ON THE ESTIMATION OF A NORMAL VARIANCE*

Sándor Csörgő and **Julian J. Faraway**

Received:

Abstract. Complementing Rukhin's [5] investigations, the mean squared error of members of a family of preliminary-test estimators and members of the basic family of shrinkage estimators are compared. The reference point is Stein's [6] estimator that is contained in both classes. It turns out that Stein's estimator has certain qualitative optimality properties among the preliminary-test estimators. But it also turns out that, while Rukhin's 4% maxim in the minimax shrinkage class is well known and frequently cited, without minimaxity greater improvements can be achieved in his local sense by estimators coming from the preliminary-test class. This underlines another aspect of Stein's original insight. Most of our findings are presented graphically.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be independent observations from a normal distribution with unknown mean $\mu \in I\!\!R$ and unknown variance $\sigma^2 > 0$, $n = 2, 3, \ldots$. Let $\overline{X}_n := \sum_{k=1}^n X_k / n$ be the sample mean, and, with any $\mu_0 \in I\!\!R$, put

$$(1.1) \qquad S_n^2 := \sum_{k=1}^n \left(X_k - \overline{X}_n\right)^2, \ \ S_n^2(\mu_0) := \sum_{k=1}^n \left(X_k - \mu_0\right)^2, \ \ \widetilde{\sigma}_n^2 := \frac{S_n^2}{n+1}, \ \ s_n^2 := \frac{S_n^2}{n-1},$$

so that $S_n^2 = S_n^2(\overline{X}_n)$. The equivariant estimator $\widetilde{\sigma}_n^2$ of σ^2 has desirable properties (cf. [3] and [4]). However, Stein [6] proved that it is inadmissible with respect to squared error loss. Specifically, he has shown that for the scaled squared error risk of the minimax estimator

$$(1.2) \qquad \sigma^2_{\mathrm{Stein},n} := \min\left\{\frac{S_n^2(0)}{n+2}, \frac{S_n^2}{n+1}\right\} = \min\left\{\frac{S_n^2(0)}{n+2}, \widetilde{\sigma}_n^2\right\}$$

one has, for every $n \geq 2$,

$$(1.3) \qquad E\left(\left[\frac{\sigma^2_{\mathrm{Stein},n}}{\sigma^2} - 1\right]^2\right) < E\left(\left[\frac{\widetilde{\sigma}_n^2}{\sigma^2} - 1\right]^2\right) = \frac{2}{n+1},$$

* Work partially supported by NSF Grant DMS–9208067.
AMS Subject Classification: 62F10, 62C15.
Key words and phrases: preliminary-test and shrinkage estimators, quadratic risk.

no matter what the true values of μ and σ^2 are, where the value of the scaled squared error of $\widetilde{\sigma}_n^2$ on the right side is well known. Stein's proof is also recorded in [7], pp. 396–397, and is accoladed and analyzed in [4].

Using a version of Stein's idea, Brown [2] has extended Stein's inadmissibility result for the problem of estimation of powers of scale parameters in general location-scale models, where the location parameter is also unknown, under a very general class of loss functions. When specialized to the normal case above, and understanding that $I(A)$ is the indicator of the event A, Brown's estimator is

$$\overline{\sigma}^2_{\mathrm{Brown},n} := \overline{\sigma}^2_{\mathrm{Brown},n}(t) = I\big(\sqrt{n}\,|\overline{X}_n| < ts_n\big)\theta_n s_n^2 + I\big(\sqrt{n}\,|\overline{X}_n| \ge ts_n\big)\widetilde{\sigma}_n^2\,, \tag{1.4}$$

where, under the scaled squared error loss, $\theta_n = \theta_n(t) \in (0,1)$ for $t > 0$ is given by

$$\theta_n := \frac{\int_0^\infty \big(2\Phi(tx)-1\big)x^n\, e^{-\frac{n-1}{2}x^2}\,dx}{\int_0^\infty \big(2\Phi(tx)-1\big)x^{n+2}\, e^{-\frac{n-1}{2}x^2}\,dx}\,,$$

where, and in what follows, $\Phi(\cdot)$ and $\varphi(\cdot) = \Phi'(\cdot)$ denote the standard normal distribution and density functions. Brown [2] has presented a limited calculation of the risk of $\overline{\sigma}^2_{\mathrm{Brown},n}(t)$ under the scaled squared error and another loss function. Under the scaled squared error loss, for appropriate choices of $t > 0$, he found 1.6% and 1.2% "maximal improvements" over $\widetilde{\sigma}_n^2$ for sample sizes $n = 2$ and $n = 10$, respectively.

Stein's estimator is a member of the family of scale-equivariant shrinkage estimators $\overline{\sigma}_n^2$, given in Rukhin's [5] notation by

$$\overline{\sigma}_n^2 := \overline{\sigma}_n^2(\phi_n) := \left[1 - \phi_n\left(\frac{\sqrt{n}\,|\overline{X}_n|}{\sqrt{n\overline{X}_n^2 + S_n^2}}\right)\right]\widetilde{\sigma}_n^2 = \left[1 - \phi_n\left(\frac{\sqrt{n}\,|\overline{X}_n|}{\sqrt{\sum_{k=1}^n X_k^2}}\right)\right]\frac{S_n^2}{n+1}\,, \tag{1.5}$$

where $\phi_n : [0,1) \mapsto [0,1]$ is a Borel measurable function. It belongs to the choice

$$\phi_n(x) = \phi_{\mathrm{Stein},n}(x) := \max\left\{0, 1 - \frac{n+1}{n+2}\,\frac{1}{1-x^2}\right\}, \quad 0 \le x < 1\,.$$

As easy considerations show, Brown's estimator is also a member of this shrinkage family.

Brewster and Zidek [1] have constructed a minimax member $\overline{\sigma}^2_{\mathrm{B-Z},n}$ of the shrinkage family and proved that their estimator is admissible within the class of all scale-equivariant estimators. Later Proskin proved that $\overline{\sigma}^2_{\mathrm{B-Z},n}$ is admissible in general. (See [4] for the exact reference.) The function $\phi_{\mathrm{B-Z},n}(\cdot)$ that produces the admissible minimax Brewster–Zidek estimator has been explicitly described by Rukhin [5].

Rukhin [5] has derived a general integral formula that can be used in principle to evaluate the scaled squared error loss of any estimator $\overline{\sigma}_n^2(\phi_n)$ in the shrinkage family as a function of $\sqrt{n}\,|\mu|/\sigma$. In this way he was able to calculate the improvement of the Brewster–Zidek estimator $\overline{\sigma}^2_{\mathrm{B-Z},n}$ over $\widetilde{\sigma}_n^2$ numerically. Furthermore, for any value of $\eta := \eta_n := \sqrt{n}\,|\mu|/\sigma$, he determined an explicit function $\phi_{\mathrm{Rukhin},n,\eta}(\cdot)$ that gives the locally optimal shrinkage estimators $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\eta)$ that provide the greatest possible improvement at η within the whole shrinkage class. Rukhin's [5] frequently cited result is that for any values of the parameters, the greatest relative improvement over $\widetilde{\sigma}_n^2$ that is possible in the whole subclass of shrinkage estimators that are minimax, is less than 4%. We have $\overline{\sigma}^2_{\mathrm{Rukhin},n}(0) = \sigma^2_{\mathrm{Stein},n}$, and Rukhin determined a critical value $\overline{\eta}_n > 0$ such that $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\eta)$ is minimax for $\eta \le \overline{\eta}_n$ and not minimax for $\eta > \overline{\eta}_n$. In the latter case, the local improvements may be greater than 4%; the precise statements are in Section 3 below.

Let us introduce the general families of preliminary-test estimators $\widehat{\sigma}^2_n(\mu_0) = \widehat{\sigma}^2_n(\mu_0; t, m_0, m)$, indexed by μ_0, given as

$$(1.6) \qquad \widehat{\sigma}^2_n(\mu_0) = I\left(\left|\frac{\sqrt{n}\,(\overline{X}_n - \mu_0)}{s_n}\right| < t\right)\frac{S^2_n(\mu_0)}{m_0} + I\left(\left|\frac{\sqrt{n}\,(\overline{X}_n - \mu_0)}{s_n}\right| \ge t\right)\frac{S^2_n}{m},$$

where $\mu_0 \in I\!R$, $t \ge 0$, $m_0 > 0$, $m > 0$, and s_n, S^2_n and $S^2_n(\mu_0)$ are as in (1.1). No members of the class $\widehat{\sigma}^2_n(\mu_0)$ are scale-equivariant if $\mu_0 \ne 0$. However, all members of the class $\widehat{\sigma}^2_n(0)$ are scale-equivariant. Moreover, $\sigma^2_{\mathrm{Stein},n}$ can be identified as

$$(1.7) \qquad \sigma^2_{\mathrm{Stein},n} = \widehat{\sigma}^2_n\big(0; t_{\mathrm{Stein},n}, n+2, n+1\big) \quad \text{where} \quad t_{\mathrm{Stein},n} = \sqrt{(n-1)/(n+1)}.$$

This is indicated by Stein [6] himself, and can be found explicitly in [3], pp. 274–275. Thus $\sigma^2_{\mathrm{Stein},n}$ results from performing a preliminary test, using Student's test with critical value $t = t_{\mathrm{Stein},n}$, for the null hypothesis that $\mu = 0$, and using $\widetilde{\sigma}^2_n$ or $S^2_n(0)/(n+2)$, depending upon whether this hypothesis is rejected or accepted. In case of the latter, $S^2_n(0)/(n+2)$ is the admissible estimator of σ^2 under the squared error loss when $\mu = 0$. (Although Brown's estimator is also based on a preliminary-test of the hypothesis $\mu = 0$, upon acceptance it does not use the value 0. Hence $\overline{\sigma}^2_{\mathrm{Brown},n}$ is not in the above class $\widehat{\sigma}^2_n(0)$ of scale-equivariant preliminary-test estimators.) Of course, no member of the class $\widehat{\sigma}^2_n(\mu_0; t, m_0, m)$ can be admissible, whether $\mu = 0$ or not. The formal choice $t = \infty$ results in the estimator $\widehat{\sigma}^2_n(\mu_0; \infty, m_0) = S^2_n(\mu_0)/m_0$.

In the recent overview of Maatta and Casella [4], the focus of attention is the shrinkage class $\overline{\sigma}^2_n(\phi_n)$ as far as point estimation goes. Most of the discussants also concentrate on this class. This is, of course, natural in view of the main developments following Stein's [6] paper. Readers interested in interval estimation, conditional properties, directions of multivariate extensions and other aspects of the problem (in all of which the above results in univariate point estimation serve as basic motivation and guideline) are referred to this excellent review that, taken together with the discussions, explores the whole literature to date.

The aim of the present paper is to calculate the scaled mean squared error of $\widehat{\sigma}^2_n(\mu_0) = \widehat{\sigma}^2_n(\mu_0; t, m_0, m)$ in (1.6) for all values of the parameters. A formula for this is derived in the next section. Since the formula obtained is very complicated, its analysis has become possible only by the use of today's computers. This is done in the third section, where we try to see what are the choices of the parameters that give the best improvements relative to $\widetilde{\sigma}^2_n$ and each other, and in relationship with Rukhin's locally optimal shrinkage estimators. Thus our investigations parallel and complement Rukhin's [5] original numerical investigations. It turns out that Stein's estimator has certain qualitative optimality properties among the preliminary-test estimators. However, it also turns out that the preliminary-test aspect of Stein's basic insight also deserves some attention: in Rukhin's local sense, far greater improvements are possible within this class if minimaxity is not required. Since Rukhin's estimator is minimax only for the limited interval $[0, \overline{\eta}_n]$ of η values, this is of some interest: $\overline{\eta}_3 = 1.38, \overline{\eta}_4 = 1.34, \ldots$; $\overline{\eta}_n$ decreases as n grows.

Stein [6] himself expressed skepticism concerning the use of quadratic loss when estimating a variance, a point dealt with by Brown [2] extensively in the general context of estimating a scale parameter and again by him and others in the discussion of [4]. But as [4] itself and many of its references show, there is still an interest in quadratic loss. This must be partly due to its general theoretical importance and its mathematical tractability. Whether the local improvements indicated in Section 3 can motivate ideas in multivariate and other settings remains to be seen. Most of our findings are presented in a graphical form.

2. The mean squared error of preliminary-test estimators

For $u, v \in \mathbb{R}$ and $k = 0, 1, 2, \ldots$, consider the functions

$$L_k(u,v) := \int_0^\infty \Phi(ux+v)x^k\, e^{-x^2/2}\, dx \quad \text{and} \quad M_k(u,v) := \int_0^\infty \varphi(ux+v)x^k\, e^{-x^2/2}\, dx\,.$$

The following theorem determines the scaled squared error risk

$$R_n(\Delta, t, m_0, m) := E\left(\left[\frac{\widehat{\sigma}_n^2(\mu_0)}{\sigma^2} - 1\right]^2\right) = E\left(\frac{\widehat{\sigma}_n^4(\mu_0)}{\sigma^4}\right) - 2E\left(\frac{\widehat{\sigma}_n^2(\mu_0)}{\sigma^2}\right) + 1 \tag{2.1}$$

in terms of the scaled first moment $E(\widehat{\sigma}_n^2/\sigma^2)$ and the scaled second moment $E(\widehat{\sigma}_n^4/\sigma^4)$, by evaluating the functions $L_k(u,v)$ and $M_k(u,v)$ at $u = c$ and $v = \pm d$, and using a constant ξ_n, where

$$\Delta := \frac{\mu - \mu_0}{\sigma}\,, \quad c := \frac{t}{\sqrt{n-1}}\,, \quad d := \sqrt{n}\,\Delta \quad \text{and} \quad \xi_n := \frac{1}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}\,, \tag{2.2}$$

and where $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x}\, dx$, $\alpha > 0$, is the usual gamma function.

THEOREM. *For $n = 2, 3, \ldots$ and $0 < t < \infty$ we have*

$$\begin{aligned} E\left(\frac{\widehat{\sigma}_n^2(\mu_0)}{\sigma^2}\right) = \frac{2\xi_n}{m_0}\Big[& L_n(c,d) + L_n(c,-d) + \left(1+d^2\right)\left\{L_{n-2}(c,d) + L_{n-2}(c,-d)\right\} \\ & - c\left\{M_{n-1}(c,d) + M_{n-1}(c,-d)\right\} + d\left\{M_{n-2}(c,d) - M_{n-2}(c,-d)\right\}\Big] \\ & - \frac{2\xi_n}{m}\Big[L_n(c,d) + L_n(c,-d)\Big] + 2\,\frac{n-1}{m} - \frac{n}{m_0} - \frac{d^2}{m_0} \end{aligned}$$

and

$$\begin{aligned} E\left(\frac{\widehat{\sigma}_n^4(\mu_0)}{\sigma^4}\right) = \frac{2\xi_n}{m_0^2}\Big[& L_{n+2}(c,d) + L_{n+2}(c,-d) + 2\left(1+d^2\right)\left\{L_n(c,d) + L_n(c,-d)\right\} \\ & + \left(3 + 6d^2 + d^4\right)\left\{L_{n-2}(c,d) + L_{n-2}(c,-d)\right\} \\ & - \left(c^3 + 2c\right)\left\{M_{n+1}(c,d) + M_{n+1}(c,-d)\right\} \\ & + d\left(c^2 + 2\right)\left\{M_n(c,d) - M_n(c,-d)\right\} \\ & - c\left(d^2 + 3\right)\left\{M_{n-1}(c,d) + M_{n-1}(c,-d)\right\} \\ & + \left(d^3 + 5d\right)\left\{M_{n-2}(c,d) - M_{n-2}(c,-d)\right\}\Big] \\ & - \frac{2\xi_n}{m^2}\Big[L_{n+2}(c,d) + L_{n+2}(c,-d)\Big] + 2\,\frac{n^2-1}{m^2} - \frac{n^2 + 2n + 2nd^2 + 4d^2 + d^4}{m_0^2}\,. \end{aligned}$$

Furthermore, for every $v \in \mathbb{R}$ and $k = 0, 1, 2, \ldots$,

$$\begin{aligned} L_{2k}(c,v) &= \frac{(2k)!}{2^k k!}\, L_0(c,v) + c \sum_{j=0}^{k-1} \frac{(2k)!\,(k-j)!}{2^j (2k-2j)!\,k!}\, M_{2k-2j-1}(c,v)\,, \\ L_{2k+1}(c,v) &= 2^k k!\,\Phi(v) + c \sum_{j=0}^{k} \frac{2^j k!}{(k-j)!}\, M_{2k-2j}(c,v) \end{aligned} \tag{2.3}$$

and, with $G^{(k)}(t)$ *denoting the* k^{th} *derivative of* $G(t) = G^{(0)}(t) = e^{t^2/2}\,\Phi(t)$, $t \in \mathbb{R}$,

$$M_k(c, \pm d) = \frac{e^{-d^2/2}}{(1+c^2)^{(k+1)/2}}\, G^{(k)}\Big(\mp d \frac{c}{\sqrt{1+c^2}}\Big)\,. \tag{2.4}$$

The formulae in (2.3) and (2.4) imply that, besides the values of Φ, the only integral in the resulting expression for $R_n(\Delta, t, m_0, m)$ in (2.1) that has to be computed by numerical integration is $L_0(c, \pm d) = \int_0^\infty \Phi(cx \pm d) e^{-x^2/2}\, dx = \sqrt{2\pi} \int_0^\infty \Phi(cx \pm d)\varphi(x)\, dx$. It is absent when n is odd. Using Leibniz's binomial rule for successive differentiation of a product, it is possible to give a (complicated) closed formula for the derivatives in (2.4) in terms of rational and exponential functions and values of Φ.

PROOF. Using the notation in (1.1) and (2.2) and setting $Z_n = \sqrt{n}\,(\overline{X}_n - \mu)/\sigma$ and $\chi^2_{n-1} = S_n^2/\sigma^2$, simple algebra gives

$$\begin{aligned}\frac{\widehat{\sigma}_n^2(\mu_0)}{\sigma^2} &= I\Big(-c\sqrt{\chi^2_{n-1}} - d < Z_n < c\sqrt{\chi^2_{n-1}} - d\Big) \frac{1}{m_0}\Big[\chi^2_{n-1} + Z_n^2 + 2dZ_n + d^2\Big] \\ &\quad + I\Big(Z_n \notin \Big(-c\sqrt{\chi^2_{n-1}} - d\,,\, c\sqrt{\chi^2_{n-1}} - d\Big)\Big) \frac{1}{m}\,\chi^2_{n-1}\end{aligned}$$

and, denoting the two indicator values here by $I\big(Z_n \in \ \big)$ and $I\big(Z_n \notin \ \big)$,

$$\begin{aligned}\frac{\widehat{\sigma}_n^4(\mu_0)}{\sigma^4} &= I\big(Z_n \in \ \big)\, \frac{1}{m_0^2}\Big[\chi^4_{n-1} + Z_n^4 + 2\chi^2_{n-1} Z_n^2 + 4d\big\{\chi^2_{n-1} Z_n + Z_n^3\big\} + 2d^2\big\{\chi^2_{n-1} + Z_n^2\big\} \\ &\qquad + 4d^2 Z_n^2 + 4d^3 Z_n + d^4\Big] + I\big(Z_n \notin \ \big)\, \frac{1}{m^2}\,\chi^4_{n-1}\,.\end{aligned}$$

Since Z_n and χ^2_{n-1} are independent, with Z_n standard normal for each n and χ^2_{n-1} having the χ^2-distribution with $n-1$ degrees of freedom, conditioning on χ^2_{n-1} we obtain

$$\begin{aligned}E\Big(\frac{\widehat{\sigma}_n^2(\mu_0)}{\sigma^2}\Big) &= \frac{\xi_n}{m_0}\int_0^\infty \Big\{\int_{-c\sqrt{y}-d}^{c\sqrt{y}-d} \Big[y + d^2 + 2dx + x^2\Big]\varphi(x)\, dx\Big\} y^{\frac{n-3}{2}}\, e^{-\frac{y}{2}}\, dy \\ &\quad + \frac{\xi_n}{m}\int_0^\infty \Big[\Phi\big(-c\sqrt{y} - d\big) + 1 - \Phi\big(c\sqrt{y} - d\big)\Big]\, y\, y^{\frac{n-3}{2}}\, e^{-\frac{y}{2}}\, dy\,,\end{aligned}$$

since for $y > 0$, the density function of χ^2_{n-1} is $h_{n-1}(y) = \xi_n y^{(n-3)/2}\, e^{-y/2}$, and

$$\begin{aligned}E\Big(\frac{\widehat{\sigma}_n^4(\mu_0)}{\sigma^4}\Big) &= \frac{\xi_n}{m_0^2}\int_0^\infty \Big\{\int_{-c\sqrt{y}-d}^{c\sqrt{y}-d} \Big[y^2 + 2d^2 y + d^4 + 4dyx + 4d^3 x + 2yx^2 + 6d^2 x^2 \\ &\qquad + 4dx^3 + x^4\Big]\varphi(x)\, dx\Big\} y^{\frac{n-3}{2}}\, e^{-\frac{y}{2}}\, dy \\ &\quad + \frac{\xi_n}{m^2}\int_0^\infty \Big[\Phi\big(-c\sqrt{y} - d\big) + 1 - \Phi\big(c\sqrt{y} - d\big)\Big]\, y^2\, y^{\frac{n-3}{2}}\, e^{-\frac{y}{2}}\, dy\,.\end{aligned}$$

In both formulae, all the inner integrals can be integrated out by using as many integrations by parts as dictated by the exponent of x. We also use the obvious symmetry properties of Φ and φ, and the facts that $\int_0^\infty y h_{n-1}(y)\, dy = n-1$ and $\int_0^\infty y^2 h_{n-1}(y)\, dy = n^2 - 1$. The latter yield some trivial terms that are integrated out completely. Disregarding the coefficients of the non-trivial terms, two types emerge. One is

$$\int_0^\infty \Phi\big(c\sqrt{y} \pm d\big) y^l \, y^{\frac{n-3}{2}} \, e^{-\frac{y}{2}} \, dy = 2 \int_0^\infty \Phi(cx \pm d) x^{n-2+2l} \, e^{-\frac{x^2}{2}} \, dx = 2L_{n-2+2l}(c, \pm d)$$

for $l = 0, 1, 2$, and the other is

$$\int_0^\infty \varphi\big(c\sqrt{y} \pm d\big) y^{\frac{r}{2}} \, y^{\frac{n-3}{2}} \, e^{-\frac{y}{2}} \, dy = 2 \int_0^\infty \varphi(cx \pm d) x^{n-2+r} \, e^{-\frac{x^2}{2}} \, dx = 2M_{n-2+r}(c, \pm d)$$

for $r = 0, 1, 2, 3$. The rest is just tedious bookkeeping of the terms and their coefficients, and both formulae for the scaled first and second moments follow.

To prove the formulae in (2.3), for $k = 1, 2, \ldots$ we set

$$\begin{aligned} f_{2k}(x) = {} & -e^{-\frac{x^2}{2}} \Big[x^{2k-1} + \sum_{j=1}^{k-1} (2k-1)(2k-3)\cdots(2k-2j+1) x^{2k-2j-1} \Big] \\ & + \big[(2k-1)(2k-3)\cdots 1\big] \sqrt{2\pi}\, \Phi(x), \quad x \in I\!R, \end{aligned}$$

and for $k = 0, 1, 2, \ldots$,

$$f_{2k+1}(x) = -e^{-\frac{x^2}{2}} \Big[x^{2k} + \sum_{j=1}^{k} (2k)(2k-2)\cdots(2k-2j+2) x^{2k-2j} \Big], \quad x \in I\!R.$$

Then it is easy to check that $f'_{2k}(x) = x^{2k} e^{-x^2/2}$ and $f'_{2k+1}(x) = x^{2k+1} e^{-x^2/2}$, $x \in I\!R$. Hence, for all $n \in I\!N$ and $v \in I\!R$,

$$L_n(c, v) = \int_0^\infty \Phi(cx + v) x^n \, e^{-x^2/2} \, dx = \Big[f_n(x) \Phi(cx + v) \Big]_{x=0}^{x=\infty} - c \int_0^\infty \varphi(cx + v) f_n(x) \, dx .$$

Substituting now $n = 2k$, $k = 1, 2, \ldots$, and $n = 2k + 1$, $k = 0, 1, 2, \ldots$, separately, both formulae in (2.3) follow by straightforward work.

Finally, to show (2.4), for $k = 0, 1, 2, \ldots$ and $v \in I\!R$,

$$\begin{aligned} M_k(c, v) &= \int_0^\infty \varphi(cx + v) x^k \, e^{-x^2/2} \, dx = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^k \, e^{-\frac{1}{2}[(cx+v)^2 + x^2]} \, dx \\ &= \frac{e^{-\frac{v^2}{2}}}{\sqrt{2\pi}} \int_0^\infty x^k \, e^{-cvx} \, e^{-\frac{1}{2}(\sqrt{1+c^2}\, x)^2} dx = \frac{e^{-\frac{v^2}{2}}}{(1+c^2)^{(k+1)/2}} \int_0^\infty y^k e^{-\frac{cv}{\sqrt{1+c^2}} y} \varphi(y) \, dy \\ &= \frac{e^{-v^2/2}}{(1+c^2)^{(k+1)/2}} \, G_k\Big(-\frac{cv}{\sqrt{1+c^2}} \Big), \end{aligned}$$

where $G_k(t) = \int_0^\infty x^k \, e^{tx} \, \varphi(x) \, dx = G_0^{(k)}(t)$, and where $G_0(t) = \int_0^\infty e^{tx} \, \varphi(x) \, dx$, $t \in I\!R$, a function related to the moment generating function of the standard normal distribution. It is routine to see that $G_0(t) = G_0^{(0)}(t) = e^{t^2/2}\, \Phi(t) = G^{(0)}(t) = G(t)$, $t \in I\!R$. □

We note that for Brown's estimator $\overline{\sigma}^2_{\mathrm{Brown},n}$ in (1.4), with the adjusted notation $c := t/\sqrt{n-1}$, $\Delta := \mu/\sigma$ and $d := \sqrt{n}\,\Delta$, and with the notation at the beginning of the proof, we have

$$\frac{\overline{\sigma}^2_{\mathrm{Brown},n}}{\sigma^2} = I\big(Z_n \in \ \big) \, \frac{\theta_n}{n-1} \, \chi^2_{n-1} + I\big(Z_n \notin \ \big) \, \frac{1}{n+1} \, \chi^2_{n-1} ,$$

and as easy special case of the above calculation we obtain $\overline{R}_n(t) := E\big(\big[(\overline{\sigma}^2_{\mathrm{Brown},n} - \sigma^2)/\sigma^2\big]^2 = E\big(\overline{\sigma}^4_{\mathrm{Brown},n}/\sigma^4\big) - 2E\big(\overline{\sigma}^2_{\mathrm{Brown},n}/\sigma^2\big) - 1$, where

$$E\left(\frac{\overline{\sigma}^2_{\mathrm{Brown},n}}{\sigma^2}\right) = 2\xi_n\left[\frac{\theta_n}{n-1} - \frac{1}{n+1}\right]\Big[L_n(c,d) + L_n(c,-d)\Big] + 2\,\frac{n-1}{n+1} - \theta_n\,,$$

$$E\left(\frac{\overline{\sigma}^4_{\mathrm{Brown},n}}{\sigma^4}\right) = 2\xi_n\left[\frac{\theta_n^2}{(n-1)^2} - \frac{1}{(n+1)^2}\right]\Big[L_{n+2}(c,d) + L_{n+2}(c,-d)\Big] + 2\,\frac{n-1}{n+1} - \theta_n^2\,\frac{n+1}{n-1}\,.$$

3. Analysis of the quadratic risk formulae

The mean squared error $R_n(\Delta, t, m_0, m)$ in (2.1) with Δ as in (2.2), given by the theorem, can be calculated to any required accuracy using a symbolic mathematics package. The one numerical integration required for $L_0(c, \pm d)$ when n is even may be computed accurately without great computational expense. The veracity of the formula was checked independently for $n = 2$ by direct numerical integration in the defining formula (2.1) and less precisely by simulation for larger sample sizes. While on issues of computation, we correct the following two misprints in [5]: the index $n-3$ in equation (3.3) should be $n+3$, and in the second line of the equation for $\Delta_{\mathrm{rel}}(\eta_0)$ on p. 927 replace ν_0 by ν in the integrand only. Also, the function $\Delta_{\mathrm{rel}}(\eta_0)$ is not monotonically increasing in η_0.

Simple inspection shows that for the risk formula of the theorem we have $R_n(\Delta, t, m_0, m) = R_n(|\Delta|, t, m_0, m)$ for any values of the other three parameters. So, unifying our notation with Rukhin's [5], we write $\eta = |d|$ below when $\mu_0 = 0$, i.e. in the scale-equivariant case, when making comparisons with shrinkage estimators.

The choice of $m = n+1$ in Stein's estimate of the normal variance in (1.7), or in any other estimator $\widehat{\sigma}^2_n(\mu_0) = \widehat{\sigma}^2_n(\mu_0; t, m_0, n+1)$ given in (1.6) is guided by the fact that the limit $\lim_{|\Delta|\to\infty} R_n(\Delta, t, m_0, m)$ is minimized by $m = n+1$ at $2/(n+1)$ of the right side of (1.3), and so any other choice of m would result in an estimator whose risk exceeds that of the estimator $\widetilde{\sigma}^2_n$ for all Δ with $|\Delta|$ large enough. Although the risk may be smaller at small values of $|\Delta|$ for other m, and this is why we included the possibility of choices different from $n+1$, we shall stay with $m = n+1$ from now on. Accordingly, we drop the dependence on $m = n+1$ in the notation: from now on $\widehat{\sigma}^2_n(\mu_0; t, m_0) := \widehat{\sigma}^2_n(\mu_0; t, m_0, n+1)$ and $R_n(\Delta, t, m_0) := R_n(\Delta, t, m_0, n+1)$.

If $m_0 \le n+1$, then $R_n(0, t, m_0) \ge 2/(n+1)$ for any $t > 0$, so from now on we deal only with the case when $m_0 > n+1$. Also we note the phenomenon that the special preliminary-test estimator $\sigma^2_{\mathrm{Stein},n}(0)$ in (1.7) can be written as a simple minimum as in (1.2) can be extended for general $m_0 > n+1$. Indeed, if we define $t^*_n(m_0) = \sqrt{(m_0 - n - 1)(n-1)/(n+1)}$, so that $t^*_n(n+2) = t_{\mathrm{Stein},n}$, with $t_{\mathrm{Stein},n} = \sqrt{(n-1)/(n+1)}$ as in (1.7), then for all $m_0 > n+1$ we have

$$\widehat{\sigma}^2_n\big(\mu_0; t^*_n(m_0), m_0\big) = \min\left\{\frac{S^2_n(\mu_0)}{m_0}, \frac{S^2_n}{n+1}\right\} = \min\left\{\frac{S^2_n(\mu_0)}{m_0}, \widetilde{\sigma}^2_n\right\}. \tag{3.1}$$

This follows by a straightforward extension of the second statement of Problem 4.3.7 in [3], p. 315. Although Stein [6] considered only the scale-equivariant case $\mu_0 = 0$, we continue to call $\widehat{\sigma}^2_n\big(\mu_0; t_{\mathrm{Stein},n}, n+2\big)$ as Stein's estimator since the idea of a preliminary test is the same for all μ_0 and, in general, the risk depends only on the distance between μ_0 and μ.

The complexity of the analytic formula for the risk precludes definitive mathematical statements. The statements below are based on clear numerical evidence; as mathematical statements they should be viewed as conjectures.

First we indicate that, in a certain sense, the Stein choices of m_0 and t are best among the preliminary-test estimators. First, consider the choice of t with $m_0 = n+1$ fixed at the Stein choice. Then computation for small n suggests the following:

(i) *For* $0 < t < t_{\text{Stein},n}$, *the estimator* $\widehat{\sigma}_n^2(\mu_0; t, n+2)$ *dominates the estimator* $\widetilde{\sigma}_n^2$ *but is dominated by Stein's estimator* $\widehat{\sigma}_n^2(\mu_0; t_{\text{Stein},n}, n+2)$, *i.e. the inequalities* $R_n(\Delta, t_{\text{Stein},n}, n+2) < R_n(\Delta, t, n+2) < R_n(\Delta, 0, n+2) = 2/(n+1)$ *hold for all* $\Delta \in I\!R$.

(ii) *For* $t > t_{\text{Stein},n}$, *the risk of the estimator* $\widehat{\sigma}_n^2(\mu_0; t, n+2)$ *exceeds that of the estimator* $\widetilde{\sigma}_n^2$ *for some* Δ, *i.e. there exists a* Δ *such that* $R_n(\Delta, t, n+2) > R_n(\Delta, 0, n+2) = 2/(n+1)$.

This is demonstrated in Figure 1, $n = 2$ and $m_0 = n+2 = 4$, where t is varied. The Stein estimate, $t = t_{\text{Stein},2} = 1/\sqrt{3}$, is best. Other choices of n exhibited similar behavior, only the differences are more pronounced when $n = 2$. In all six figures below, the percentage risk improvement over $\widetilde{\sigma}_n^2$ is plotted on the vertical axes.

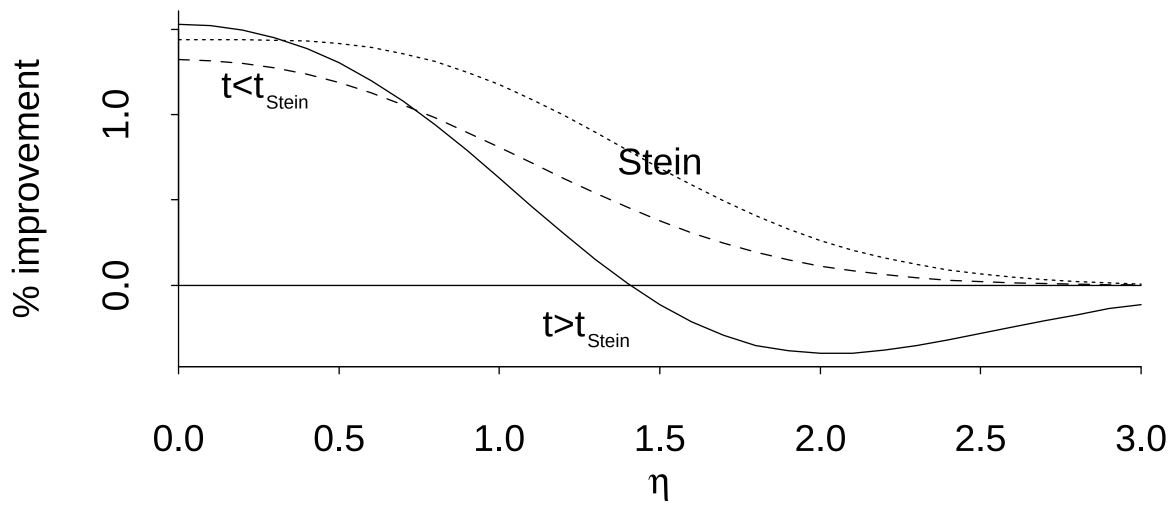

Figure 1. Varying t from $t_{\text{Stein},2} = 1/\sqrt{3}$: $n = 2$, $m_0 = 4$, $t = 0.6/\sqrt{3}$, $t = 1.4/\sqrt{3}$.

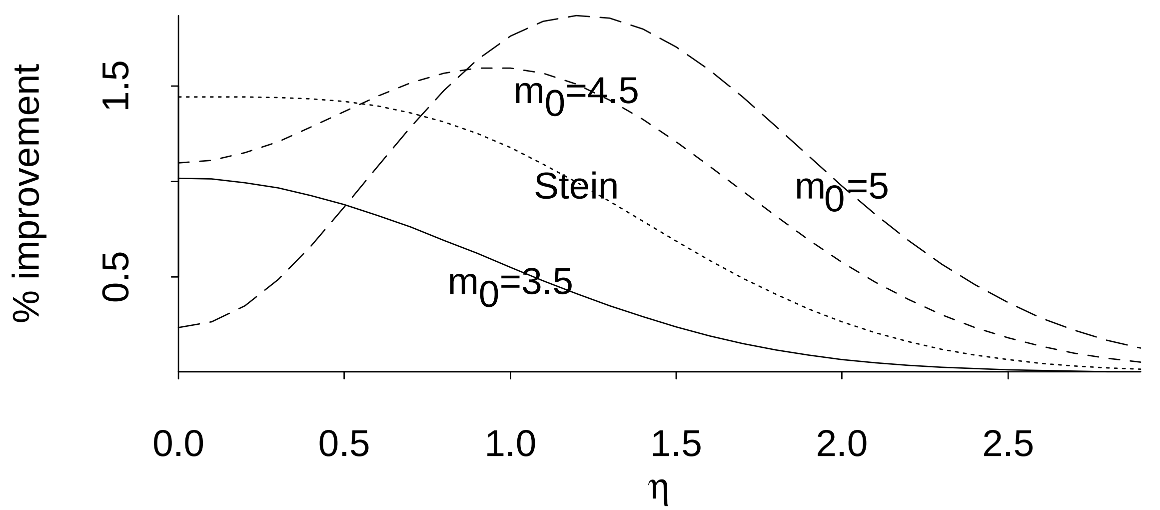

Figure 2. Varying m_0 from Stein's $m_0 = 4$: $n = 2$ and $t = t_2^*(m_0)$.

Next we investigated varying m_0, for which computation for small n suggests the following:

(iii) *For $n+1 < m_0 < n+2$, the estimator $\widehat{\sigma}_n^2(\mu_0; t_n^*(m_0), m_0)$ is dominated by Stein's estimator $\widehat{\sigma}_n^2(\mu_0; t_{\text{Stein},n}, n+2)$, i.e. the inequality $R_n(\Delta, t_n^*(m_0), m_0) > R_n(\Delta, t_{\text{Stein},n}, n+2)$ holds for all $\Delta \in I\!R$.*

(iv) *The risk when $\Delta = 0$, that is when the preliminary test hypothesis is true, is minimized by $m_0 = n+2$.*

In Figure 2 above, $n = 2$, $t = t_2^*(m_0)$ and m_0 is varied. The Stein choice of $m_0 = 4$ is best in the sense described above but larger choices of m_0 can result in a smaller quadratic risk for some values of Δ. Again, other choices of n tell the same story.

Uncertain knowledge of the value of μ can motivate the use of a preliminary-test estimator, so we should require its best performance when $\Delta = \mu - \mu_0 = 0$. The Stein choice of the parameters minimizes the risk at $\Delta = 0$ while still dominating the estimator $\widetilde{\sigma}_n^2$. However, the risk of Stein's estimator is at best 1.44% smaller than the risk of $\widetilde{\sigma}_n^2$ for $n = 2$. The improvement drops off to 1.30% for $n = 10$.

Since, as explained in [4], the starting point for the development of the admissible estimator of Brewster and Zidek [1] is Brown's [2] modification of Stein's idea, it is of some historical interest to consider a comparison of Brown's estimator and our general class of preliminary-test estimators separately before we go over to more general shrinkage estimators.

If the parameter $K := t/\sqrt{n}$ of $\overline{\sigma}^2_{\text{Brown},n}(t)$ in (1.4) is adjusted in a manner to minimize the risk at $\Delta = 0$, the improvement over $\widetilde{\sigma}_n^2$ is only 1.36% for $n = 2$, obtained at $K = 0.31$. Brown's estimator achieves its maximum improvement over $\widetilde{\sigma}_n^2$ at values of $\Delta \neq 0$, claiming gains of 1.6% and 1.2% for sample sizes 2 and 10, respectively. It is possible to adjust m_0 to obtain gains of 1.93% and 1.66%, respectively, using the minimax preliminary-test estimator with $t = t_n^*(m_0)$. Figure 3 compares the performance of Brown's estimator against the preliminary-test estimator for $n = 2$. If minimizing the risk at $\Delta = 0$ is the objective, then Brown's best choice of $K = 0.31$ is dominated by the Stein estimator, but if minimizing the risk for any Δ while still dominating $\widetilde{\sigma}_n^2$ is the aim, then Brown's best choice of $K = 0.55$ is beaten by the preliminary-test estimator with $m_0 = 5.11$.

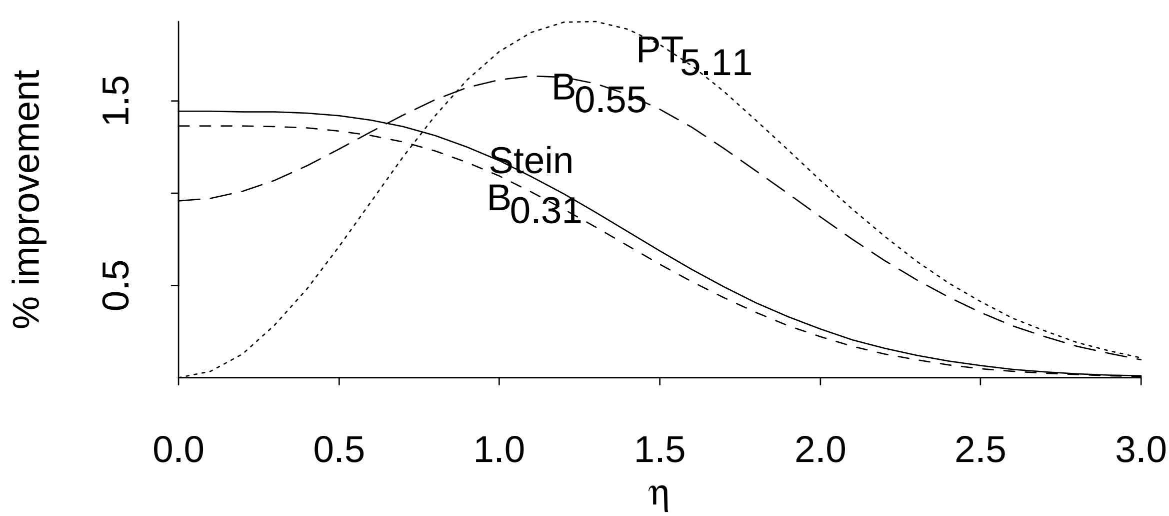

Figure 3. Comparing Brown's estimator to the preliminary-test estimator: $n = 2$, Stein's estimate, Brown's estimate with $K = 0.31$ and $K = 0.55$, and the minimax preliminary-test (PT) estimate with $m_0 = 5.11$.

The preliminary-test estimator should be compared to those of Brewster–Zidek and Rukhin under the same rules. If we restrict ourselves to minimax estimators, the Stein estimate, which is a special case of both the preliminary-test and Rukhin's locally optimal estimators $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\eta)$, achieves its minimum risk at $\eta = 0$. If the game is to achieve the minimum risk for any value of η, then the Rukhin estimator is superior, followed by the Brewster–Zidek estimator. Since the preliminary-test estimator falls in the shrinkage class when it is expressed in the form of a minimum as in (3.1), Rukhin's considerations show that it will be inferior in the locally optimal sense. A comparison of the estimators under this minimax restriction is shown in Figure 4.

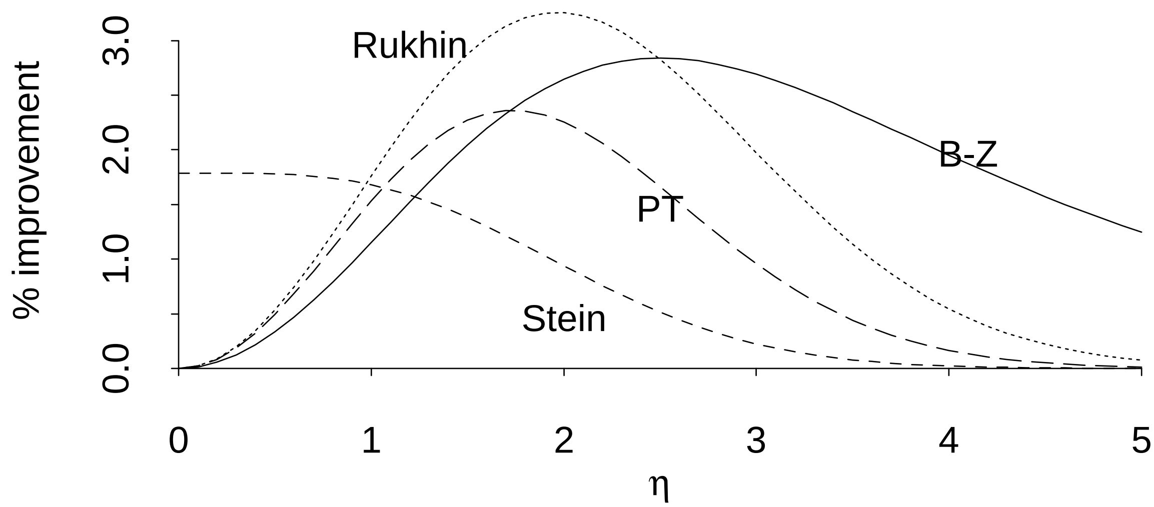

Figure 4. A comparison of Rukhin's locally best minimax estimator $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\overline{\eta}_n)$, the Brewster–Zidek estimator (B-Z), the Stein estimator and the locally best minimax preliminary-test estimator (PT) for $n = 3$.

The most improvement we can hope for using Rukhin's minimax estimator $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\overline{\eta}_n)$ is less than 4% and that will only occur when the true value of η is where the minimum risk occurs. Here $\overline{\eta}_n$ is Rukhin's critical value of $\eta = \sqrt{n}|\mu|/\sigma$, such that $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\eta)$ is minimax for $\eta \leq \overline{\eta}_n$ and not minimax for $\eta > \overline{\eta}_n$. According to Rukhin's Table 3, $\overline{\eta}_3 = 1.38, \overline{\eta}_4 = 1.34, \ldots, \overline{\eta}_8 = 1.27$, so that $\overline{\eta}_n$ is a decreasing function of n. If we happened to have prior information about the value of η and this is away from the minimum risk, we might hope to realize the maximum possible gain by shifting the location of the observations so that the new, shifted value of η is at the minimum. Unfortunately, to determine the right amount of location shifting required, we would need to know σ, so such location shifting is not realizable. Moreover, it seems more likely that the prior information will be about the mean, in which case the Stein estimate will be the best realizable minimax choice.

If we do happen to have prior information about η we may not care about minimaxity – local optimality will be the primary goal. In this situation, the preliminary-test estimator can do substantially better than Rukhin's estimator. Consider the preliminary-test estimator with $t = \infty$, i.e. the estimator $\widehat{\sigma}^2_n(\mu_0; \infty, m_0) = S^2_n(\mu_0)/m_0$. By simple calculation, its risk is

$$\frac{4d^2 + 2n + (n + d^2)^2}{m_0^2} - \frac{2n + 2d^2}{m_0} + 1$$

which is minimized at $|d|$ by the choice of $m_0 = [4d^2 + 2n + (n + d^2)^2]/[n + d^2]$, where the risk is $[4d^2 + 2n]/[4d^2 + 2n + (n + d^2)^2]$. Note that this tends to zero as $|d|$ tends to infinity. Choosing $\mu_0 = 0$, so that $\widehat{\sigma}^2_n(0; \infty, m_0) = S^2_n(0)/m_0$ is scale-equivariant, $|d| = \eta$ as above. In Figure

5, we plot the maximum possible improvement in the risk at a given η over $\widetilde{\sigma}_n^2$ for Rukhin's estimator and this estimator. This estimator seems to be the best that can be found within the preliminary-test class. Again, we use $n = 3$, so that $\overline{\eta}_3 = 1.38$.

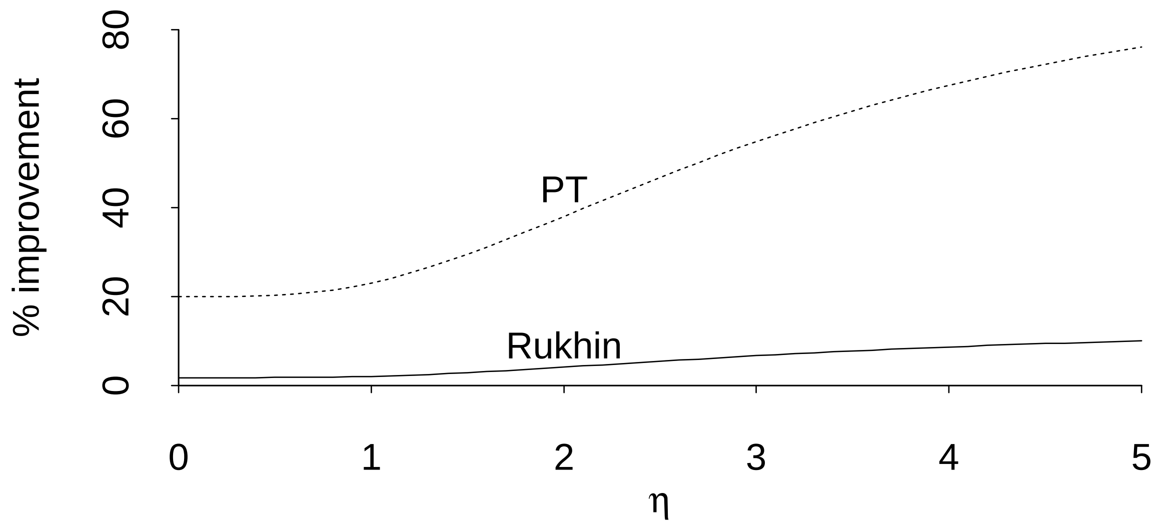

Figure 5. The maximum possible improvement at a given η for Rukhin's locally optimal estimator $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\eta)$ and the locally optimal preliminary-test estimator for $n = 3$.

Of course, if we happen to be wrong about the true value of η this estimator may do very poorly. We can provide some insurance against an unexpected value of η by using another member of the preliminary test family. In Figure 6, we choose a member of the preliminary-test family by choosing m_0 as above at the η where the risk of $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\overline{\eta}_n)$ is minimized and then adjust $t \in (0, \infty)$ so that the risk of the corresponding preliminary-test estimator and the Rukhin's best estimator are equal at that η. As can be seen, neither estimator dominates.

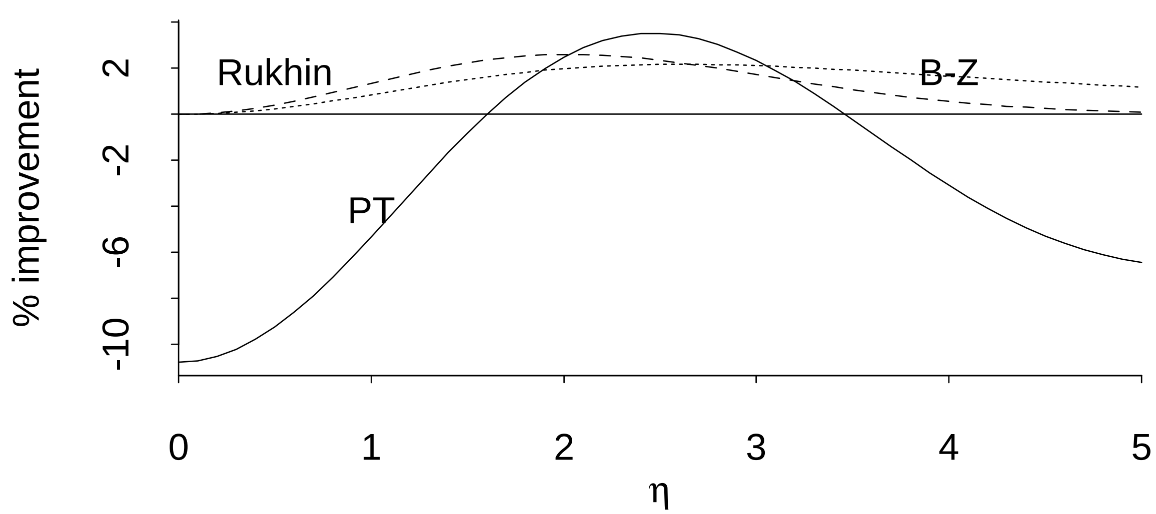

Figure 6. Comparison of the risks of Rukhin's locally best minimax estimator $\overline{\sigma}^2_{\mathrm{Rukhin},n}(\overline{\eta}_n)$, the Brewster–Zidek estimator (B-Z) and a selected preliminary-test estimator (PT) for $n = 3$.

To conclude, if minimaxity is important then the estimator of Rukhin is a good choice. Even so the maximum possible gain over $\widetilde{\sigma}_n^2$ is small and we will be lucky to see much gain at all.

However, if local optimality is important (as it would be with some prior knowledge) then we can realize far greater improvements within the preliminary-test family which lies outside the shrinkage class that has been the focus of most prior work.

References

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

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

[3] Lehmann, E.: Theory of Point Estimation, New York, Wiley (1983)

[4] Maatta, J.M. and Casella, G.: Developments in decision-theoretic variance estimation (With discussion by J.O. Berger, L.D. Brown, A. Cohen, E.I. George, J.T. Hwang, K.B. MacGibbon and G.E. Shorrock, A.L. Rukhin, and W.E. Strawderman). Statist. Sci. **5**, 90–120 (1990)

[5] Rukhin, A.L.: How much better are better estimators of a normal varianceΓ J. Amer. Statist Assoc. **82**, 925–928 (1987)

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

[7] Zacks, S.: The Theory of Statistical Inference, New York, Wiley (1971)