Ann. Inst. Statist. Math.
Vol. 47, No. 2, 273–286 (1995)

SOME MODIFICATIONS OF IMPROVED ESTIMATORS OF A NORMAL VARIANCE

NOBUO SHINOZAKI

Department of Administration Engineering, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku, Yokohama 223, Japan

(Received May 12, 1994; revised November 24, 1994)

Abstract. Consider the problem of estimating a normal variance based on a random sample when the mean is unknown. Scale equivariant estimators which improve upon the best scale and translation equivariant one have been proposed by several authors for various loss functions including quadratic loss. However, at least for quadratic loss function, improvement is not much. Herein, some methods are proposed to construct improving estimators which are not scale equivariant and are expected to be considerably better when the true variance value is close to the specified one. The idea behind the methods is to modify improving equivariant shrinkage estimators, so that the resulting ones shrink little when the usual estimate is less than the specified value and shrink much more otherwise. Sufficient conditions are given for the estimators to dominate the best scale and translation equivariant rule under the quadratic loss and the entropy loss. Further, some results of a Monte Carlo experiment are reported which show the significant improvements by the proposed estimators.

Key words and phrases: Entropy loss, quadratic loss, shrinkage estimator, Stein estimator, uniform risk improvement.

1. Introduction

Suppose that we want to estimate the variance, σ^2, based on a random sample $X_1, X_2, \ldots, X_n$ from a normal distribution with unknown mean μ. For quadratic loss function

$$L_1(\hat{\sigma}^2, \sigma^2) = \left(\frac{\hat{\sigma}^2}{\sigma^2} - 1\right)^2, \tag{1.1}$$

the estimator

$$\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1.2}$$

with $\bar{X} = \sum_{i=1}^{n} X_i/n$ is best among all scale and translation equivariant procedures and is minimax with constant risk. However, Stein (1964) showed that it is inadmissible and is dominated by the shrinkage estimator

$$\min\left\{\frac{1}{n+1}\sum_{i=1}^{n}(X_i-\bar{X})^2, \frac{1}{n+2}\sum_{i=1}^{n}X_i^2\right\}. \tag{1.3}$$

Extending Brown's technique (1968), Brewster and Zidek (1974) constructed a minimax and generalized Bayes estimator which has been shown to be admissible (Proskin (1985)). Using a different technique, Strawderman (1974) obtained a class of minimax estimators some of which are generalized Bayes. Rukhin (1987) has investigated the relative risk improvement of the Brewster-Zidek estimator and locally optimal minimax shrinkage ones given by himself. He has observed that the maximum relative risk improvement is at most 3 or 4 % for both estimators. A good account of these results can be found in Maatta and Casella (1990), where confidence estimation is also considered.

Brown (1968) derived estimators which improve upon the best equivariant estimator for more general loss functions. In particular, for the entropy loss function

$$L_2(\hat{\sigma}^2, \sigma^2) = \frac{\hat{\sigma}^2}{\sigma^2} - \log\left(\frac{\hat{\sigma}^2}{\sigma^2}\right) - 1, \tag{1.4}$$

the best equivariant estimator

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \tag{1.5}$$

is improved. Brewster and Zidek (1974) showed that a Stein-type shrinkage estimator dominates (1.5) for the entropy loss (1.4). They also derived improving generalized Bayes estimator for a variety of loss functions. Madi (1993) has derived a class of improving estimators based on several independent samples for a large class of loss functions. They are similar to those of Brewster and Zidek (1974).

Recently, Kubokawa (1994) has given a unified approach to improving equivariant estimators by using a definite integral, and has provided a class of improving procedures in both point and interval estimation for a wide class of distributions and loss functions.

The improving estimators proposed so far are all scale equivariant. However, in here we introduce better estimators which are not scale equivariant. We construct them by modifying scale equivariant improving estimators so that they are much better in some specified region at the expense of smaller improvement outside it.

As a matter of fact, we propose three methods to modify improving estimators when σ^2 is thought to be near σ_0^2. Without loss of generality we set $\sigma_0^2 = 1$. The idea which is common to the three methods is that, we should shrink a little if the usual estimate is less than 1 and much more if the estimate is much larger than 1. This will make the improvement larger for σ^2 in the neighborhood of 1 at the expense of smaller improvement outside it.

In Section 2, we describe these methods for the quadratic loss and give minimaxity results for the resulting estimators. We show that a similar argument holds for the entropy loss in Section 3. In Section 4, we give some results of a Monte Carlo experiment which demonstrate the risk behavior of the proposed estimators.

2. Three methods to modify estimators—quadratic loss case

In the main part of this paper, we are concerned with a more general situation than the one discussed in Section 1. Let S and T be independent statistics and suppose that the distribution of S/σ^2 is a central χ^2 distribution with ν degrees of freedom and that of T/σ^2 is a non-central χ^2 distribution with k degrees of freedom and non-centrality parameter λ. The situation discussed in Section 1 corresponds to $S = \sum_{i=1}^{n}(X_i - \bar{X})^2$, $\nu = n-1$, $T = n\bar{X}^2$, $k = 1$ and $\lambda = n\mu^2/(2\sigma^2)$. Also, one can reduce the variance estimation problem in the usual linear model to the situation above . An alternative approach in linear regression model can be found in Gelfand and Dey (1988). The estimator of σ^2 to be improved upon is

$$\delta_0 = S/(\nu+2).$$

2.1 *Method I*

The estimator given by Stein (1964) which improves upon δ_0 can be expressed as

$$\delta^* = \min\left(\frac{S}{\nu+2}, \frac{S+T}{\nu+k+2}\right) = \frac{S}{\nu+2} - \frac{k}{\nu+k+2}\left(\frac{S}{\nu+2} - \frac{T}{k}\right)^+, \tag{2.1}$$

where $a^+ = \max(0, a)$. Therefore, it shrinks δ_0 towards T/k when $S/(\nu{+}2) > T/k$. However, if the value of σ^2 is thought to be not less than 1, there is no reason to shrink δ_0 when $\delta_0 < 1$. So we may consider estimators which shrink δ_0 towards $\max(1, T/k)$ when $\delta_0 > \max(1, T/k)$. By introducing an absolutely continuous function $r(\cdot)$, we consider the following more general class of estimators:

$$\delta_1 = \begin{cases} \dfrac{S}{\nu+2} - \dfrac{k}{\nu+k+2}\left\{r\left(\dfrac{S}{S+T}\right)\right\}\left(\dfrac{S}{\nu+2} - 1\right), & \text{if } \dfrac{T}{k} < 1 < \dfrac{S}{\nu+2}, \\ \dfrac{S}{\nu+2} - \dfrac{k}{\nu+k+2}\left\{r\left(\dfrac{S}{S+T}\right)\right\}\left(\dfrac{S}{\nu+2} - \dfrac{T}{k}\right), & \text{if } 1 < \dfrac{T}{k} < \dfrac{S}{\nu+2}, \\ \dfrac{S}{\nu+2}, & \text{otherwise.} \end{cases}$$

Then, we can prove the following domination result.

THEOREM 2.1. *Under quadratic loss function* (1.1), δ_1 *dominates* δ_0 *if* $r(\cdot)$ *is an absolutely continuous function which satisfies* $0 < r(\cdot) \leq 2$.

The proof is given in the Appendix.

Remark. The condition $0 < r(S/(S+T))$ for all S and T is unnecessary and we only need that $r(S/(S+T)) > 0$ in some nonempty set in $A \cup B$, where A and B are the sets $T/k < 1 < S/(\nu+2)$ and $1 < T/k < S/(\nu+2)$ respectively.

2.2 *Method II*

First, we give the modification of Stein's estimator δ^* by this method. By introducing a function $h(\cdot)$, we consider the following estimator

$$\delta_2 = \frac{S}{\nu+2} - \frac{k}{\nu+k+2} h(S) \left(\frac{S}{\nu+2} - \frac{T}{k} \right)^+ .$$

We will discuss the choice of $h(\cdot)$ after we give a sufficient condition on $h(\cdot)$ for the estimator δ_2 to dominate δ_0.

THEOREM 2.2. *If $h(\cdot)$ is monotone non-decreasing, not identically 0 and satisfies $0 \le h(\cdot) \le 2$, then δ_2 dominates δ_0 under quadratic loss* (1.1).

We can prove this similarly to Theorem 2.1 if $h(\cdot)$ is differentiable. Even if $h(\cdot)$ is not differentiable, the proof goes through by using Riemann integration. We omit the details.

Remark. If $h(S)$ is replaced by $h(S+T)$ in the definition of δ_2, one obtains

$$\delta_2' = \frac{S}{\nu+2} - \frac{k}{\nu+k+2} h(S+T) \left(\frac{S}{\nu+2} - \frac{T}{k} \right)^+ .$$

For this δ_2' one can prove the same proposition as Theorem 2.2.

Now, we discuss how to choose $h(S)$ in order to get larger risk improvement for (μ, σ^2) in the neighborhood of $(\mu = 0, \sigma^2 = 1)$. Assuming that $h(\cdot)$ is differentiable and applying Lemma A.1, we can easily check that the risk difference $E(\delta_0 - \sigma^2)^2/\sigma^2 - E(\delta_2 - \sigma^2)^2/\sigma^2$ is the expected value of

$$\begin{aligned}
&\frac{2}{\nu+2}\frac{k}{\nu+k+2} h(S)\{2 - h(S)\}\frac{T}{k} I_+ + \frac{2kL}{\nu+k+2} h^2(S) \left(\frac{S}{\nu+2} - \frac{T}{k} \right)^+ \\
&\quad + \frac{2}{\nu+2}\frac{k}{\nu+k+2} S h'(S) \left(\frac{S}{\nu+2} - \frac{T}{k} \right)^+ \left\{ 2 - \frac{2k}{\nu+k+2} h(S) \right\},
\end{aligned}$$

where I_+ is the indicator function of the set $S/(\nu+2) > T/k$ and L is a Poisson variable as given in the proof of Theorem 2.1. The first term will be the main one. If $h(S)$ is nearly equal to 0 for $S \ll \nu + 2$, this will mean lack of improvement for $\sigma^2 \ll 1$. If $h(S)$ is nearly equal to 2 for $S \gg \nu + 2$, this will mean lack of improvement for $\sigma^2 \gg 1$. It seems natural that $h(\cdot)$ should increase from 0 to 2 in the neighborhood of $S = \nu + 2$. One drastic choice is

$$h(S) = \begin{cases} 0, & S \le \nu + 2, \\ 2, & S > \nu + 2. \end{cases}$$

A more moderate choice might be $h(S) = 2 \times F(S/(\nu+2))$, where $F(\cdot)$ is a distribution function of a continuous random variable whose mean is 1. More specifically, one can choose $F(S/(\nu+2))$ as if it were a prior distribution of σ^2. In the simulation study about which we will report in Section 4, we have chosen the distribution function of an inverse gamma variable with mean 1 as $F(\cdot)$.

If we apply methods I and II simultaneously to δ^* with $r(S/(S+T)) \equiv 1$, we get the estimator

$$\frac{S}{\nu+2} - \frac{k}{\nu+k+2} h(S) \left\{ \frac{S}{\nu+2} - \max\left(1, \frac{T}{k}\right) \right\}^+ . \tag{2.2}$$

Choosing the function $h(S)$ so that it increases from 0 to 2 in the neighborhood of $S/(\nu+2) = \sigma_0^2 (> 1)$ results in the estimator, we should use when we have prior information that $\sigma^2 \cong \sigma_0^2 > 1$. We can prove that the estimator (2.2) improves upon δ_0 if $h(\cdot)$ is non-decreasing, not identically 0 and $0 \le h(\cdot) \le 2$.

We can apply method II to other estimators. For example, for the modified Strawderman's estimator with $h = h(S+T)$, we have the following result.

THEOREM 2.3. *Let $h(\cdot)$ be a non-decreasing function which satisfies $0 \le h(\cdot) \le 1$. Then the estimator*

$$\frac{S}{\nu+2} - \frac{S}{\nu+2} h(S+T) \epsilon\left(\frac{S}{S+T}\right) \left(\frac{S}{S+T}\right)^{\delta}$$

is at least as good as δ_0 under quadratic loss if $\delta \ge 0$ and $0 \le \epsilon(\cdot) \le D(\delta)$, where

$$\begin{aligned} D(\delta) = \min\Bigg[\frac{2}{1+\delta}, 2\Bigg\{ & \beta\left(\frac{n+3}{2}+\delta, \frac{1}{2}\right)\beta\left(\frac{n+1}{2}, 1\right) \\ & - \beta\left(\frac{n+1}{2}+\delta, \frac{1}{2}\right)\beta\left(\frac{n+2}{2}, 1\right)\Bigg\} \\ & \div \left\{\beta\left(\frac{n+1}{2}, 1\right)\beta\left(\frac{n+3}{2}+2\delta, \frac{1}{2}\right)\right\}\Bigg]. \end{aligned}$$

The proof is similar to that of Strawderman (1974).

By modifying Brown's estimator one obtains

$$\frac{S}{\nu+2} - h(S+T)\left(\frac{1}{\nu+2} - c\right) S I_r, \tag{2.3}$$

where $r > 0$, $Z = \sqrt{T/S}$, I_r is the indicator function of the set $Z \le r$ and

$$c = E_0[S/\sigma^2 \mid Z \le r] / E_0[(S/\sigma^2)^2 \mid Z \le r],$$

where E_0 denotes the expected value when $\lambda = 0$. We can show that the estimator (2.3) dominates δ_0 if $h(\cdot)$ is non-decreasing, not identically 0 and satisfies $0 \le h(\cdot) \le 2$.

If we denote

$$\phi(z) = \frac{E_0[S/\sigma^2 \mid Z \le z]}{E_0[(S/\sigma^2)^2 \mid Z \le z]},$$

we can verify that a modification of the estimator given by Brewster and Zidek (1974)

$$\frac{S}{\nu+2} - h(S+T)\left(\frac{S}{\nu+2} - \phi(Z)S\right) \tag{2.4}$$

is at least as good as δ_0 if $h(\cdot)$ is non-decreasing and $0 \le h(\cdot) \le 1$.

Brewster-Zidek estimator is admissible and generalized Bayes and some of Strawderman's are also generalized Bayes. It is quite desirable but seems difficult to give similar results about some estimators derived here.

2.3 *Method III*

This method is the most drastic one and can be used to construct modified dominating estimators by truncating shrinkage, when the original dominating shrinkage estimator is less than 1. Suppose that the estimator $\delta(S,T)$ improves upon δ_0 under quadratic loss and that $\delta(S,T) \le \delta_0$ for all (S,T). Then, we consider the following estimator:

$$\delta_3 = \begin{cases} \dfrac{S}{\nu+2}, & \text{if } \dfrac{S}{\nu+2} < 1, \\ 1, & \text{if } \delta(S,T) < 1 \le \dfrac{S}{\nu+2}, \\ \delta(S,T), & \text{if } 1 \le \delta(S,T) \le \dfrac{S}{\nu+2}. \end{cases}$$

We can verify that δ_3 dominates δ_0 under the quadratic loss as follows: first, we note that $P\{1 < S/(\nu+2) \text{ and } \delta(S,T) < S/(\nu+2)\} > 0$ because the risk of $\delta(S,T)$ will be larger than that of δ_0 for $\sigma^2 \ge 1$ if $\delta(S,T) = S/(\nu+2)$ for all $S > \nu+2$. If $\sigma^2 \le 1$, the risk of δ_3 is less than that of δ_0 because δ_3 is not farther from σ^2 than $S/(\nu+2)$ for all (S,T) and $P\{S/(\nu+2) > 1 \text{ and } \delta(S,T) < S/(\nu+2)\} > 0$. If $\sigma^2 > 1$, the risk of δ_3 is not larger than that of $\delta(S,T)$ because δ_3 is not farther from σ^2 than $\delta(S,T)$ for all (S,T). Since $\delta(S,T)$ dominates δ_0, δ_3 dominates δ_0.

If we want to make the improvement of δ_3 large for $\sigma^2 = 1$, we may choose $\delta(S,T)$, the estimator which shrinks a lot.

3. Entropy loss case

It is shown here that one can apply three methods described in Section 2 to modify dominating estimators under entropy loss function (1.4). The best equivariant estimator is

$$\phi_0 = S/\nu,$$

and is the one to be improved upon. Although Brewster and Zidek (1974) have given a generalized Bayes improving estimator, in here, we only discuss modifications of the following improving one

$$\phi^* = \min\left(\frac{S}{\nu}, \frac{S+T}{\nu+k}\right) = \frac{S}{\nu} - \frac{k}{\nu+k}\left(\frac{S}{\nu} - \frac{T}{k}\right)^+, \tag{3.1}$$

which is of the same type as (2.1).

3.1 *Method I*

Modifying ϕ^* we consider estimators which shrink ϕ_0 towards $\max(1, T/k)$ when $\phi_0 > \max(1, T/k)$:

$$\phi_1 = \frac{S}{\nu} - \frac{k}{\nu+k} r\left(\frac{S}{S+T}\right)\left\{\frac{S}{\nu} - \max\left(1, \frac{T}{k}\right)\right\}^+ .$$

Then, we have the following result.

THEOREM 3.1. *Under entropy loss function* (1.4) ϕ_1 *dominates* ϕ_0 *if* $r(\cdot)$ *is an absolutely continuous function which satisfies* $0 < r(\cdot) \le a_{\nu,k}$, *where*

$$a_{\nu,k} = 2\Big/\left[1 + \frac{\nu}{\nu+k}\left\{\sqrt{\frac{1}{4} + \frac{1}{3}\frac{k}{\nu} + \frac{k^2}{\nu^2}} - \frac{1}{2}\right\}\right]. \tag{3.2}$$

The proof is given in the Appendix. The remark given at the end of Subsection 2.1 also applies to this case.

3.2 *Method II*

By introducing a function $h(\cdot)$ we modify ϕ^* as

$$\phi_2 = \frac{S}{\nu} - \frac{k}{\nu+k} h(S)\left(\frac{S}{\nu} - \frac{T}{k}\right)^+ .$$

Then, we have the following.

THEOREM 3.2. *If* $h(\cdot)$ *is monotone non-decreasing, not identically* 0 *and* $0 \le h(\cdot) \le a_{\nu,k}$, *then* ϕ_2 *dominates* ϕ_0 *under entropy loss* (1.4), *where* $a_{\nu,k}$ *is given by* (3.2).

We can prove this in the same way as Theorem 3.1. The remark similar to the one given after Theorem 2.2 also applies to this case.

If methods I and II are applied simultaneously to ϕ^*, we get the improving estimator

$$\frac{S}{\nu} - \frac{k}{\nu+k} h(S)\left(\frac{S}{\nu} - \max(T/k, 1)\right)^+ ,$$

where $h(\cdot)$ is non-increasing, not identically 0 and $0 \le h(\cdot) \le a_{\nu,k}$.

3.3 *Method III*

Let $\phi(S,T)$ be an estimator which improves upon ϕ_0 under the entropy loss and $\phi(S,T) \le \phi_0$ for all (S,T). We construct a modified dominating estimator by truncating the shrinkage when $\phi(S,T) < 1$:

$$\phi_3 = \begin{cases} S/\nu, & \text{if } S/\nu < 1, \\ 1, & \text{if } \phi(S,T) < 1 < S/\nu, \\ \phi(S,T), & \text{if } 1 \le \phi(S,T) \le S/\nu. \end{cases}$$

Then, we can show that ϕ_3 dominates ϕ_0 under the entropy loss as in the quadratic loss case.

Table 1. The relative risk improvements of δ_i and ϕ_i when $\nu = 1$ (in percent). (In the left parts of Tables 1–3 three figures in each cell give the relative risk improvements of δ_1, δ_2 and δ_3 from the top, and in the right parts those of ϕ_1, ϕ_2 and ϕ_3. The relative risk improvements of δ^* and ϕ^* when $\sigma^2 = 1$ is given at the bottom.)

	δ_i			ϕ_i		
σ^2 \ μ	0.0	0.5	1.0	0.0	0.5	1.0
	2.25	1.86	0.91	9.85	8.57	5.06
0.5	4.84	3.75	1.57	14.57	11.94	6.30
	3.49	2.81	1.25	12.17	10.28	5.72
	4.29	4.10	2.30	13.08	12.10	8.19
0.8	5.82	5.39	2.91	14.50	13.27	8.79
	5.60	5.21	2.82	14.83	13.51	8.88
	5.18	4.65	3.05	14.25	13.05	9.71
1.0	5.77	5.12	3.33	14.09	12.95	9.69
	6.14	5.42	3.48	15.43	14.01	10.24
	5.55	4.94	3.91	14.75	13.62	11.34
1.5	4.34	3.89	3.23	12.21	11.50	9.97
	5.34	4.74	3.78	14.64	13.51	11.27
	5.51	4.78	4.07	14.48	13.41	11.78
2.0	3.61	3.05	2.82	10.95	10.26	9.58
	4.63	3.96	3.48	13.57	12.57	11.19
δ^* or ϕ^*	1.44	1.42	1.18	6.50	6.44	5.55

4. A Monte Carlo experiment

In here, we give some Monte Carlo simulation results to demonstrate the performance of the proposed estimators. We only consider a random sample of size n from a normal distribution with mean μ. Therefore, $k = 1$, $\nu = n - 1$ and $\lambda = n\mu^2/(2\sigma^2)$. The following three estimators are considered for quadratic and entropy losses respectively.

Quadratic loss:

$\delta_1 = S/(\nu+2) - 2/(\nu+3)\{S/(\nu+2) - \max(1,T)\}^+$.

$\delta_2 = S/(\nu+2) - 2F(S/(\nu+2))/(\nu+3)\{S/(\nu+2) - T\}^+$, where $F(\cdot)$ is the distribution function of the Inverse Gamma distribution $IG(101, 1/100)$ (see Berger (1985)) whose mean and variance are 1 and 1/99 respectively.

$\delta_3 = S/(\nu+2) - \{S/(\nu+2) - \max(1,\delta^{**})\}^+$, where $\delta^{**} = S/(\nu+2) - 2/(\nu+3)(S/(\nu+2) - T)^+$.

Entropy loss:

$\phi_1 = S/\nu - a_{\nu,1}/(\nu+1)\{S/\nu - \max(1,T)\}^+$, where $a_{\nu,1}$ is defined by (3.1).

$\phi_2 = S/\nu - a_{\nu,1}F(S/\nu)/(\nu+1)(S/\nu - T)^+$, where $F(\cdot)$ is the same as in the case

Table 2. The relative risk improvements of δ_i and ϕ_i when $\nu = 3$ (in percent).

σ^2 \ μ	δ_i 0.0	0.5	1.0	ϕ_i 0.0	0.5	1.0
	2.09	1.41	0.26	5.87	4.03	0.89
0.5	6.29	3.67	0.50	14.23	8.31	1.41
	4.36	2.74	0.40	10.44	6.50	1.22
	5.93	4.17	1.44	12.14	9.05	3.18
0.8	10.17	6.85	2.04	18.33	12.98	4.05
	9.50	6.44	1.96	17.54	12.48	3.94
	7.24	5.54	2.33	14.22	11.03	4.77
1.0	9.55	7.07	2.80	17.37	13.14	5.40
	9.93	7.32	2.86	18.24	13.71	5.55
	8.46	7.25	4.10	16.03	13.46	7.76
1.5	6.80	6.03	3.62	13.50	11.60	7.02
	8.14	7.04	4.03	16.06	13.48	7.78
	7.85	6.89	4.53	15.44	13.37	8.97
2.0	4.35	4.12	3.13	10.20	9.11	6.84
	5.65	5.15	3.65	12.83	11.23	7.90
δ^* or ϕ^*	1.71	1.60	0.83	4.23	3.98	2.27

of δ_2.

$\phi_3 = S/\nu - \{S/\nu - \max(1, \phi^{**})\}^+$, where $\phi^{**} = S/\nu - a_{\nu,1}/(\nu+1)(S/\nu - T)^+$.
All the estimators are designed so that they give most improvement when the value of σ^2 is close to 1. We also note that in all the estimators, we have taken constants so that they shrink as much as possible.

We take $\nu = 1, 3, 5$, $\mu = 0.0, 0.5, 1.0$ and $\sigma^2 = 0.5, 0.8, 1.0, 1.5$, 2.0. Based on 100,000 iterations, the risks of the estimators were estimated by averaged losses. Using risk estimators, we estimated the relative risk improvements:

$$[E\{L_1(\delta_0,\sigma^2)\} - E\{L_1(\delta_i,\sigma^2)\}]/E\{L_1(\delta_0,\sigma^2)\}, \qquad i = 1, 2, 3,$$

and

$$[E\{L_2(\phi_0,\sigma^2)\} - E\{L_2(\phi_i,\sigma^2)\}]/E\{L_2(\phi_0,\sigma^2)\}, \qquad i = 1, 2, 3.$$

These are given in Tables 1–3. (We verified some entries in Tables 1–3 by using numerical integration. The differences are at most 0.15% in all cases.) For comparison, the relative risk improvements of δ^* or ϕ^* (given by (2.1) or (3.1)) are also given.

We may summarize the results as follows:

(i) For both loss functions, the relative risk improvements of the proposed estimators are much larger than those of the equivariant improving estimator δ^*

Table 3. The relative risk improvements of δ_i and ϕ_i when $\nu = 5$ (in percent).

σ^2 \ μ	δ_i 0.0	δ_i 0.5	δ_i 1.0	ϕ_i 0.0	ϕ_i 0.5	ϕ_i 1.0
	1.23	0.68	0.08	3.00	1.60	0.13
0.5	5.07	2.10	0.14	10.23	4.16	0.25
	3.26	1.52	0.12	7.07	3.17	0.22
	5.41	3.09	0.58	9.21	5.50	1.01
0.8	10.93	5.81	0.87	16.99	9.29	1.41
	10.13	5.43	0.84	15.91	8.79	1.37
	7.36	4.85	1.14	12.00	7.86	1.93
1.0	10.92	6.82	1.44	16.91	10.53	2.34
	11.25	6.99	1.47	17.46	10.81	2.38
	9.23	6.45	2.41	14.35	10.43	4.06
1.5	7.51	5.32	2.11	12.24	9.03	3.74
	8.84	6.15	2.32	14.31	10.37	4.07
	8.13	6.36	3.25	13.46	10.61	5.46
2.0	4.07	3.50	2.23	7.91	6.59	4.10
	5.17	4.28	2.50	9.90	8.04	4.58
δ^* or ϕ^*	1.66	1.45	0.48	3.01	2.63	0.90

or ϕ^* in the region of (μ, σ^2), which is of interest. In many cases, they are more than three times larger than those of δ^* or ϕ^*. Even if we compare the relative risk improvements, with those of Brewster-Zidek estimator and locally optimal shrinkage estimator, given by Rukhin (1987) for quadratic loss function, we can repeat the same statement (see also Rukhin and Ananda (1992)). We should not neglect such significant improvements. This especially applies to the entropy loss case.

(ii) The estimators δ_1 and ϕ_1 give larger improvements for some $\sigma^2 > 1$ rather than $\sigma^2 = 1$. On the contrary, δ_2 and ϕ_2 give larger improvements for some $\sigma^2 < 1$ rather than $\sigma^2 = 1$. There is no big difference between the maximum improvements of the three estimators. In any case, one should examine the behavior of the estimators more closely to design the estimators.

(iii) If the value of ν gets larger, the range of μ gets smaller in which the proposed estimators give significant improvements. This is due to the fact that the non-centrality parameter $\lambda = n\mu^2/(2\sigma^2)$ gets larger if $\nu = n - 1$ increases.

Acknowledgements

The author would like to thank the referees for their helpful comments and for carefully reading the manuscript.

Appendix

Lemmas and proofs of theorems

The following well-known lemma is useful to prove dominance (see Efron and Morris (1976)).

LEMMA A.1. *Let V/σ^2 be distributed as χ^2 with f degrees of freedom and let $h(\cdot)$ be an absolutely continuous function. If both expectations exist,*

$$E\left\{\frac{V}{\sigma^2}h(V)\right\} = E\{fh(V) + 2Vh'(V)\}.$$

PROOF OF THEOREM 2.1. Let A and B be the sets $T/k < 1 < S/(\nu+2)$ and $1 < T/k < S/(\nu+2)$ respectively and let I_C denote the indicator function of a set C. We show that δ_1 dominates δ_0 for the loss function $(\hat{\sigma}^2 - \sigma^2)^2/\sigma^2$. We calculate the risk difference between δ_0 and δ_1 as in Stein (1964), by introducing an auxiliary variable L distributed as a Poisson variable with parameter λ such that L is independent of S and that T given S and L is central χ^2 with $k+2L$ degrees of freedom. Putting $f(S,T) = \{S/(\nu+2) - \max(1, T/k)\}^+$, we have

$$\begin{aligned}
&(\delta_0 - \sigma^2)^2/\sigma^2 - (\delta_1 - \sigma^2)^2/\sigma^2 \\
&= \frac{2k}{\nu+k+2}\frac{1}{\nu+2}\frac{S}{\sigma^2}r\left(\frac{S}{S+T}\right)f(S,T) - \frac{2k}{\nu+k+2}r\left(\frac{S}{S+T}\right)f(S,T) \\
&\quad - \frac{k^2}{(\nu+k+2)^2}\frac{1}{\sigma^2}r^2\left(\frac{S}{S+T}\right)f^2(S,T) \\
&\equiv D_1 - D_2 - D_3,
\end{aligned}$$

say. To evaluate $E(D_1)$ we apply Lemma A.1 with $V = S$. Then

$$\begin{aligned}
E(D_1 - D_2) = E\Bigg[\frac{4k}{(\nu+k+2)(\nu+2)}\bigg\{&r\left(\frac{S}{S+T}\right)\max\left(1, \frac{T}{k}\right)I_{A\cup B} \\
&+ \frac{ST}{(S+T)^2}r'\left(\frac{S}{S+T}\right)f(S,T)\bigg\}\Bigg].
\end{aligned}$$

Since $r(\cdot) \le 2$ and $\max(1, T/k) \ge T/k$,

$$\begin{aligned}
\text{(A.1)}\qquad -E(D_3) \ge E\bigg\{&-\frac{2k^2}{(\nu+k+2)^2}\frac{1}{\nu+2}\frac{S}{\sigma^2}r\left(\frac{S}{S+T}\right)f(S,T) \\
&+\frac{2k}{(\nu+k+2)^2}\frac{T}{\sigma^2}r\left(\frac{S}{S+T}\right)f(S,T)\bigg\}.
\end{aligned}$$

We apply Lemma A.1 to the first term of the right-hand side of (A.1) with $V = S$ and also to the second term with $V = T$ since T is central χ^2 with $k + 2L$ degrees of freedom given S and L. Then

$$\begin{aligned}
-E(D_3) \geq & -\frac{2k^2}{(\nu+k+2)^2} E\Bigg\{ r\left(\frac{S}{S+T}\right) f(S,T) \\
& \qquad + \frac{2}{\nu+2} r\left(\frac{S}{S+T}\right) \max\left(1, \frac{T}{k}\right) I_{A\cup B} \\
& \qquad + \frac{2}{\nu+2} \frac{ST}{(S+T)^2} r'\left(\frac{S}{S+T}\right) f(S,T) \Bigg\} \\
& + \frac{2k}{(\nu+k+2)^2} E\Bigg\{ (k+2L) r\left(\frac{S}{S+T}\right) f(S,T) \\
& \qquad - 2r\left(\frac{S}{S+T}\right) \frac{T}{k} I_B \\
& \qquad - 2\frac{ST}{(S+T)^2} r'\left(\frac{S}{S+T}\right) f(S,T) \Bigg\} \\
\geq & -E(D_1 - D_2)
\end{aligned}$$

with strict inequality for some parameter values. This completes the proof.

The following lemma is similar to the one used in simultaneous estimation problem of gamma parameters (see Dey and Srinivasan (1985)).

LEMMA A.2. *If* $0 \leq v < 1$, *then*

$$\log(1-v) \geq -v - v^2/6 - v^2/\{3(1-v)\}.$$

PROOF OF THEOREM 3.1. We again calculate the risk difference by introducing an auxiliary Poisson variable L such that T given L is central χ^2 with $k+2L$ degrees of freedom. Since ϕ_1 can be expressed as

$$\phi_1 = \frac{S}{\nu}\left[1 - r\left(\frac{S}{S+T}\right)\{1 - f(S,T)\}^+\right],$$

with

$$f(S,T) = \left[1 - \frac{k}{\nu+k}\left\{\frac{\nu}{k} + \frac{\nu}{S}\max\left(1, \frac{T}{k}\right)\right\}\right]^+,$$

we have

$$\begin{aligned}
&\{\phi_0/\sigma^2 - \log(\phi_0/\sigma^2) - 1\} - \{\phi_1/\sigma^2 - \log(\phi_1/\sigma^2) - 1\} \\
&\quad = \frac{S+T}{\sigma^2}\frac{1}{\nu}\frac{S}{S+T} r\left(\frac{S}{S+T}\right) f(S,T) + \log\left\{1 - r\left(\frac{S}{S+T}\right) f(S,T)\right\} \\
&\quad \equiv D_1 + D_2,
\end{aligned}$$

say. To evaluate the expected value of D_1, we apply Lemma A.1 with $V = S$ and also with $V = T$. Then

$$E(D_1) \geq \frac{\nu+k}{\nu} E\left\{ \frac{S}{S+T} r\left(\frac{S}{S+T}\right) f(S,T) \right\}.$$

It follows from Lemma A.2 that

$$\begin{aligned} D_2 \geq & -r\left(\frac{S}{S+T}\right) f(S,T) \\ & -r^2\left(\frac{S}{S+T}\right) f^2(S,T) \left\{ \frac{1}{6} + \frac{1}{3} \frac{1}{1 - r(S/(S+T)) f(S,T)} \right\}. \end{aligned}$$

Therefore, $E(D_1 + D_2)$ is not smaller than the expected value of

$$\begin{aligned} r\left(\frac{S}{S+T}\right) f^2(S,T) \Bigg[& \frac{\nu+k}{\nu} \frac{S}{S+T} \\ & - r\left(\frac{S}{S+T}\right) \left\{ \frac{1}{6} + \frac{1}{3} \frac{1}{1 - r(S/(S+T)) f(S,T)} \right\} \Bigg]. \end{aligned}$$

Since $r(S/(S+T)) \leq a_{\nu,k}$ and $f(S,T) \leq 1 - \nu/(\nu+k) \cdot (S+T)/S$, one needs only to show that

$$z - \frac{1}{6} a_{\nu,k} - \frac{1}{3} \frac{a_{\nu,k} z}{z - a_{\nu,k}(z-1)} \geq 0, \quad \text{for } 1 < z < \frac{\nu+k}{\nu}.$$

We can easily prove the above inequality and this completes the proof.

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer, New York.

Brewster, J. F. and Zidek, J. V. (1974). Improving on equivariant estimators, *Ann. Statist.*, **2**, 21–38.

Brown, L. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters, *Ann. Math. Statist.*, **39**, 29–48.

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss, *Ann. Statist.*, **13**, 1581–1591.

Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distributions, *Ann. Statist.*, **4**, 11–21.

Gelfand, A. E. and Dey, D. K. (1988). Improved estimation of the disturbance variance in a linear regression model, *J. Econometrics*, **39**, 387–395.

Kubokawa, T. (1994). A unified approach to improving equivariant estimators, *Ann. Statist.*, **22**, 290–299.

Maatta, J. M. and Casella, G. (1990). Developments in decision-theoretic variance estimation, *Statist. Sci.*, **5**, 90–120.

Madi, M. T. (1993). Alternative estimators for the variance of several normal populations, *Statist. Probab. Lett.*, **17**, 321–328.

Proskin, H. M. (1985). An admissibility theorem with applications to the estimation of the variance of the normal distribution, Ph. D. dissertation, Department of Statistics, Rutgers University, New Jersey.

Rukhin, A. L. (1987). How much better are better estimators of a normal variance?, *J. Amer. Statist. Assoc.*, **77**, 159–162.

Rukhin, A. L. and Ananda, M. M. A. (1992). Risk behavior of variance estimators in multivariate normal distribution, *Statist. Probab. Lett.*, **13**, 159–166.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, *Ann. Inst. Statist. Math.*, **16**, 155–160.

Strawderman, W. E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss, *Ann. Statist.*, **2**, 190–198.