually obtained from expert knowledge or previous studies, $F = \frac{\tilde{\beta} - \beta_0}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \tilde{\beta} x_i)^2}$, and $F_{1,\nu}(\alpha)$ is the α-level critical value of the F statistic with 1 and $\nu = n - 1$ degrees of freedom (d.f.) to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$. Under the alternative hypothesis, the distribution of F is a non central F with $(1, \nu)$ d.f. and non centrality parameter Δ^2 where $\Delta = \frac{(\beta - \beta_0)\sqrt{\mathbf{x}^T \mathbf{x}}}{\sigma}$.

Following Khan and Saleh (2001) we define the SE of β which has the form

$$\hat{\beta}_c^{\text{SE}} = \tilde{\beta} - \frac{c\, S_n (\tilde{\beta} - \beta_0)}{\sqrt{\mathbf{x}^T \mathbf{x}} |\tilde{\beta} - \beta_0|}, \tag{4}$$

where $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ and c is the shrinkage constant. The optimum value of c is obtained by minimizing the risk of the estimator.

Our interest in Section 3 focuses on the properties of the SE under an asymmetric LINEX loss criterion and in Section 4 we compare numerically performance of the SE, PTE and LSE under the LINEX loss. The LINEX loss function is defined as

$$\text{L}(\delta) = b[\exp(a\delta) - a\delta - 1], \qquad \forall\, a \neq 0,\ b > 0 \tag{5}$$

where $\delta = (\theta^* - \theta)$ is the estimation error. The two parameters a and b in $\text{L}(\delta)$ serve to determine the shape and scale, respectively, of $\text{L}(\delta)$. A positive value of a indicates that overestimation is more serious than underestimation and a negative value of a represents the reverse situation. The magnitude of a reflects the degree of asymmetry about $\delta = 0$. This asymmetric loss function grows approximately linearly on one side of $\delta = 0$ and grows approximately exponentially on the other side. If $a \to 0$, then the LINEX function reduces to the squared error loss function. Without loss of generality, throughout the paper we assume that the scale parameter $b = 1$. As a approaches 0, the growth pattern of LINEX loss becomes similar for both positive and negative errors of estimation and approaches the quadratic loss. Hence, the LINEX loss function is more general than the quadratic loss function. Further discussion about the properties of this loss function is in Zellner (1986).

In the following section we show the derivation of the LINEX risk function of the SE and study its properties. In Section 4 we recall already known form of the risk function of the PTE and the LSE and we are interested in the comparison of the performance of the three estimators, SE, PTE and LSE under the LINEX loss function.

3. The risk functions

The focus of this paper is in the performance of SE under LINEX loss function. Therefore, we are interested in the LINEX risk function of the SE, which is defined as

$$\mathrm{R}_a[\hat{\beta}_c^{\mathrm{SE}}] = \mathbb{E}\, e^{a\left(\hat{\beta}_c^{\mathrm{SE}} - \beta\right)} - a\mathbb{E}(\hat{\beta}_c^{\mathrm{SE}} - \beta) - 1 \tag{6}$$

where $a \in \mathbb{R}$ is the shape parameter of the LINEX loss function.

We first recall some basic properties of the random variables and functions related to the SE and its properties.

- The random variable

$$U := \frac{\sqrt{\mathbf{x}^T\mathbf{x}}}{\sigma}(\tilde{\beta} - \beta), \tag{7}$$

where $\tilde{\beta}$ is defined in (2), has the standard normal distribution, $U \sim \mathrm{N}(0, 1)$;

- the random variable

$$W^2 := \frac{(n-1)S_n^2}{\sigma^2} \tag{8}$$

has the chi-square distribution with $n-1$ d.f., $W^2 \sim \chi^2_{n-1}$;

- the random variables U and W are independent, where W is defined as the positive square root of the sum of squares of independent standard normal random variables. Therefore, $W \sim \chi_{n-1}$ which is known as Chi distribution and its MGF (see, e.g., Johnson, Kotz, and Balakrishnan 1994, Ch. 18 for the formulas for the moments) has the form

$$\mathbb{E}\, e^{sW} = \sum_{k=0}^{\infty} \frac{\mathbb{E}\, W^k}{k!} s^k = \sum_{k=0}^{\infty} \frac{2^{k/2}\Gamma\left(\frac{n-1+k}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)k!} s^k,$$

and, consequently,

$$\mathbb{E}\, e^{sW} = {}_1F_1\left(\tfrac{n-1}{2}, \frac{1}{2}, \tfrac{s^2}{2}\right) + sM(n)\, {}_1F_1\left(\tfrac{n}{2}, \frac{3}{2}, \tfrac{s^2}{2}\right), \quad s \in \mathbb{R}, \tag{9}$$

where

$$M(n) = \sqrt{2}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

and ${}_1F_1(a, b, x)$ is a confluent hypergeometric function which is also known as the Kummer function (see Abramovitz and Stegun 1964, Ch. 13).

- Since

$$ {}_1F_1(a, b, x) = \sum_{k=0}^{\infty} \frac{(a)^k}{(b)^k\, k!}\, x^k, \quad x \in \mathbb{R}, \tag{10} $$

where $(r)^k = r(r+1)\dots(r+k-1)$, $k \geq 1$, and $(r)^0 = 1$ is the ascending Pochhammer symbol; differentiating (10) with respect to x we immediately get (see formula 13.4.8 in Ch. 13 of Abramovitz and Stegun 1964).

$$ \frac{d\, {}_1F_1(a, b, x)}{dx} = \frac{a}{b}\, {}_1F_1(a+1, b+1, x), \qquad x \in \mathbb{R}. \tag{11} $$

Now we are in a position to derive the risk function of the SE under the LINEX loss function.

Proposition 3.1. *In the model (1) the LINEX risk of the SE (4) has the form*

$$ \begin{aligned} R_a[\hat{\beta}_c^{SE}] = e^{\frac{a^2K^2}{2}} \Bigg[{}_1F_1\left(\frac{n-1}{2}, \frac{1}{2}, \frac{a^2c^2K^2}{2(n-1)}\right) \\ + \frac{acKM(n)\,(1 - 2\Phi\,(aK + \Delta))}{\sqrt{n-1}}\, {}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{a^2c^2K^2}{2(n-1)}\right) \Bigg] \\ - \frac{acKM(n)(1 - 2\Phi(\Delta))}{\sqrt{n-1}} - 1, \end{aligned} \tag{12} $$

where $K = \frac{\sigma}{\sqrt{\mathbf{x}^T\mathbf{x}}}$, $\Delta = \frac{\beta - \beta_0}{K}$, *and* Φ *is the cumulative distribution function of the standard normal distribution.*

Proof. Due to (7), we can write

$$ \hat{\beta}_c^{SE} - \beta = \tilde{\beta} - \beta - c\frac{S_n(\hat{\beta} - \beta + \beta - \beta_0)}{\sqrt{\mathbf{x}^T\mathbf{x}}(\hat{\beta} - \beta + \beta - \beta_0)} = \frac{\sigma}{\sqrt{\mathbf{x}^T\mathbf{x}}}\, U - c\frac{S_n\left(\frac{\sigma}{\sqrt{\mathbf{x}^T\mathbf{x}}}U + \beta - \beta_0\right)}{\sqrt{\mathbf{x}^T\mathbf{x}}\left|\frac{\sigma}{\sqrt{\mathbf{x}^T\mathbf{x}}}U + \beta - \beta_0\right|}. $$

Consequently,

$$ \hat{\beta}_c^{SE} - \beta = KU - \frac{cKW}{\sqrt{n-1}}\, \frac{U + \Delta}{|U + \Delta|}. $$

Therefore,

$$ I(a) = \mathbb{E}\, e^{a(\hat{\beta}_c^{SE} - \beta)} = \mathbb{E}\, e^{aKU}\, \exp\left(-\frac{acKW}{\sqrt{n-1}}\, \frac{U + \Delta}{|U + \Delta|}\right). \tag{13} $$

Recall (see, e.g., Hoque, Khan, and Wesolowski 2009) that if Z and S are independent, $Z \sim \mathrm{N}(0, 1)$, $\phi : \mathbb{R}^2 \to \mathbb{R}$ is a Borel function and $c \in \mathbb{R}$ then

$$ \mathbb{E}\, e^{cZ}\phi(Z, S) = e^{\frac{c^2}{2}}\, \mathbb{E}\, \phi(Z + c, S). \tag{14} $$

Due to independence of U and W, we may apply (14) to $I(a)$ as given in (13). It yields

$$ I(a) = e^{\frac{a^2K^2}{2}}\, \mathbb{E}\, \exp\left(-\frac{acKW}{\sqrt{n-1}}\, \frac{U + aK + \Delta}{|U + aK + \Delta|}\right). $$

Again using the independence of U and W, we see that

$$\begin{aligned}\mathbb{E}\exp&\left(-\frac{acKW}{\sqrt{n-1}}\frac{U+aK+\Delta}{|U+aK+\Delta|}\right)\\ &=\Phi(aK+\Delta)\mathbb{E}\exp\left(-\frac{acKW}{\sqrt{n-1}}\right)+(1-\Phi(aK+\Delta))\mathbb{E}\exp\left(\frac{acKW}{\sqrt{n-1}}\right). \quad (15)\end{aligned}$$

Note that the right-hand side of (15) is a convex combination of the MGF of the χ_{n-1} distribution evaluated at z and $-z$, where $z=\frac{acK}{\sqrt{n-1}}$. Therefore, we apply (9) and get

$$I(a)=e^{\frac{a^2K^2}{2}}\left[{}_1F_1\left(\frac{n-1}{2},\frac{1}{2},\frac{z^2}{2}\right)+M(n)(1-2\Phi(aK+\Delta))z\,{}_1F_1\left(\frac{n}{2},\frac{3}{2},\frac{z^2}{2}\right)\right].$$

The derivative of I with respect to a evaluated at $a=0$ gives (up to scaling by a) the second term in the expression (6) for the LINEX risk. Remembering the fact that z depends on a, we get (we rely here again on (11))

$$\begin{aligned}\frac{dI}{da}&=aK^2I_1(a)+e^{\frac{a^2K^2}{2}}\left[aK^2c^2\,{}_1F_1\left(\frac{n+1}{2},\frac{3}{2},\frac{a^2K^2c^2}{2(n-1)}\right)+M(n)\right.\\ &\quad\times(1-2\Phi(aK+\Delta))\frac{Kc}{\sqrt{n-1}}\,{}_1F_1\left(\frac{n}{2},\frac{3}{2},\frac{a^2K^2c^2}{2(n-1)}\right)\\ &\quad+\frac{na^2K^2c^2}{3(n-1)}M(n)(1-2\Phi(aK+\Delta))\\ &\quad\left.\times{}_1F_1\left(\frac{n+2}{2},\frac{5}{2},\frac{a^2K^2c^2}{2(n-1)}\right)-\frac{2aK^2cM(n)}{\sqrt{n-1}}\,{}_1F_1\left(\frac{n}{2},\frac{3}{2},\frac{a^2K^2c^2}{2(n-1)}\right)\phi(aK+\Delta\right], \quad (16)\end{aligned}$$

where ϕ is the probability distribution function of the standard normal distribution.

Plugging $a=0$ in (16), we get

$$\mathbb{E}\left(\hat{\beta}_c^{SE}-\beta\right)=\left.\frac{dI}{da}\right|_{a=0}=\frac{cKM(n)(1-2\Phi(\Delta))}{\sqrt{n-1}},$$

where we used also the fact that ${}_1F_1(a,b,0)=1$. For another proof of the above formula for the bias of the SE, the reader may refer to Khan, Hoque, and Saleh (2002).

Combining the formula for $I(a)$ with the above expression, we get the form of the LINEX risk of the SE, $\hat{\beta}_c^{SE}$, as given in (12). □

Remark 3.1. We note that the MSE of the SE can be easily obtained from the form of the LINEX risk by evaluating its second derivative at $a=0$. Therefore, we differentiate (16), use again (11) and plug $a=0$, which finally gives

$$\left.\frac{d^2I}{da^2}\right|_{a=0}=K^2+K^2c^2-4\frac{K^2cM(n)}{\sqrt{n-1}}\phi(\Delta)=\frac{\sigma^2}{\mathbf{x}^T\mathbf{x}}\left(1+c^2-4\frac{cM(n)}{\sqrt{n-1}}\phi(\Delta)\right).$$

This form of the MSE of the SE of β has already been derived by a different method in Khan, Hoque, and Saleh (2002).

Remark 3.2. Of course, it is of interest to know the approximate value of the LINEX risk for large expert evaluation mistakes, that is for $\Delta \to \pm\infty$. Directly from (12), we get

$$\lim_{\Delta\to\pm\infty} R_a[\hat{\beta}_c^{SE}] = e^{\frac{a^2K^2}{2}} \left[{}_1F_1\left(\frac{n-1}{2}, \frac{1}{2}, \frac{a^2c^2K^2}{2(n-1)}\right) \mp \frac{acKM(n)}{\sqrt{(n-1)}}\, {}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{a^2c^2K^2}{2(n-1)}\right)\right] \pm \frac{acKM(n)}{\sqrt{n-1}} - 1.$$

3.1. Optimal shrinkage constant

In this subsection, we are interested in a search of the shrinkage constant c which minimizes the LINEX risk of SE for a given shape parameter a of the loss function.

Proposition 3.2. *For a given shape parameter a of the LINEX loss function, there exists a unique shrinkage constant $c^* = c^*(a, K, n, \Delta)$ which minimizes $R_a[\hat{\beta}_c^{SE}]$. It is the unique root of the equation*

$$M(n)(1-2\Phi(\Delta)) = e^{\frac{a^2K^2}{2}} \left[acK\sqrt{n-1}\,{}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{a^2c^2K^2}{2(n-1)}\right) + M(n)\left(1-2\Phi\left(aK+\Delta\right)\right) \times \left({}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{a^2c^2K^2}{2(n-1)}\right) + \frac{a^2c^2K^2}{n-1}\,{}_1F_1\left(\frac{n+2}{2}, \frac{5}{2}, \frac{a^2c^2K^2}{2(n-1)}\right)\right)\right].$$

In the case $\Delta = 0$, the optimal $c^* = c_0^*(a)$ is the unique root of the equation

$$acK\sqrt{n-1}\,{}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{a^2c^2K^2}{2(n-1)}\right) = M(n)\left(2\Phi\left(aK\right)-1\right)\left({}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{a^2c^2K^2}{2(n-1)}\right) + \frac{a^2c^2K^2}{n-1}\,{}_1F_1\left(\frac{n+2}{2}, \frac{5}{2}, \frac{a^2c^2K^2}{2(n-1)}\right)\right).$$

Proof. Denote

$$F(a, K, n, \Delta, z) = e^{\frac{a^2K^2}{2}} \left[{}_1F_1\left(\frac{n-1}{2}, \frac{1}{2}, \frac{z^2}{2}\right) + M(n)\left(1-2\Phi\left(aK+\Delta\right)\right) z\,{}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{z^2}{2}\right)\right] - zM(n)(1-2\Phi(\Delta)),$$

which is, up to a shift by -1, the risk function.

We are looking for a minimum of F as a function of z. To this aim, we differentiate F with respect to z and use the identity (11) for derivatives of ${}_1F_1$ functions. Consequently, we get

$$F'(z) := \frac{\partial F}{\partial z} = -M(n)(1-2\Phi(\Delta)) + e^{\frac{a^2K^2}{2}} \left[z(n-1)\,{}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{z^2}{2}\right) + M(n)\left(1-2\Phi\left(aK+\Delta\right)\right)\left({}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{z^2}{2}\right) + z^2\,{}_1F_1\left(\frac{n+2}{2}, \frac{5}{2}, \frac{z^2}{2}\right)\right)\right]. \quad (17)$$

We will now show that the function F' is continuous and strictly increasing from $-\infty$ to $+\infty$.

Consider $\mathbb{E}\exp(a(R_a[\hat{\beta}_c^{SE}] - \beta))$ as a function of $z = \frac{acK}{\sqrt{n-1}}$. Denote this function as G. Note that the form of the risk function, as given in (12), implies that G' and F' differ only by a constant shift. Therefore, it is sufficient to show that G' is continuous and strictly increasing from $-\infty$ to $+\infty$. Due to (15), we see that this derivative equals

$$G'(z) = -\Phi(aK+\Delta)\mathbb{E}\,We^{-zW} + (1-\Phi(aK+\Delta))\mathbb{E}\,We^{zW},$$

and hence, G' is a continuous function. Moreover, the second derivative

$$G''(z) = \Phi(aK + \Delta)\mathbb{E}\, W^2 e^{-zW} + (1 - \Phi(aK + \Delta))\mathbb{E}\, W^2 e^{zW}$$

is strictly positive. That is, G' is strictly increasing. Since W follows the Chi distribution, it follows immediately from the formula for $G'(z)$ that $\lim_{z\to\pm\infty} G'(z) = \pm\infty$.

Consequently, F' has a single zero which can be obtained numerically. Denote this zero by $z^*(a)$. Due to the relation between z and c, we see that the optimal shrinkage constant is

$$c^* = c^*(a) = \frac{z^*(a)\sqrt{n-1}}{aK}. \tag{18}$$

□

Remark 3.3. Consider the case of $\Delta = 0$. For the unique zeroes of $F'(a, z) := F'(z)$, we have $z^*(a) = -z^*(-a)$.

To see this fact using (17), we calculate

$$F'(-a, -z^*(a)) = e^{\frac{a^2K^2}{2}} \left[-z^*(a)(n-1)\,{}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{(z^*(a))^2}{2}\right) + M(n)\,(1 - 2\Phi\,(-aK)) \right.$$
$$\left. \times \left({}_1F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{(z^*(a))^2}{2}\right) + (z^*(a))^2\, {}_1F_1\left(\frac{n+2}{2}, \frac{5}{2}, \frac{(z^*(a))^2}{2}\right)\right)\right].$$

Now, by the identity $1 - 2\Phi(-aK) = -(1 - 2\Phi(aK))$, the above equality leads to

$$F'(-a, -z^*(a)) = -F'(a, z^*(a)) = 0$$

and thus, by the uniqueness of the root, we see that $-z^*(a) = z^*(-a)$.

Therefore, by the formula relating c^* and z^*, we obtain

$$c_0^*(a) = c_0^*(-a).$$

4. Numerical analysis of the risk functions

In this section, we pursue a comparative analysis of the LINEX risk performance of the SE with the PTE and the LSE.

To this end, we need to recall two formulas for the LINEX loss for the PTE and LSE. Actually, they were derived in a slightly more general linear regression model (with non zero intercept) in Hoque, Khan, and Wesolowski (2009). The modification for the model with the zero intercept is straightforward.

Remark 4.1. The risk function of the PTE of the parameter β under the LINEX loss function is given by

$$\mathrm{R}[\hat{\beta}^{\mathrm{PTE}}; \beta] = \exp(-aK\Delta)G_{1,\nu}(c;\, \Delta^2) + \exp(a^2K^2/2)[1 - G_{1,\nu}(c;\, (\Delta + aK)^2)]$$
$$+ aK\Delta\, G_{3,\nu}(c/3;\ \Delta^2) - 1 \tag{19}$$

where $c = F_{1,\nu}(\alpha)$ and $G_{\mu,\nu}(q;\ \theta)$ is the cumulative distribution function of the non central F distribution with $(\mu,\ \nu)$ d.f., non centrality parameter θ, and evaluated at q.

Remark 4.2. The risk function of the LSE of the parameter β under the LINEX loss function is given by

$$\mathrm{R}[\hat{\beta}^{\mathrm{LSE}}; \beta] = \exp(a^2K^2/2) - 1. \tag{20}$$

From Equation (18), it is evident that the optimal value of the shrinkage constant c is obtainable only through numerical calculation, and hence the analytical analysis of the risk

Table 1. Maximum and minimum efficiencies of SE of β and values of Δ at which the minimum efficiency occurs for $a = 3$ and selected α and n.

		Sample size, n						
α		20	25	30	35	40	45	50
0.05	†Efl_{1*}	37.15	31.66	28.19	26.21	24.87	23.89	23.15
	‡Efl_{2*}	2.75	2.74	2.74	2.74	2.74	2.74	2.74
	secEfl_{1o}	1.56	1.58	1.59	1.60	1.60	1.614	1.61
	¶Efl_{2o}	0.36	0.37	0.37	0.37	0.37	0.37	0.37
	††Δ_{1o}	− 1.08	− 1.08	− 1.08	− 1.08	− 1.08	− 1.10	− 1.10
	‡‡Δ_{2o}	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00
0.10	Efl_{1*}	11.34	10.27	9.65	9.25	8.97	8.76	8.60
	Efl_{2*}	2.75	2.74	2.74	2.74	2.74	2.74	2.74
	Efl_{1o}	1.81	1.82	1.83	1.83	1.84	1.84	1.84
	Efl_{2o}	0.36	0.37	0.37	0.37	0.37	0.37	0.37
	Δ_{1o}	− 1.23	− 1.23	− 1.23	− 1.25	− 1.25	− 1.25	− 1.25
	Δ_{2o}	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00
0.15	Efl_{1*}	5.62	5.28	5.07	4.93	4.83	4.76	4.70
	Efl_{2*}	2.75	2.75	2.74	2.74	2.74	2.74	2.74
	Efl_{1o}	1.91	1.91	1.92	1.92	11.92	1.92	1.92
	Efl_{2o}	0.36	0.37	0.37	0.37	0.37	0.37	0.37
	Δ_{1o}	− 1.40	− 1.40	− 1.40	− 1.40	− 1.43	− 1.43	− 1.43
	Δ_{2o}	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00
0.20	Efl_{1*}	3.43	3.29	3.20	3.11	3.10	3.07	3.05
	Efl_{2*}	2.75	2.74	2.74	2.74	2.74	2.74	2.74
	Efl_{1o}	1.92	1.92	1.92	1.92	1.92	1.92	1.90
	Efl_{2o}	0.36	0.37	0.37	0.37	0.37	0.37	0.37
	Δ_{1o}	− 1.60	− 1.60	− 1.63	− 1.63	− 1.63	− 1.63	− 1.63
	Δ_{2o}	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00	− 5.00

†maximum efficiency of SE relative to PTE.
‡maximum efficiency of SE relative to LSE.
secminimum efficiency of SE relative to PTE.
¶minimum efficiency of SE relative to LSE.
††the value of Δ at which the minimum efficiency of SE relative to PTE occurs.
‡‡the value of Δ at which the minimum efficiency of SE relative to LSE occurs.

function of SE is mathematically intractable. Therefore, we proceed with numerical analysis. We use the Newton–Raphson method to calculate the optimum value of c for different values of n and a, and proceed with the numerical calculation of the risk of SE. The calculation has been implemented in R programming. For simplicity, without any loss of generality throughout the article, we assume $K = 1$ for calculation of c as well as the calculation of the risks.

For easy understanding, we define the efficiency of SE relative to PTE and LSE. For selected values of α and n, the efficiencies of SE relative to PTE and LSE are calculated and presented in Table 1 and plotted in Figure 1.

From Figure 1, it is clear that for any value of Δ, SE is more efficient than PTE and LSE regardless of the values of a as long as Δ is not too far from 0. For $a = \pm 2$, when $\Delta = 0$, SE outperforms both PTE and LSE. As Δ grows from 0, the efficiency of SE starts decreasing for both positive and negative values of Δ. However, the rate of decrease is higher in the negative side of Δ for positive a and opposite for negative a. For large value of Δ, the performance of SE is worse than that of the PTE and LSE. Nevertheless, as the value of Δ depends on the accuracy of the expert knowledge and/or previous studies, any large absolute value of Δ is very unlikely. Therefore, typically, the performance of SE is better than that of both PTE and LSE.

To reconfirm the dominance of SE over PTE and LSE, we present the efficiency of SE relative to PTE (for different values of α) and LSE in Table 1 for various sample sizes.

In the table, the maximum and minimum efficiencies of SE relative to PTE and LSE are presented for $a = 3$. The efficiency was calculated for various sample sizes and selected values of

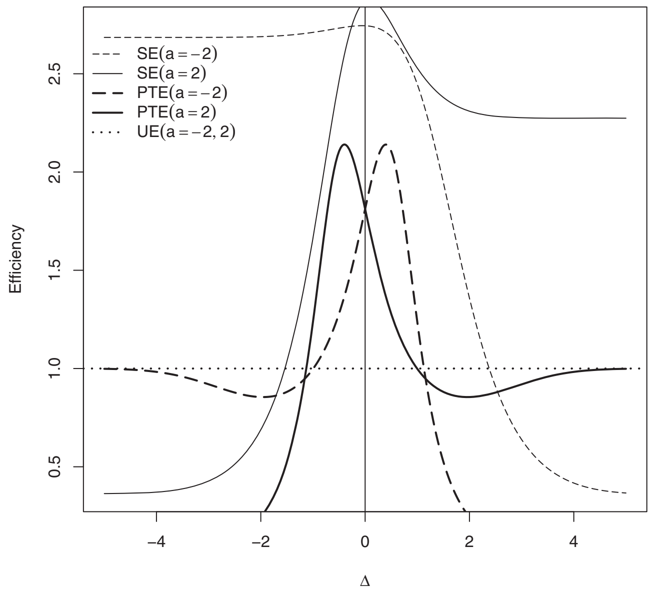

Figure 1. The plot of the efficiency of SE and PTE relative to LSE for selected values of a.

α, the level of significance of the PTE. We also present the minimum guaranteed efficiency of SE (relative to both PTE and LSE) for given sample sizes and the values of Δ at which the minimum guaranteed efficiency occurs. For example, when $\alpha = 0.05$ and $n = 25$, the maximum efficiencies of SE relative to PTE and LSE are 31.66 and 2.74, respectively. On the other hand, for the same α and n, the minimum guaranteed efficiencies of SE relative to PTE and LSE are 1.58 and 0.37, respectively. These minimum efficiencies occur at $\Delta = -1.08$ and -5, respectively. For given α and n, both the maximum and minimum efficiency of SE relative to PTE are higher than that relative to the LSE. As the sample size becomes larger, the maximum efficiency decreases but the minimum guaranteed efficiency increases. It is clear that in the neighborhood of $\Delta = 0$, both SE and PTE outperform LSE, but SE performs the best among the three estimators. From Figure 1, it is also clearly evident that SE outperforms PTE in a wider neighborhood of $\Delta = 0$. The LINEX loss function and the efficiency of the estimators under LINEX loss behave alike which is asymmetric in nature.

5. Concluding remarks

The three estimators considered in this article have been compared for their performance under the asymmetric LINEX loss function. The foregoing analysis reaffirms the superiority of SE over the PTE and LSE that was shown by the pioneering work of Stein (1956) and Bancroft (1944) and many thereafter under the symmetric squared error loss. Therefore, the practitioners of SE can use it without incurring higher risk of the estimator in both symmetric and asymmetric loss criteria.

References

Abramovitz, M., and I. A. Stegun. 1964. *Handbook of mathematical functions*. US National Bureau of Standards.

Arashi, M., and S. M. M. Tabatabaey. 2010. Estimation of the location parameter under LINEX loss function: multivariate case. *Metrika* 72 (1):51–7.

Arashi, M., and S. M. M. Tabatabaey. 2010a. A note on classical Stein-type estimators in elliptically contoured models. *Journal of Statistical Planning and Inference* 140:1206–13.

Bancroft, T. A. 1944. On biases in estimation due to the use of the preliminary test of significance. *Annals of Mathematical Statistics* 15:190–204.

Christoffersen, P. F., and F. X. Diebold. 1997. Optimal prediction under asymmetric loss. *Econometric Theory* 13:808–17.

Fuqi, C., and S. Nkurunziza. 2016. A class of Stein-rules in multivariate regression model with structural changes. *Scandinavian Journal of Statistics, Theory and Applications* 43:83–102.

Gruber, M. H. J. 1998. *Improving efficiency by shrinkage: The James–Stein and ridge regression estimators*. New York: Mercel Dekker.

Hoque, Z., S. Khan, and J. Wesolowski. 2009. Performance of preliminary test estimator under linex loss function. *Communications in Statistics, Theory and Methods* 38:252–61.

Hoque, Z., and S. Hossain. 2012. Improved estimation in regression with varying penalty. *Journal of Statistical Theory and Practice* 6:260–73.

James, W., and C. Stein. 1961. Estimation with quadratic loss. In J. Neyman (ed.), *Proceedings of the fourth berkeley symposium on mathematical statistics and probability*, 1, 361–79. Berkeley, California: University of California Press.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*, Vol. 1. New York: Wiley.

Khan, S., and A. K. Md. E. Saleh. 2001. On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. *Statistical Papers* 42 (4):451–73.

Khan, S., Z. Hoque, and A. K. Md. E. Saleh. 2002. Estimation of the slope parameter for linear regression model with uncertain prior information. *Journal of Statistical Research* 36 (1):55–73.

Ma, T., and S. Liu. 2013. Estimation of order-restricted means of two normal populations under the LINEX loss function. *Metrika* 76 (3):409–52.

Misra, N., S. K. Iyer, and H. Singh. 2004. The LINEX risk of maximum likelihood estimators of parameters of normal populations having order restricted means. *Sankhya* 66 (4):652–77.

Ohyauchi, N. 2013. Comparison of risks of estimators under the LINEX loss for a family of truncated distributions. *Statistics* 47 (3):590–604.

Rojo, J. 1987. On the admissibility of $c\bar{X} + d$ with respect to the linex loss function. *Communications in Statistics, Theory and Methods* 16 (12):3745–8.

Saleh, A. K. Md. E. 2005. *Theory of preliminary test and stein-type estimation with applications*. New Jersey: John Wiley & Sons.

Saleh, A. K. Md. E., A. Arashi, and S. M. M. Tabatabaey. 2014. *Statistical inference for models with multivariate t-distributed errors*. New Jersey: John Wiley & Sons.

Saleh, A. K. Md. E., and P. K. Sen. 1985. Shrinkage least squares estimation in a general multivariate linear model. In *Proceedings of the Fifth Pannonian Symposium on mathematical statistics*, 307–25. Amsterdam, 275–297.

Saleh, A. K. Md. E., and P. K. Sen. 1986. On shrinkage *R*-estimation in a multiple regression model. *Communication in Statistics, Theory and Methods* 15 (7):2229–44.

Saleh, A. K. Md. E., and C. P. Han. 1990. Shrinkage estimation in regression analysis. *Estadistica* 42:40–63.

Stein, C. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics, and probability*, J. Neyman, ed., 1, 197–206. Berkeley, California: University of California Press.

Varian, H. R. 1975. A Bayesian approach to real estate assessment. In *Studies in Bayesian econometrics and statistics in honor of L. J. Savage*, ed. Feinberg and A. Zellner, 195–208. Amsterdam: North Holland.

Zellner, A. 1986. Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association, Theory and Methods* 81 (394):446–51.

Zieliński, R. 2005. Estimating quantiles with LINEX loss function. Applications to VaR estimation. *Applications Mathematics* 32 (4):367–73.

Zou, G. 1997. Admissible estimation for finite population under the LINEX loss function. *Journal of Statistical Planning and Inference* 61 (2):373–84.