Ann Inst Stat Math (2014) 66:811–832
DOI 10.1007/s10463-013-0425-x

Estimation of a non-negative location parameter with unknown scale

Mohammad Jafari Jozani · Éric Marchand · William E. Strawderman

Received: 8 November 2012 / Revised: 1 April 2013 / Published online: 5 September 2013

Abstract For a vast array of general spherically symmetric location-scale models with a residual vector, we consider estimating the (univariate) location parameter when it is lower bounded. We provide conditions for estimators to dominate the benchmark minimax MRE estimator, and thus be minimax under scale invariant loss. These minimax estimators include the generalized Bayes estimator with respect to the truncation of the common non-informative prior onto the restricted parameter space for normal models under general convex symmetric loss, as well as non-normal models under scale invariant L^p loss with $p > 0$. We cover many other situations when the loss is asymmetric, and where other generalized Bayes estimators, obtained with different powers of the scale parameter in the prior measure, are proven to be minimax. We rely on various novel representations, sharp sign change analyses, as well as capitalize on Kubokawa's integral expression for risk difference technique. Several properties such as robustness of the generalized Bayes estimators under various loss functions are obtained.

Keywords Dominance · Generalized Bayes · Lower bounded mean · L^p loss · Minimax · Robustness

M. Jafari Jozani (✉)
Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
e-mail: m_jafari_jozani@umanitoba.ca

É. Marchand
Département de mathématiques, Université de Sherbrooke,
Sherbrooke, QC J1K 2R1, Canada

W. E. Strawderman
Department of Statistics and Biostatistics, Rutgers University,
561 Hill Center, Busch Campus, Piscataway, NJ 08854-8019, USA

1 Introduction

1.1 Preamble

We begin with the normal model in canonical form

$$X \sim N(\mu, \sigma^2),\, S^2 \sim \sigma^2 \chi_n^2, \quad \text{independent } (n \geq 1), \tag{1}$$

which plays a central role in both statistical theory and practice. Consider situations where additional information on (μ, σ) is available in terms of parametric restrictions. Bayesian inference in such restricted parameter space problems does not, conceptually, present any difficulties as both the prior and the resulting posterior will be adapted and will adapt to the constraints. Assessing the frequentist performance of Bayesian estimators in such situations is, however, considerably more challenging. Such assessments may include, for instance, testing for minimaxity, an evaluation in comparison to a benchmark procedure such as minimum risk equivariant (MRE) estimator or a maximum likelihood estimator (mle), or a study of the frequentist performance of associated Bayesian confidence intervals.

As an illustration, consider model (1) with known σ and the non-negative mean restriction $\mu \geq 0$. Despite early discoveries by Katz (1961) and Sacks (1963) that the generalized Bayes estimator with respect to the flat prior on $[0, \infty)$ is minimax and dominates the MRE estimator $\delta_0(X, S) = X$ under squared error loss, despite various generalizations to other models and location invariant losses (Farrell 1964; Kubokawa 2004; Marchand and Strawderman 2005), no other Bayes minimax estimators were known until the Maruyama and Iwasaki (2005) findings which provide other Bayes minimax estimators under squared error loss. Even then, little has been obtained for estimating μ in (1) for $\mu \geq 0$ and unknown σ. In this case, Kubokawa (2004) obtained, for scale invariant squared error loss, a class of minimax improvements on δ_0, which includes the generalized Bayes estimator $\delta_{\pi_0}(X, S)$ with respect to the truncation of the usual non-informative prior onto the restricted parameter space (see expression 6).

Our main motivation for his work has been to generalize and better understand Kubokawa's findings. The paper consists of various extensions with respect to the loss, the model, and the prior; which bypass in a unified way the specific normal case-squared error loss calculations by Kubokawa. Several new technical aspects have been developed to meet such challenges.

1.2 The problem

As an extension of model (1), we consider spherically symmetric models for an observable $(X, U) = (X, U_1, \ldots, U_n)$ with density proportional to

$$\frac{1}{\sigma^{n+1}} f\left(\frac{(x-\mu)^2 + \|u\|^2}{\sigma^2}\right), \tag{2}$$

and with $n \geq 1,\ \mu \geq 0,\ \sigma > 0$. The function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is known, and it assumed throughout that:

$$f' < 0, \quad \text{and} \quad \frac{tf'(t)}{f(t)} \text{ decreases in } t \text{ for } t > 0. \tag{3}$$

Hereafter, for conciseness, reference to model (2) shall be understood to encompass these assumptions on f. Multivariate (for X) versions of (2) have been previously considered, namely in recent work where robust minimax generalized Bayes estimators of μ without constraints are provided (see Fourdrinier and Strawderman 2010). Various other features of such models are described in Sect. 2.1.

We consider estimating μ where it is assumed that $(\mu, \sigma) \in \Theta = \{(\mu, \sigma) : \mu \geq 0, \sigma > 0\}$ under location and scale invariant loss

$$\rho\left(\frac{d-\mu}{\sigma}\right), \tag{4}$$

with (i) ρ absolutely continuous a.e., (ii) ρ strictly bowled shaped with $\rho(t) \geq \rho(0) = 0$ for all $t \in \mathbb{R}$, $\rho' < 0$ on $(-\infty, 0)$ and $\rho' > 0$ on $(0, \infty)$. We also assume that the pair (f, ρ) leads to risk finiteness, namely that there exists a unique minimum risk equivariant estimator for the unconstrained problem. In such cases, it is given by $\delta_0(X, S) = X + c_0 S$ with constant risk $R((\mu, \sigma), \delta_0) = E_{0,1}[\rho(X + c_0 S)]$, and with (also see Remark 3)

$$c_0 = \operatorname{argmin}_c\{E_{0,1}[\rho(X + cS)]\}, \tag{5}$$

which is uniquely determined by $E_{0,1}[S\rho'(X + c_0 S)] = 0$. It is also worth pointing out that $c_0 = 0$ for symmetric losses ρ, and consequently that the MRE estimator coincides with the unbiased estimator X, and is robust with respect to the choice of the underlying model density f. It follows from Kiefer (1957) that δ_0 is minimax for the unconstrained problem. With the constraint on μ, $\delta_0(X, S)$ produces indeed implausible estimates, but it remains minimax [see Marchand and Strawderman (2012), and references therein] for general ρ, and its constant risk thus matches the minimax risk. The challenge here is to search for good improvements on $\delta_0(X, S)$ that capitalize on the parametric information, and we focus on potential Bayesian improvements such as the generalized Bayes estimators δ_{π_l} with respect to the prior measures

$$\pi_l(\mu, \sigma) = \frac{1}{\sigma^{l+1}}\, \mathbb{I}_{[0,\infty)}(\mu) 1_{(0,\infty)}(\sigma); \quad l \geq -(n-1); \tag{6}$$

the lower bound on l required for the posterior density to be well defined. The class includes the choice π_0 which is of intrinsic interest as it represents a plausible adaptation, or truncation onto Θ of the right Haar invariant measure π_{rh} with the MRE estimator (also) being the generalized Bayes estimator $\delta_{\pi_{\text{rh}}}$ with respect to π_{rh}. Moreover, the study of frequentist properties on the restricted parameter space of Bayesian procedures associated with π_0 or, moregenerally, truncations of the right Haar invariant prior measure has recently surfaced in interval estimation problems (Zhang and Woodroofe 2003; Marchand and Strawderman 2006, 2008, 2013).

In Sect. 2, we further describe features of the underlying model and present various expressions, properties, and illustrations relative to the Bayes estimators δ_{π_l}. Namely, we establish a robustness property, applicable to scale invariant L^p loss with $\rho(t) = |t|^p$, $p > 0$, and asymmetrized versions as given in (13), stating that the Bayes estimator δ_{π_l} does not depend on the underlying f in (2).

The developments of Sect. 3 make use of Kubokawa (1994) Integral Expression of Risk Difference (IERD) technique to derive classes of dominating (minimax) estimators of $\delta_0(X, S) = X + c_0 S$. With further analyses, which bring into play novel technical arguments of interest on their own, we provide several instances of (f, ρ) where these classes of minimax estimators include Bayesian estimators of the type δ_{π_l}. Namely, we establish in Sects. 4 and 5 that:

(A) The Bayes estimators δ_{π_l} with $l \geq 0$ dominate δ_0 for normal models in (1) and general convex ρ's such that ρ is even. The estimator δ_{π_0} also dominates δ_0 for asymmetric $\rho' s$ such that $|\rho'(u)| \geq |\rho'(-u)|$ for all $u > 0$;

(B) The Bayes estimators δ_{π_l} with $l \geq 0$ dominate δ_0 for all (fixed) f in (2) satisfying assumption (3), and whenever the loss is scale invariant L^p, $p \geq 1$. The estimator δ_{π_0} also dominates δ_0 for asymmetrized versions as given in (13) (where $|\rho'(u)| \geq |\rho(-u)|$ for all $u > 0$ as in (A);

(C) The Bayes estimator δ_{π_0} dominates δ_0 for all (fixed) f in (2) satisfying assumption (3), and whenever the loss is scale invariant L^p with $p \in (0, 1)$.

The ensemble of results provide extensions of Kubokawa's normal case, scale invariant squared error loss result applicable to δ_{π_0} in three directions: choice of f, choice of ρ, and applicability to other Bayesian estimators δ_{π_l}'s. Moreover, the developments relative to (A), (B), and (C) are unified and contain two alternative proofs replicating Kubokawa's result. It is also notable that (C) involves the case of a concave in $|\frac{d-\mu}{\sigma}|$ (and hence non-convex) loss. Finally, various other observations, including non-minimaxity results, are also given throughout the exposition and in Sect. 6.

2 Preliminary results and properties of the estimator δ_{π_l}

2.1 The underlying model

In (2) and (3), the density of (X, U) is unimodal with central location parameter $(\mu, 0, \ldots, 0)$ and scale parameter σ. Our parameter of interest is the non-negative μ, or median, of X, while U is a residual vector. Condition (3) is equivalent to an increasing monotone likelihood ratio (mlr) in $(X - \mu)^2 + \|U\|^2$ of the family of densities in (2) when viewed as a scale family (parameter σ) with known μ. Assumption (3) is, for unimodal and symmetric densities, weaker than both (a) the logconcavity of $f(y)$ and (b) the logconcavity of $f(y^2)$ for $y > 0$, with (a) implying (b), and with (b) equivalent to an increasing mlr property in X of the family of densities in (2) when viewed as a location family (parameter μ) with known σ.

The most important and best known case covered by (2) and (3) is the normal case where $(X, U) \sim N_{n+1}((\mu, 0, \ldots, 0), \sigma^2 I_{n+1})$ and $f(t) \propto e^{-t/2}$. However, our inference results will also apply to many other models such as (i) exponential power densities with $f(t) \propto e^{-\alpha t^p}$, $p > 0$, $\alpha > 0$, including Laplace densities arising for $p = 1/2$; (ii) the Kotz distribution with $f(t) \propto t^m e^{-\alpha t}$, $m \in (-1/2, 0), \alpha > 0$; as well as for (iii) Student densities with $f(t) \propto (1 + t/\nu)^{-(\nu+n+1)/2}$, $\nu \geq 1$ degrees of freedom. The Student example illustrates a non-logconcave f (in fact, it is logconvex) which satisfies the weaker assumptions required here. The Student distributions, which are scale mixtures of normals, often serve as useful, alternative models to the normal model. Here is an interesting general situation for which scale mixtures inherit assumption (3).

Lemma 1 *A scale mixture of the form* $f(t) = \int_0^\infty v f_0(tv) h(v) dv$ *satisfies assumption* (3) *as soon as both* f_0 *and* h *satisfy assumption* (3).

Proof See Appendix. □

Remark 1 In the Student case above, both f_0 (a normal density) and h (a gamma density), are log-concave and satisfy (3).

Remark 2 We note that model (2) arises for observables $Y_1, \ldots, Y_{n+1}$ having joint density

$$\frac{1}{\sigma^{n+1}} f\left(\frac{\sum_i (y_i - \theta)^2}{\sigma^2}\right),$$

through an orthogonal transformation

$$(Y_1, \ldots, Y_{n+1}) \to (X = \sqrt{n}\bar{Y}, U_1, \ldots, U_n),$$

with $\mu = \sqrt{n}\theta$.

For model (2), $(X, S = \|U\|)$ is a sufficient statistic with joint density $f_{X,S}$ on $\mathbb{R} \times \mathbb{R}^+$ which we take as equal to:

$$\frac{s^{n-1}}{\sigma^{n+1}} f\left(\frac{(x-\mu)^2 + s^2}{\sigma^2}\right). \tag{7}$$

For the normal model canonical form in (1), we will write the joint density of (X, S) in (7) as $\frac{1}{\sigma^2}\phi(\frac{x-\mu}{\sigma})h(\frac{s}{\sigma})$, with

$$\phi(u) = (2\pi)^{-1/2} e^{-u^2/2}, \quad \text{and} \quad h(v) = \frac{v^{n-1}\, e^{-v^2/2}}{\Gamma(n/2)\, 2^{n/2-1}}. \tag{8}$$

2.2 Properties of the Bayes estimators δ_{π_l}

We proceed with various preliminary results, observations, and illustrations concerning the generalized Bayes estimators $\delta_{\pi_l}(X, S)$ with respect to the improper priors in (6). As previously mentioned, one can verify that the lower bound on the power l in (6) guarantees that the posterior density of (μ, σ) is well defined given that (7) is a density. Even with a well-defined posterior density, we further assume, and not necessarily emphasize (mainly in Sects. 3, 4, and 5), that the pair (f, ρ) leads to the existence of the Bayes estimator δ_{π_l}.

We define for $m > 0$, $w \in \mathbb{R}$, $z \in \mathbb{R}$,

$$B_m(w, z) = \int_0^\infty \int_{-\infty}^{vw} \rho'(u + c_0 v + zv) v^m f(u^2 + v^2) \mathrm{d}u \mathrm{d}v, \tag{9}$$

provided it exists. The function $B_m(w, z)$, as well as some of its properties (see for instance Lemma 4) will play a key role below, namely in the following representation of the Bayes estimator $\delta_{\pi_l}(X, S)$.

Lemma 2 *Under model* (7)*, provided existence of the Bayes estimator δ_{π_l}, we have $\delta_{\pi_l}(X, S) = X + c_0 S + g_{\pi_l}(\frac{X}{S})S$, where $g_{\pi_l}(y)$ satisfies, for all $y \in \mathbb{R}$, $l \geq -(n-1)$,*

$$B_{n+l}(y, g_{\pi_l}(y)) = 0. \tag{10}$$

Proof Writing an estimator as $X + c_0 S + g(X, S)$, we have that the Bayes estimate $\delta_{\pi_l}(x, s)$ minimizes in $g(x, s)$ the expected posterior loss:

$$E\left[\rho\left(\frac{x + c_0 s + g(x, s) - \mu}{\sigma}\right) \middle| (X, S) = (x, s)\right],$$

or, equivalently,

$$\int_0^\infty \int_0^\infty \rho\left(\frac{x + c_0 s + g(x, s) - \mu}{\sigma}\right) \frac{s^{n-1}}{\sigma^{n+1}} f\left(\frac{(x-\mu)^2 + s^2}{\sigma^2}\right) \frac{1}{\sigma^{l+1}} \mathrm{d}\mu \mathrm{d}\sigma.$$

With the change of variables $(\mu, \sigma) \to (u = \frac{(x-\mu)}{\sigma}, v = \frac{s}{\sigma})$, the Bayes estimate $\delta_{\pi_l}(x, s)$ is seen to minimize in $g(x, s)$:

$$\int_0^\infty \int_{-\infty}^{vx/s} \rho\left(u + c_0 v + \frac{v}{s} g(x, s)\right) f(u^2 + v^2)\, v^{n+l-1} \mathrm{d}u \mathrm{d}v.$$

Now, observe that $\frac{1}{s} g(x, s)$ depends on (x, s) only through the function $y = x/s$, which implies that the estimator $\delta_{\pi_l}(X, S)$ is of the form $X + c_0 S + g_{\pi_l}(\frac{X}{S})\, S$ with $g_{\pi_l}(y)$ minimizing in $g(y)$ the quantity

$$\int_0^\infty \int_{-\infty}^{vy} \rho(u + c_0 v + g(y) v)\, f(u^2 + v^2)\, v^{n+l-1} \mathrm{d}u \mathrm{d}v. \tag{11}$$

Finally, the result is obtained by differentiation. □

We point out that $g_{\pi_l}(y)$ is uniquely determined (Lemma 4), and is a continuous function of y such that

$$g_{\pi_l}(y) \geq -y - c_0 \quad \text{for all } y \in \mathbb{R}. \tag{12}$$

This must indeed be the case as the Bayes estimates $\delta_{\pi_l}(x,s)$ are necessarily non-negative, and $\frac{1}{s}\delta_{\pi_l}(x,s) \geq 0 \iff \frac{x}{s} + c_0 + g_{\pi_l}(\frac{x}{s}) \geq 0$. We pursue with an intriguing robustness property, and alternative representation, of the Bayes estimators δ_{π_l} for scale invariant L^p loss, and their asymmetrized versions given by

$$\rho_{c_1,c_2}(t) = c_1\,|t|^p\,\mathbb{I}_{(-\infty,0)}(t) \;+\; c_2\,|t|^p\,\mathbb{I}_{[0,\infty)}(t), \tag{13}$$

with $p > 0,\ c_1 > 0$, and $c_2 > 0$.

Lemma 3 *For losses ρ_{c_1,c_2} as in (13), the Bayes estimators δ_{π_l}, given in Lemma 2, do not depend on the underlying model density f provided they exist.*

Proof From (11), we have

$$\begin{aligned}
c_0 + g_{\pi_l}(y) &= \underset{h}{\operatorname{argmin}} \int_0^\infty \int_{-\infty}^{vy} \rho_{c_1,c_2}(u+hv)\, f(u^2+v^2)\, v^{n+l-1}\mathrm{d}u\mathrm{d}v \\
&= \underset{h}{\operatorname{argmin}} \int_0^\infty \int_{-\infty}^{vy} \rho_{c_1,c_2}\left(\frac{u}{v}+h\right)\, f(u^2+v^2)\, v^{n+l+p-1}\mathrm{d}u\mathrm{d}v \\
&= \underset{h}{\operatorname{argmin}} \left(\int_0^\infty x^{(n+l+p-1)/2} f(x)\,\mathrm{d}x\right)\left(\int_{-\infty}^{y} \frac{\rho_{c_1,c_2}(t+h)}{(1+t^2)^{(n+l+p+1)/2}}\mathrm{d}t\right) \\
&= \underset{h}{\operatorname{argmin}} \int_{-\infty}^{y} \frac{\rho_{c_1,c_2}(t+h)}{(1+t^2)^{(n+l+p+1)/2}}\mathrm{d}t,
\end{aligned} \tag{14}$$

by making use of the homogeneity of ρ_{c_1,c_2} and the change of variables $(u,v) \to (t = u/v, x = u^2+v^2)$. Finally, expression (14) tells us that $\delta_{\pi_l}(x,s) = x + s(c_0 + g_{\pi_l}(x/s))$ is independent of f. □

This type of property seems to have first been noticed by Maruyama (see Maruyama 2003; Maruyama and Iwasaki 2005) in a multivariate setting under L^2 loss.

Remark 3 (Minimum risk equivariant estimator)

(a) Proceeding as in the proof of Lemma 2, we obtain the useful representation $X + c_0(n)S$ for the MRE estimator, with the defining equation

$$\int_0^\infty \int_{-\infty}^\infty \rho'(u + c_0(m)v)\, v^m\, f(u^2+v^2)\mathrm{d}u\mathrm{d}v = 0, \tag{15}$$

for $c_0(m),\ m \geq 1$.

(b) A robustness property similar to Lemma 3 (also illustrated in Example 1, part C) is shared by the MRE estimators with respect to losses ρ_{c_1,c_2} and can be established by expanding (5) showing that

$$c_0 = \underset{c}{\operatorname{argmin}} \int_{-\infty}^{\infty} \frac{\rho_{c_1,c_2}(t+c)}{(1+t^2)^{(n+p+1)/2}} \mathrm{d}t. \tag{16}$$

Example 1 (scale invariant L^2 loss, scale invariant L^1 loss and their asymmetrized versions)

(A) For scale invariant squared error loss with $\rho(t) = t^2$ in (4), the MRE estimator is $\delta_0(X) = X$, provided the second moment of X under (2) exists. Lemma 2 as well as (14) provide representations $X + g_{\pi_l}(\frac{X}{S})S$ for the Bayes estimator $\delta_{\pi_l}(X, S)$; $l > -(n-1)$. Differentiating (14) with respect to h, we obtain directly for $y \in \mathbb{R}$

$$g_{\pi_l}(y) = -E[T|T \le y], \tag{17}$$

where T has density on $\mathbb{R}$ proportional to $(1+t^2)^{-(n+l+3)/2}$. Here, the distribution of T is a multiple of a Student distribution with $n+l+1$ degrees of freedom. Equivalently from (10), we have

$$\begin{aligned} & B_{n+l}(y, g_{\pi_l}(y)) = 0 \\ & \iff \int_0^{\infty} \int_{-\infty}^{vy} (u + g_{\pi_l}(y)v)\, v^{n+l}\, f(u^2+v^2) \mathrm{d}u\mathrm{d}v = 0 \\ & \iff g_{\pi_l}(y) = -\frac{\int_0^{\infty} \int_{-\infty}^{vy} \frac{u}{v}\, v^{n+l+1}\, f(u^2+v^2)\mathrm{d}u\mathrm{d}v}{\int_0^{\infty} \int_{-\infty}^{vy} v^{n+l+1}\, f(u^2+v^2)\mathrm{d}u\mathrm{d}v}, \end{aligned} \tag{18}$$

illustrating the fact that the distribution of T arises as the (independent of f) distribution of the ratio $\frac{U}{V}$, with (U, V) having joint density on $\mathbb{R} \times \mathbb{R}^+$ proportional to $v^{n+l+1} f(u^2+v^2)$. From representation (17), observe that $g_{\pi_l}(\cdot)$ decreases on $\mathbb{R}$ with $\lim_{y\to\infty} g_{\pi_l}(y) = 0$ (since $\int_{-\infty}^{\infty} uf(u^2+v^2)du = 0$ for all $v > 0$), and hence that $g_{\pi_l}(\cdot)$ is positive, i.e., δ_{π_l} expands on the MRE δ_0. Such properties are of interest as they indicate that the amplitude of the expansion $\delta_{\pi_l}(x, s) - \delta_0(x, s)$ decreases in x for fixed s, and increases in s for fixed x (in fact $(\delta_{\pi_l}(x, s) - \delta_0(x, s))/s$ increases in s). Such a property resonates back to Katz (1961) where in the normal case with known σ, the Bayes estimator with respect to a flat prior for μ on $(0, \infty)$ expands X by the amount $\sigma \frac{\phi(x/\sigma)}{\Phi(x/\sigma)}$ which decreases in x and increases in σ. Below, we establish such properties for general convex ρ in Lemma 5, as well as scale invariant L^p concave loss with $p \in (0, 1)$ in Lemma 5. Finally, we point out that alternative expressions for δ_{π_1} in the above normal case were given by Kubokawa (2004), as well as Marchand et al. (2012).

(B) As above, for scale invariant absolute value error loss with $\rho(t) = |t|$ in (4), the MRE estimator is $\delta_0(X) = X$. For $l \ge -(n-1)$, $\delta_{\pi_l}(X, S) = X + g_{\pi_l}(\frac{X}{S})S$ is obtainable from (14) yielding

$$g_{\pi_l}(y) = -\text{median}[T|T \le y] = -F_{n+l}^{-1}\left(\frac{F_{n+l}(y)}{2}\right), \tag{19}$$

where F_m and F_m^{-1} are the cdf and inverse cdf of T having density on $\mathbb{R}$ proportional to $(1+t^2)^{-(m+2)/2}$. As above, it is easily seen directly that such a $g_{\pi_l}(\cdot)$ decreases on $\mathbb{R}$, that $\lim_{y\to\infty} g_{\pi_l}(y) = 0$, that $\delta_{\pi_l}(x,s)$ expands once again on $\delta_0(x,s)$ for all $(x,s) \in \mathbb{R} \times \mathbb{R}^+$, and the difference between these estimates decreases in x/s.

(C) Consider now asymmetrized L^1 losses ρ_{c_1,c_2} in (13) with $p=1$. By making use of Remark 3, the MRE estimator is given by $\delta_0(X) = X + c_0 S$, with c_0 independent of f, and $c_0(n) = -F_n^{-1}(\frac{c_2}{c_1+c_2})$ and F_n^{-1} the inverse cdf given in part (B). For $l \ge -(n-1)$, we obtain from (14) $\delta_{\pi_l}(X,S) = X + c_0(n)S + g_{\pi_l}(\frac{X}{S})S$ with $g_{\pi_l}(y) = -c_0(n) - F_{n+l}^{-1}(\frac{c_2}{c_1+c_2}F_{n+l}(y))$, thus extending (19) which occurs for $c_1 = c_2$. Observe here that $\lim_{y\to\infty} g_{\pi_l}(y) = -c_0(n) + c_0(n+l)$, which does not equal 0 in general, the exception being precisely $l=0$, and/or $c_1 = c_2$. This property is more general as seen below in Lemma 5.

We pursue with further properties relative to $B_m(\cdot,\cdot)$ and g_{π_l} (applicable when these quantities exist).

Lemma 4 *For all* $a>0$, $y\in\mathbb{R}$, $l \ge -(n-1)$, *and strictly bowled-shaped* ρ,

(a) $B_{n+l}(y+a, g_{\pi_l}(y)) > 0$;
(b) $B_{n+l}(y,z)$ *is non-decreasing in* z *whenever* ρ *is also convex;*
(c) $\lim_{y\to\infty} B_{n+l}(y,0) = 0$ *whenever* $l=0$; *or whenever* $l \ne 0$ *and* ρ *is an even function.*

Proof Part (b) is obvious given the convexity of ρ, while part (c) follows from the given representations (15) and (10). For establishing (a), suppose, to arrive at a contradiction, that $B_{n+l}(y+a, g_{\pi_l}(y)) \le 0$. This would imply $C_1 \le 0$, where

$$C_1 = \int_0^\infty \int_{vy}^{v(y+a)} \rho'(u + c_0 v + g_{\pi_l}(y)v)\; v^{n+l} f(u^2+v^2)\mathrm{d}u\mathrm{d}v.$$

Now, observe that for $(u,v) \in I(u,v) = \{(u,v) : vy < u < v(y+a)\}$, we have by (12): $u+c_0v+g_{\pi_l}(y)v > vy+c_0v+g_{\pi_l}(y)v \ge 0$, implying $\rho'(u+c_0v+g_{\pi_l}(y)v) > 0$, (for such $(u,v)'s \in I(u,v)$). This renders $C_1 \le 0$ impossible, and yields the result. □

The strictly decreasing property of g_{π_l} that follows in Lemma 5 is a critical property that we will exploit later for the risk comparisons. We do not know how far the property can be extended for non-convex ρ, but we do establish here, and use later, such a property for L_p losses and their asymmetrized versions for the non-convex choices $p \in (0,1)$.

Lemma 5 *For* $l \ge -(n-1)$,

(a) $g_{\pi_l}(y)$ *is strictly decreasing in* y *whenever* ρ *is convex;*
(b) $g_{\pi_l}(y)$ *is strictly decreasing in* y *whenever the loss is* ρ_{c_1,c_2} *as in* (13) *with* $p \in (0,1)$.

(c) *For strictly bowled-shaped* ρ, $\lim_{y\to\infty} g_{\pi_l}(y) = -c_0(n) + c_0(n+l)$, *where* $c_0(m)$ *is defined in* (15). *Consequently,* $\lim_{y\to\infty} g_{\pi_l}(y) = 0$ *whenever* $l = 0$, *or* $l \neq 0$ *and* ρ *is even.*

Proof (a) It suffices to show that we cannot have $g_{\pi_l}(y+\epsilon) \geq g_{\pi_l}(y)$ for some $y \in \mathbb{R}$, $\epsilon > 0$. Indeed, if this were the case, it would follow, using defining equation (10) and part (a) of Lemma 4, that

$$0 = B_{n+l}(y+\epsilon, g_{\pi_l}(y+\epsilon)) \geq B_{n+l}(y+\epsilon, g_{\pi_l}(y)) > 0,$$

which is not possible.

(b) Set $s(y) = -c_0 - g_{\pi_l}(y)$ and rewrite representation (14) as

$$s(y) = \underset{s}{\operatorname{argmin}}\, E[\rho_{c_1,c_2}(T-s)|T \leq y], \tag{20}$$

with T having density proportional to $(1+t^2)^{-(n+l+p+1)/2}$ on $\mathbb{R}$. Observe that the family of densities for $T|T \leq y$ has strictly increasing monotone likelihood ratio in T with parameter y. Now, consider, for $a_1 < a_2$, the function $\rho_{c_1,c_2}(t-a_1) - \rho_{c_1,c_2}(t-a_2)$, which changes signs once from $-$ to $+$ as a function of t as t increases on $\mathbb{R}$, and infer that

$$H(a_1, a_2, y) = E[\rho_{c_1,c_2}(T-a_1) - \rho_{c_1,c_2}(T-a_2)]$$

has a single root, and changes signs once from $-$ to $+$, as a function of y, as y increases on $\mathbb{R}$, given the mlr property (e.g., Lehmann 1986). Suppose now, to arrive at a contradiction that g_{π_l} is not strictly decreasing, i.e., s is not strictly increasing and there exists $y_2 < y_1$ such that $a_2 = s(y_2) \geq s(y_1) = a_1$. Then, we would have with the definition of $s(y)$ in (20) and the properties of H: $H(s(y_1), s(y_2), y_2) > 0$ and $H(s(y_1), s(y_2), y_1) < 0$ which leads to a contradiction and establishes the result.

(c) This follows by matching expression (10) when $y \to \infty$ with (15). □

Remark 4 The above proof in (b) goes through for all losses ρ_{c_1,c_2}, including the convex cases with $p \geq 1$.

The following results permit the ordering of Bayes estimators δ_{π_l} in terms of the power l in the prior measure π_l in (6).

Lemma 6 *For the normal model in (1),* $y \in \mathbb{R}$, *and convex and even* ρ, *the quantities* $g_{\pi_l}(y)$ *decrease in* l, $l \geq -(n-1)$, *provided they exist.*

Proof See Appendix. □

Corollary 1 *For models* (2) *with* f *satisfying assumption* (3), $y \in \mathbb{R}$, *and scale invariant* L^p *loss with* $p > 0$, $g_{\pi_l}(y)$ *decreases in* l, $l \geq -(n-1)$, *provided existence.*

Proof Lemma 3 tells us that $g_{\pi_l}(y)$ is independent of f and, thus, matches the normal model $g_{\pi_l}(y)$ and Lemma 6 tells us that such $g_{\pi_l}(y)$'s decrease in l whenever ρ is even as for the L^p loss here. □

3 Minimax conditions for general ρ and f

For estimating $\mu \geq 0$ in (2) or in (7) with unknown $\sigma > 0$ under strictly bowled-shaped loss $\rho(\frac{d-\mu}{\sigma})$, we establish here useful sufficient conditions for an estimator $\delta(X, S)$ to be minimax. We first make use of Kubokawa's IERD technique in Theorem 1. Proposition 1 (below) then extracts a sign varying condition for minimaxity which will serve as the basis for further analysis for the specific cases of normal models and general convex ρ in Sect. 4, and for L^p losses and their asymmetric versions ρ_{c_1,c_2} with general f satisfying assumptions (3) in Sect. 5. Various other technical results and remarks, including a condition for non-minimaxity with applications, are also introduced in this section. We consider the following subclass of scale invariant estimators.

Definition 1 $C = \{\delta_g(X, S) : \delta_g(X, S) = \delta_0(X, S) + g(\frac{X}{S})\, S$, with g absolutely continuous a.e., non-increasing, non-constant, and $\lim_{t\to\infty} g(t) = 0\}$.

These estimators in C expand upon δ_0, in view of the restriction $\mu \geq 0$, include δ_{π_0} and the generalized Bayes estimators δ_{π_l}; $l \neq 0,\ l \geq -(n-1)$; for even ρ as seen by the properties given in Lemma 5. Under invariant losses as in (4), such estimators will have frequentist risk $R(\theta, \delta_g)$ depending on $\theta = (\mu, \sigma)$ only through the maximal invariant $\lambda = \mu/\sigma$, and we seek conditions for which such a risk falls below the constant risk of the MRE estimator δ_0 for all $\lambda \geq 0$. As mentioned above, such improvements will necessarily be minimax estimators since δ_0 is minimax. Hereafter, we will just refer, for the most part, to such improvements as being minimax estimators. The focus is largely on the generalized Bayes estimator δ_{π_0}, which will be seen to be minimax for various settings of (f, ρ) and which provides a benchmark in the sense that estimators $\delta_g \in C$ will be minimax for convex ρ under the simple condition that δ_g does not expand on δ_0 as much as δ_{π_0} [Theorem 1, (ii)]. In turn, for various choices of (f, ρ) with ρ even, and by appealing to Lemma 6, these classes of minimax estimators will contain the generalized Bayes estimators δ_{π_l}'s, $l > 0$. We now pursue with an intermediate dominance condition.

Theorem 1 *For estimating μ in (2) or (7) with $\mu \geq 0, \sigma > 0$, an estimator $\delta_g \in C$ is minimax, under strictly bowled-shaped loss $\rho(\frac{d-\mu}{\sigma})$ whenever either one of the following conditions holds for all $\lambda \geq 0$ and $y \in \{y : g'(y) < 0\}$:*

(i)

$$\int_0^\infty \int_{-\infty}^{vy-\lambda} \rho'(u + c_0 v + g(y)\, v) v^n f(u^2 + v^2) \mathrm{d}u \mathrm{d}v \ \leq 0,$$

or

(ii) *ρ is convex, $g \leq g_{\pi_0}$ and $\psi_\rho(\lambda, y) \leq 0$, where*

$$\psi_\rho(\lambda, y) = \int_0^\infty \int_{-\infty}^{vy-\lambda} \rho'(u + c_0 v + g_{\pi_0}(y)\, v) v^n f(u^2 + v^2) \mathrm{d}u \mathrm{d}v.$$

Proof With $\rho'(\cdot)$ increasing by the assumption of convexity, condition (ii) implies (i) so that we only need to establish the sufficiency of (i). Following Kubokawa (1994), write for $\delta_g(X, S) \in C$,

$$\begin{aligned}
\rho\left(\frac{x + c_0 s - \mu}{\sigma}\right) &- \rho\left(\frac{x + c_0 s + g(\frac{x}{s})s - \mu}{\sigma}\right) \\
&= \rho\left(\frac{x + c_0 s + g(y)s - \mu}{\sigma}\right)\Bigg|_{y=x/s}^{y=\infty} \\
&= \int_{x/s}^{\infty} \frac{s}{\sigma}\rho'\left(\frac{x + c_0 s + g(y)s - \mu}{\sigma}\right) g'(y)\mathrm{d}y.
\end{aligned}$$

Now, use the above expression for the difference in losses to write the difference in risks at $\theta = (\mu, \sigma)$ as:

$$\begin{aligned}
\Delta_g(\theta) &= R(\theta, \delta_0) - R(\theta, \delta_g) \\
&= \frac{1}{\sigma}\int_0^{\infty} s \int_{-\infty}^{\infty} \left\{\int_{x/s}^{\infty} g'(y)\rho'\left(\frac{x + c_0 s + g(y)s - \mu}{\sigma}\right) \mathrm{d}y\right\} f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s \\
&= \int_{\{g'(y)<0\}} g'(y) \left\{\int_0^{\infty} \int_{-\infty}^{sy} \rho'\left(\frac{x + c_0 s + g(y)s - \mu}{\sigma}\right)\right. \\
&\quad \left. \times \frac{s^n}{\sigma^{n+2}} f\left(\frac{(x-\mu)^2 + s^2}{\sigma^2}\right) \mathrm{d}x\mathrm{d}s\right\} \mathrm{d}y, \qquad (21)
\end{aligned}$$

since $g' \leq 0$ a.e. Now, the difference in risks $\Delta_g(\theta)$ will be non-negative for all $\theta \in \Theta$ as long as for all $y \in \mathbb{R}$ such that $g'(y) < 0$, $\mu \geq 0, \sigma > 0$, the bracketed term in (21) is less than or equal to 0, which is equivalent to **(i)** with the change of variables $(x, s) \to (u = \frac{x-\mu}{\sigma}, v = \frac{s}{\sigma})$. □

Remark 5 Notice that $\psi_\rho(0, y) = B_n(y, g_{\pi_0}(y)) = 0$ for all $y \in \mathbb{R}$ by virtue of the definition of g_{π_0} in (10). Therefore, the risks of δ_{π_0} and δ_0 match at the boundary of Θ where $\mu = 0$, $\sigma > 0$. Moreover, if δ_g expands more than δ_{π_0} (whether or not $\delta_g \in C$), then the risk at the boundary of δ_g will exceed that of δ_0, hence giving a condition for non-minimaxity. This is so given that

$$\begin{aligned}
R((0, \sigma), \delta_g) &= E_{(0,1)}(\rho(\delta_g(X, S))) \\
&> E_{(0,1)}(\rho(\delta_{\pi_0}(X, S))) \\
&= R((0, \sigma), \delta_{\pi_0}) \\
&= R((0, \sigma), \delta_0),
\end{aligned}$$

since $\delta_{\pi_0}(X, S) \geq 0$ with probability one, and ρ is increasing on $(0, \infty)$. As a consequence of the above, and of Lemma 6 and Corollary 1, we have the following non-minimaxity result.

Corollary 2 *For estimating μ in* (2) *or* (7) *with $\mu \geq 0$, $\sigma > 0$, the generalized Bayes estimators δ_{π_l} with $-(n-1) \leq l < 0$ are not minimax whenever* (a) *f is normal* *and ρ is even and convex, or whenever* (b) *f satisfies assumption* (3) *and the loss is invariant L^p with $p > 0$.*

Analogously, we point out that δ_{π_0} does not dominate any other minimax estimator $\delta_g \in C$ taking non-negative values and satisfying (ii) of Theorem 1, since such δ_g's shrink δ_{π_0} and $R((0,\sigma),\delta_g) = E_{(0,1)}(\rho(\delta_g(X,S))) < E_{(0,1)}(\rho(\delta_{\pi_0}(X,S))) = R((0,\sigma),\delta_{\pi_0})$.

Remark 6 A plausible alternative to the MRE estimator δ_0 is, of course, its truncation $\delta_0^T(X,S) = \max(0, \delta_0(X,S))$. Clearly δ_0^T improves upon δ_0 for bowl-shaped ρ, since for all $\mu \geq 0, \sigma > 0$, $\rho(\frac{\delta_0^T(x,s)-\mu}{\sigma}) \leq \rho(\frac{\delta_0(x,s)-\mu}{\sigma})$ for all $(x,s) \in \mathbb{R}\times\mathbb{R}^+$, with strict inequality occurring with positive probability. Moreover, the estimator δ_0^T belongs to the class C with $g_0^T(y) = \max(0, -y-c_0)$, and satisfies the condition (i) of Theorem 1 with $\{y : (g_0^T)'(y) < 0\} = (-\infty, -c_0)$ since

$$
\begin{aligned}
&\int_0^\infty \int_{-\infty}^{vy-\lambda} \rho'\left(u + c_0 v + g_0^T(y)v\right) v^n f(u^2+v^2)\mathrm{d}u\mathrm{d}v \\
&= \int_0^\infty \int_{-\infty}^{vy-\lambda} \rho'(u - vy)v^n f(u^2+v^2)\mathrm{d}u\mathrm{d}v \\
&\leq \int_0^\infty \int_{-\infty}^{vy-\lambda} \rho'(-\lambda)v^n f(u^2+v^2)\mathrm{d}u\mathrm{d}v \\
&\leq 0,
\end{aligned}
$$

for all $\lambda \geq 0$. Finally, the observations of Remark 5 apply to δ_0^T, with δ_0^T a shrinker of δ_{π_0}, and δ_{π_0} not dominating δ_0^T.

With Theorem 1, our attention focuses on the quantity $\psi_\rho(\lambda, y)$ and testing the condition $\psi_\rho(\cdot,\cdot) \leq 0$ on $\mathbb{R}^+ \times \mathbb{R}$ for various choices of ρ. Now, since

$$
\psi_\rho(0,y) = 0, \text{ and } \lim_{\lambda\to\infty} \psi_\rho(\lambda, y) = 0 \quad \text{for all } y \in \mathbb{R}, \tag{22}
$$

$\psi_\rho(\cdot, y)$ cannot be monotone on $[0,\infty)$ for any ρ and $y \in \mathbb{R}$. We are thus led to analyze the behavior of $\frac{\partial}{\partial\lambda}\psi_\rho(\lambda, y)$.

Proposition 1 *Let $k(y) = y + c_0 + g_{\pi_0}(y)$, $f_{\lambda,y}(t)$ be a Lebesgue density on* $(0,\infty)$ *proportional to*

$$
t^n f\left(\lambda^2(1+y^2)\left\{\left(t - \frac{y}{1+y^2}\right)^2 + \frac{1}{(1+y^2)^2}\right\}\right),
$$

and

$$
D_\rho(\lambda,y) = \int_0^{1/k(y)} |\rho'(\lambda(tk(y)-1))| f_{\lambda,y}(t)\mathrm{d}t - \int_{1/k(y)}^\infty |\rho'(\lambda(tk(y)-1))| f_{\lambda,y}(t)\mathrm{d}t. \tag{23}
$$

Suppose further that $D_\rho(\cdot, y)$ changes signs once from $-$ to $+$ on $[0, \infty)$ *for all* $y \in \mathbb{R}$. *Then, for estimating μ in* (2) *or* (7) *under assumption* (3) *with* $\mu \geq 0, \sigma > 0$,

(i) *the generalized Bayes estimator δ_{π_0} is minimax, for strictly bowled-shaped loss $\rho(\frac{d-\mu}{\sigma})$ as long as $\delta_{\pi_0} \in C$;*
(ii) *for $\delta_g \in C$, the condition $g \leq g_{\pi_0}$ is sufficient for δ_g to be minimax under convex loss $\rho(\frac{d-\mu}{\sigma})$.*

Proof We have

$$\begin{aligned} \frac{\partial}{\partial \lambda}\psi_\rho(\lambda, y) &= -\int_0^\infty \rho'(vk(y)-\lambda)v^n f((vy-\lambda)^2+v^2)\mathrm{d}v \\ &\propto -\lambda^{n+1}\int_0^\infty \rho'(\lambda t k(y)-\lambda) f_{\lambda,y}(t)\mathrm{d}t \qquad (24)\\ &\propto D_\rho(\lambda, y). \qquad (25)\end{aligned}$$

Therefore, under the given assumptions on the sign changes of $D_\rho(\cdot, y)$, we infer that, for all $y \in \mathbb{R}$, $\psi_\rho(\lambda, y)$ decreases, then increases as λ varies on $[0, \infty)$. Finally, the result follows from Theorem 1 and property (22). □

Remark 7 (i) From (24), note that

$$\frac{\partial}{\partial \lambda}\psi_\rho(\lambda, y)|_{\lambda=0^+} = -\int_0^\infty \rho'(vk(y))v^n f(v^2(y^2+1))\mathrm{d}v \leq 0,$$

since $k(\cdot) \geq 0$ from (12), and $\rho'(\cdot) \geq 0$ on $[0, \infty)$. Hence, Proposition 1's sign change assumption on $D_\rho(\cdot, y)$ is consistent, for any strictly bowled-shaped ρ, with the behavior of $\psi_\rho(\lambda, y)$ for λ near 0.

(ii) Turning to the families of densities $\{f_{\lambda,y}(\cdot), \lambda \in [0, \infty), y \in \mathbb{R}\}$, they can be shown for $y \leq 0$ to possess a decreasing monotone likelihood ratio (mlr) in T, or equivalently in $W = (T - \frac{y}{y^2+1})^2$, with λ viewed as the parameter. Indeed, for $\lambda_1 > \lambda_0 \geq 0$, setting $\alpha_i = \lambda_i^2(y^2+1)$ and $\epsilon = (y^2+1)^{-2}$, we have

$$\frac{f_{\lambda_1,y}(t)}{f_{\lambda_0,y}(t)} \propto \frac{f(\alpha_1(w+\epsilon))}{f(\alpha_0(w+\epsilon))}$$

which decreases in w, $w > \frac{y^2}{(y^2+1)^2}$, given assumption (3).

(iii) In the normal case, the densities $f_{\lambda,y}(\cdot)$ may be described as weighted (by the factor t^n) positively truncated $N(y/(1+y^2), 1/(\lambda^2(1+y^2)))$ densities. They have been recently studied in related work of Marchand et al. (2012) where quantiles are estimated under the restriction $\mu \geq 0$.

We conclude this section with a very useful technical result.

Lemma 7 *Let $k(y) = y + c_0 + g_{\pi_0}(y)$ as in Proposition 1 and let ρ be either an even function, or more generally satisfy $|\rho'(-u)| \leq |\rho'(u)|$ for all $u > 0$. Then we have* $\frac{1}{k(y)} > \max\{0, \frac{y}{1+y^2}\}$.

Proof The positivity of $k(y)$, $y \in \mathbb{R}$ follows from (12). To establish that $\frac{1}{k(y)} > \frac{y}{1+y^2}$, we assume the contrary and show that this would imply $D_\rho(\lambda, y) \leq 0$ for all $\lambda \geq 0$ which is not possible given (24) and (25). Indeed, we would have, for all $\lambda \geq 0$, $y \in \mathbb{R}$, under the given assumption on ρ

$$\begin{aligned}
D_\rho(\lambda, y) \leq & \int_0^{1/k(y)} \left|\rho'\left(\lambda k(y)\left(t - \frac{1}{k(y)}\right)\right)\right| f_{\lambda,y}(t)\mathrm{d}t \\
& - \int_{1/k(y)}^{2/k(y)} \left|\rho'\left(\lambda k(y)\left(t - \frac{1}{k(y)}\right)\right)\right| f_{\lambda,y}(t)\mathrm{d}t \\
\leq & \int_0^{1/k(y)} \left|\rho'\left(\lambda k(y)\left(t - \frac{1}{k(y)}\right)\right)\right| \left(f_{\lambda,y}(t) - f_{\lambda,y}\left(\frac{2}{k(y)} - t\right)\right) \mathrm{d}t \\
\leq & \; 0,
\end{aligned}$$

given that $f_{\lambda,y}(t) \leq f_{\lambda,y}(\frac{2}{k(y)} - t)$ for all $t \in (0, 1/k(y))$ whenever $\frac{1}{k(y)} \leq \frac{y}{1+y^2}$. □

The inequality $\frac{1}{k(y)} > \max\{0, \frac{y}{1+y^2}\}$ will be exploited as a technical result, but it also provides an interesting upper bound for the generalized Bayes estimator δ_{π_0}, namely

$$\delta_{\pi_0}(x, s) = sk\left(\frac{x}{s}\right) < x + \frac{s^2}{x}, \quad \text{for } x > 0,$$

applicable to all pairs (f, ρ) for which δ_{π_0} exists, with f satisfying (3), ρ satisfying the conditions of Lemma 7.

4 Minimax results for the normal case

Here is a minimax result applicable in the normal case, to Bayes estimators δ_{π_l}, and for general convex losses that are either even functions or, more generally, that penalize the rate of overestimation more sharply than the rate of underestimation in the sense

$$|\rho'(-u)| \leq \rho'(u), \quad \text{for all } u \geq 0. \tag{26}$$

Theorem 2 *For estimating μ in the normal case in (1) with $\mu \geq 0, \sigma > 0$ under convex ρ in (4),*

(a) *the condition $g \leq g_{\pi_0}$ suffices for an estimator $\delta_g \in C$ to be minimax in cases where ρ satisfies condition (26);*
(b) *such minimax estimators include the generalized Bayes estimator δ_{π_0} under losses ρ satisfying (26), and all δ_{π_l} with $l > 0$ when ρ is even.*

Proof Given that $\delta_{\pi_l} \in C$ for $l = 0$, and for $l > 0$ when ρ is even by virtue of Lemma 5, the first part of (b) is simply a restatement of (a) for the Bayes estimator δ_{π_0}, while the part relating to δ_{π_l} with $l > 0$ follows also from (a) and Lemma 6. The rest of the proof concerns part (a) and we apply Proposition 1. From (23), we have with the change of variables $u = \lambda(tk(y) - 1)$:

$$D_\rho(\lambda, y) = \frac{1}{\lambda k(y)} E[-\rho'(U)],$$

where U has density proportional to

$$f_{\lambda,y}\left(\left(\frac{u}{\lambda}+1\right)\frac{1}{k(y)}\right) 1_{(-\lambda,\infty)}(u). \tag{27}$$

Since $-\rho'$ changes signs once from $+$ to $-$ on $\mathbb{R}$, a decreasing in u monotone likelihood ratio property of the densities in (27) with respect to the parameter λ will suffice to establish that $D_\rho(\lambda, y)$ changes signs from $-$ to $+$ on $[0, \infty)$ as a function of $\lambda \geq 0$ and permit us to apply Proposition 1.[1] Now, the densities in (27) may be written as

$$h_\lambda(u) \propto \left(\frac{u+\lambda}{\lambda}\right)^n f\left(c(u+\lambda b)^2 + d\lambda^2\right) 1_{(-\lambda,\infty)}(u),$$

with $c = (1+y^2)/k^2(y)$, $b = 1 - \left(yk(y)/(1+y^2)\right)$, and $d = (1+y^2)^{-1}$. Notice that we have $c > 0$ by virtue of (12), and $b > 0$ by assumption (26) and Lemma 7. Finally, in the normal case with $f(t) = (2\pi)^{(n+1)/2} e^{-t/2}$, the ratio $\frac{h_{\lambda_1}(u)}{h_{\lambda_0}(u)}$ is, for $\lambda_1 > \lambda_0 \geq 0$, undetermined for $u \leq -\lambda_1$, equal to $+\infty$ for $u \in (-\lambda_1, -\lambda_0]$, and otherwise proportional to

$$\left(\frac{u+\lambda_1}{u+\lambda_0}\right)^n e^{-bcu(\lambda_1-\lambda_0)},$$

which is indeed decreasing in u for $u > -\lambda_0$, and which establishes the result. □

The normal case minimax results of Theorem 2 in part (a), and applicable to the generalized Bayes estimator δ_{π_0}, were previously obtained for the specific case of scale invariant L^2 loss by Kubokawa (2004). He works directly with the Bayes estimator in Example 1 to derive the key required analytical properties, namely the monotonicity of g_{π_0} in Lemma 5 and inequality (i) of Theorem 1. With Kubokawa's analysis specific to scale invariant L^2 loss, the normal model and the estimator δ_{π_0}, our unified development above contrasts and provides extensions with respect to the loss and the prior. In the next section, we give extensions with respect to the model for scale invariant L^p losses and asymmetric versions.

5 Minimax results for scale invariant L^p losses and their asymmetric versions

The minimax results of this section are applicable for the wider class of models, or choices of f, in (2) with assumptions (3). As well, these findings concern scale invariant L^p losses $|\frac{d-\mu}{\sigma}|^p$, $p > 0$, and the more general ρ_{c_1,c_2} in (13) with $c_2 \geq c_1$.

[1] It is interesting to point out that the arguments here apply as well to strictly bowled-shaped losses. As well, only a stochastic increasing property for the densities f is required. The monotonicity of g_{π_0}, however, is guaranteed by the convexity of ρ (Lemma 5), which is assumed here.

For these losses, (23) reduces to $D_\rho(\lambda, y) = p\lambda^{p-1} E_\lambda[g_y(T)]$, with $T \sim f_{\lambda,y}$ and $g_y(t) = c_1\,(1 - tk(y))^{p-1}\mathbb{I}_{(0,1/k(y))}(t) - c_2\,(tk(y) - 1)^{p-1}\mathbb{I}_{(1/k(y),\infty)}(t)$. With this representation, observe that $g_y(\cdot)$ changes sign once on $(0, \infty)$ from $+$ to $-$, so that $E_\lambda[g_y(T)]$ changes signs from $-$ to $+$ as λ varies on $[0, \infty)$, in view of sign change properties and the mlr property of Remark 7 (ii). Therefore, $D_\rho(\lambda, y)$ varies indeed, as a function of $\lambda \in [0, \infty)$ from $-$ to $+$ as prescribed in Proposition 1 for $y \le 0$ and losses ρ_{c_1,c_2}. For $y > 0$, however, the situation is more delicate. We continue with the non-convex case with $p \in (0, 1)$, and this will be followed by the convex case with $p \ge 1$.

Theorem 3 *For estimating μ in (2) or (7) with $\mu \ge 0, \sigma > 0$ under scale invariant L^p loss $|\frac{d-\mu}{\sigma}|^p$ with $p \in (0, 1)$, the generalized Bayes estimator δ_{π_0} is minimax.*

Proof With $\delta_{\pi_0} \in C$ by virtue of Lemma 5, we seek to apply part (i) of Proposition 1 to show that δ_{π_0} is minimax. For $\rho(t) = |t|^p$ with $p > 0$, we reexpress (23) as

$$D_\rho(\lambda, y) \propto E[A(T)\,B(T)] = E[G(S)],$$

with

$$G(s) = E[A(T)B(T)|S = s], \quad A(t) = t^n \left|t - \frac{1}{k(y)}\right|^{p-1}, \quad B(t) = -1 + 2\mathbb{I}_{(0,\frac{1}{k(y)}]}(t),$$

$$T \sim f\left(\lambda^2(1 + y^2)\left\{\left(t - \frac{y}{1 + y^2}\right)^2 + \frac{1}{(1 + y^2)^2}\right\}\right),$$

and

$$S \overset{d}{=} \left(T - \frac{y}{1 + y^2}\right)^2 + \frac{1}{(1 + y^2)^2}.$$

Given assumption (3), the family of densities of S are seen to have a decreasing monotone likelihood ratio in S with parameter $\lambda^2(1 + y^2)$. Therefore, in accordance with Karlin's sign change analysis, to prove the result, it will suffice to show that

$$G(s) \text{ changes signs once from } + \text{ to } - \text{ as } s \text{ varies on } \left(\frac{1}{(1 + y^2)^2}, \infty\right) \tag{28}$$

to establish that $D_\rho(\lambda, y)$ changes signs as prescribed by Proposition 1. We proceed by treating separately the cases: (i) $\frac{1}{k(y)} - \frac{y}{1+y^2} \ge \frac{y}{1+y^2}$ and (ii) $0 \le \frac{1}{k(y)} - \frac{y}{1+y^2} \le \frac{y}{1+y^2}$. Here, we have made use of Lemma 7 to discount the remaining possibility $\frac{1}{k(y)} - \frac{y}{1+y^2} < 0$.

Case (i): Set $s_0 = (\frac{1}{k(y)} - \frac{y}{1+y^2})^2 + \frac{1}{(1+y^2)^2}$. Observe that, whenever $s \ge s_0$, $P(T \ge \frac{1}{k(y)}|S = s) = 1$ implying $P(B(T) = -1|S = s) = 1$ and $G(s) \le 0$. Similarly, if $s < s_0$, then $P(B(T) = 1|S = s) = 1$ and $G(s) \ge 0$. Hence, the above establishes (28) for case (i).

Case (ii): Here, we set $s_1 = \frac{y^2}{(1+y^2)^2} + \frac{1}{1+y^2}$, so that $s_0 \le s_1$. As in (i), we verify that $G(s) \ge 0$ for $s \le s_0$, and $G(s) \le 0$ for $s \ge s_1$. Finally, for $s \in (s_0, s_1)$, the conditional distribution of $T|S = s$ is a two-point uniform discrete distribution on $\{t_1, t_2\}$, with $t_1 = \frac{y}{1+y^2} + \Delta$, $t_2 = \frac{y}{1+y^2} - \Delta$ and $\Delta = \sqrt{s - \frac{1}{(1+y^2)^2}}$.

We hence obtain

$$G(s) = \frac{1}{2}(A(t_2) - A(t_1))$$
$$= \frac{1}{2}\left[t_2^n \left|t_2 - \frac{1}{k(y)}\right|^{p-1} - t_1^n \left|t_1 - \frac{1}{k(y)}\right|^{p-1}\right] < 0,$$

since $t_2 < t_1$, $n > 1$; $|t_1 - \frac{1}{k(y)}| = \frac{y}{1+y^2} - \frac{1}{k(y)} + \Delta < -\frac{y}{1+y^2} + \frac{1}{k(y)} + \Delta = |t_2 - \frac{1}{k(y)}|$, and $p-1 < 0$. Hence, the above establishes (28) for case (ii) and completes the proof. □

Theorem 4 *For estimating μ in (2) or (7) under assumptions (3), with $\mu \ge 0, \sigma > 0$ and with loss ρ_{c_1,c_2}, $p \ge 1$ and $c_2 \ge c_1$,*

(a) *the condition $g \le g_{\pi_0}$ suffices for an estimator $\delta_g \in C$ to be minimax;*
(b) *such minimax estimators include the generalized Bayes estimator δ_{π_0}, as well as all generalized Bayes estimators δ_{π_l}, $l > 0$ for the symmetric case $c_1 = c_2$.*

Proof For losses ρ as in (13), we may write

$$D_\rho(\lambda, y) = p\lambda^{p-1}\left\{c_1 \int_0^{\frac{1}{k(y)}} h_{\lambda,y}(w)\mathrm{d}w - c_2 \int_{\frac{1}{k(y)}}^{\infty} h_{\lambda,y}(w)\mathrm{d}w\right\},$$

with $h_{\lambda,y}(\cdot)$ a probability density function on $(0,\infty)$ proportional to $|wk(y) - 1|^{p-1} f_{\lambda,y}(w)$. From this, we see that $D_\rho(\lambda, y)$ is positive iff $P_\lambda(W > 1/k(y)) < c_1/(c_1+c_2)$, where W is a random variable with pdf $h_{\lambda,y}$. We show below in Sect. 7.4 of the Appendix that, whenever $\frac{1}{k(y)} > \frac{y}{1+y^2}$, the quantity $P_\lambda(W > 1/k(y))$ decreases in λ on $[0,\infty)$, which means that $D_\rho(\cdot, y)$ changes signs from $-$ to $+$ on $[0,\infty)$. The result then follows from Proposition 1 and Lemma 7. □

6 Concluding remarks

We have considered the problem of estimating a lower bounded location parameter for a wide array of spherically symmetric location-scale models with a residual vector as represented in model (2), with unknown scale, and under scale invariant loss as given by (4). With a relative paucity of findings for such problems when the scale parameter is unknown, we have established the minimaxity of the generalized Bayes estimator δ_{π_0} for normal models and convex loss, as well for more general models and scale invariant L^p loss and asymmetric versions ρ_{c_1,c_2} given in (13). Moreover, we have shown the role of δ_{π_0} to be pivotal, in the sense that it provides an upper threshold condition necessary for the minimaxity of many estimators. Other minimax estimators are also obtained, including generalized Bayes estimators δ_{π_l} when $l > 0$ and the loss is convex and even in the above situations. The results represent extensions of Kubokawa's results (2004) applicable to scale invariant L^2 loss. Much of the treatment is unified and exploits general features of the model and the loss with incisive analysis and novel representations. Various other observations are given, including the robustness of the Bayes estimator δ_{π_l} with respect to the choice of f in model (2).

As illustrated by Marchand et al. (2012), the normal case improvements provided for scale invariant L^2 loss yield applications for two-sample problems where $Y_i \sim N(\mu_i, \sigma^2); i = 1, 2$ with unknown μ_1, μ_2, σ^2, where the objective is to estimate μ_1 (or μ_2) with the additional information of the ordering $\mu_1 \leq \mu_2$. Despite the advances presented here, minimax extensions to other strictly bowled-shaped losses, although plausible, are still lacking. Furthermore, numerous questions remain unanswered such as the admissibility of the above minimax estimators, the investigation of wider classes of Bayes estimators for minimaxity, and related tests of minimaxity for multivariate location-scale problems with order restrictions.

7 Appendix

7.1 Proof of Lemma 1

We have

$$\frac{t\, f'(t)}{f(t)} = \frac{\int_0^\infty t v^2\, f_0'(tv)\, h(v) \mathrm{d}v}{\int_0^\infty v\, f_0(tv)\, h(v) \mathrm{d}v} = \frac{\int_0^\infty z^2\, f_0'(z)\, h(z/t) \mathrm{d}z}{\int_0^\infty z\, f_0(z)\, h(z/t) \mathrm{d}z} = E_t\left[\frac{Z\, f_0'(Z)}{f_0(Z)}\right], \tag{29}$$

where Z has density proportional to $z\, f_0'(z)\, h(z/t)$ on $\mathbb{R}^+$. Now, observe that this scale family of densities for Z have increasing monotone likelihood ratio in Z, with parameter t, as a consequence of assumption (3) for h. Finally, the result follows from representation (29) with this monotone likelihood ratio and since $\frac{z f_0'(z)}{f_0(z)}$ decreases in z by assumption (3) for f_0. □

7.2 Proof of Lemma 6

We fix $y \in \mathbb{R}$ throughout and set $c_0 = 0$ given that ρ is assumed even. First, observe that by differentiating (10) for the normal case with $f(u^2 + v^2) = \phi(u)h(v)$ in (8), we have $\frac{\partial}{\partial l} B_{n+l}(y, g_{\pi_l}(y) = 0$ which implies

$$\int_0^\infty \left[\frac{\partial}{\partial l}\left\{\int_{-\infty}^{vy} \rho'(u + g_{\pi_l}(y)v)\phi(u)\mathrm{d}u\right\}\right.$$
$$\left. + \left\{\int_{-\infty}^{vy} \rho'(u + g_{\pi_l}(y)v)\phi(u)\mathrm{d}u\right\}\log(v)\right] v^{n+l}h(v)\mathrm{d}v = 0.$$

Hence, given that ρ' is increasing, to show that $g_{\pi_l}(y)$ decreases in l, it will suffice to show that $I \geq 0$, where

$$I = \int_0^\infty (\log v)A_y(v)v^{n+l}h(v)\mathrm{d}v, \quad \text{and} \quad A_y(v) = \int_{-\infty}^{vy} \rho'(u + g_{\pi_l}(y)v)\phi(u)\mathrm{d}u. \tag{30}$$

Now, we will show below that

$$A_y(v) \text{ changes signs once as a function of } v \text{ from} - \text{ to } +. \tag{31}$$

Applying Lemma 8, which is stated in the Appendix, with $\xi \sim \xi^{n+l}\, h(\xi)\, 1_{(0,\infty)}(\xi)$, $r(\xi) = \log(\xi)$, and $s(\xi) = A_y(\xi)$, we infer that $I \geq 0$, since $E[A_y(\xi)] = 0$ given the definition of $g_{\pi_l}(y)$ in (10). There remains to establish (31), which we proceed to do separating the cases: (i) $y \leq 0$ and (ii) $y > 0$.

(i) Case $y < 0$. Let v_0 be such that $A_y(v_0) = 0$. Such a value exists since the average value of $A_y(\xi)$ under the above density for ξ is equal to 0. For $\epsilon \geq 0$, we have

$$\begin{aligned}\frac{A_y(v_0+\epsilon)}{\Phi((v_0+\epsilon)y)} &= \int_0^{(v_0+\epsilon)y} \rho'(u - \epsilon y + \epsilon(y + g_{\pi_l}(y)) + g_{\pi_l}(y)v_0)\frac{\phi(u)}{\Phi((v_0+\epsilon)y)}\mathrm{d}u \\ &\geq \int_0^{v_0 y} \rho'(u' + g_{\pi_l}(y)v_0)\frac{\phi(u'+\epsilon y)}{\Phi((v_0+\epsilon)y)}\mathrm{d}u' \\ &= C_y(v_0, \epsilon) \quad \text{(say)},\end{aligned}$$

with equality if and only if $\epsilon = 0$, given (12) and since ρ' is increasing. Now, observe that the ratio of densities $\frac{(\phi(u'+\epsilon y)/\Phi((v_0+\epsilon)y))}{\phi(u')/\Phi(v_0+y)}$ is increasing in u', for $u' \in (-\infty, v_0 y)$ and $\epsilon y < 0$. Hence, this monotone likelihood ratio property implies that $\frac{A_y(v_0+\epsilon)}{\Phi((v_0+\epsilon)y)} \geq C_y(v_0, \epsilon) \geq C_y(v_0, 0) = 0$, for $\epsilon > 0$ and $y < 0$, yielding (31) for $y < 0$.

(ii) Case $y \geq 0$. As in (i), let v_0 be such that $A_y(v_0) = 0$. Using this, as well as property (12), the non-negativity of $g_{\pi_l}(\cdot)$ (Lemma 5) (since $c_0 = 0$), and the convexity of ρ, we have for $\epsilon > 0$:

$$\begin{aligned}A_y(v_0+\epsilon) &= \int_{-\infty}^{(v_0+\epsilon)y} \rho'\{u + (g_{\pi_l}(y))(v_0+\epsilon)\}\phi(u)\mathrm{d}u \\ &\geq A_y(v_0) + \int_{v_0 y}^{(v_0+\epsilon)y} \rho'\{u + (g_{\pi_l}(y))v_0\}\phi(u)\mathrm{d}u \\ &\geq 0.\end{aligned}$$

$\square$

7.3 Lemma used within the proof of Lemma 6

The following result is well known and its proof is left to the reader.

Lemma 8 *Let ξ be a continuous random variable, let $r(\cdot)$ be a continuous and increasing function on the support of ξ, and let $s(\cdot)$ be a continuous function which changes signs once from $-$ to $+$ at s_0 on the support of ξ. Then, we have $E[\,r(\xi)\,s(\xi)\,] \geq r(s_0)\,E[s(\xi)]$, and, in particular if $E[s(\xi)] = 0$, then $E[\,r(\xi)\,s(\xi)\,] \geq 0$.*

7.4 Theorem 4: proof of a monotonicity property for $P_\lambda(W > 1/k(y))$

We wish to show that

$$P_\lambda\left(W > \frac{1}{k(y)}\right) \quad \text{decreases in } \alpha \text{ whenever } \frac{1}{k(y)} > a, \tag{32}$$

with W having pdf $h_{\lambda,y}(w)$ on $(0,\infty)$ proportional to $|wk(y) - 1|^{p-1}\,w^n\,f(\alpha\{(w-a)^2 + \epsilon\}$, $a = \frac{y}{1+y^2}$, $\alpha = \lambda^2(1+y^2)$, and $\epsilon = (y^2+1)^{-2}$. We have

$$\begin{aligned}
&\frac{\partial}{\partial\lambda} P_\lambda\left(W > \frac{1}{k(y)}\right) = 2\lambda(1+y^2)\frac{\partial}{\partial\alpha}\frac{\int_{\frac{1}{k(y)}}^{\infty} |wk(y)-1|^{p-1}\,w^n\,f\left(\alpha\left\{(w-a)^2+\epsilon\right\}\right)\mathrm{d}w}{\int_0^\infty |wk(y)-1|^{p-1}w^n f\left(\alpha\left\{(w-a)^2+\epsilon\right\}\right)\mathrm{d}w} \leq 0 \\
&\iff E\left[\gamma\left(\alpha\left\{(W-a)^2+\epsilon\right\}\right)\middle| W > \frac{1}{k(y)}\right] \geq E\left[\gamma\left(\alpha\left\{(W-a)^2+\epsilon\right\}\right)\right],
\end{aligned} \tag{33}$$

under pdf $h_{\lambda,y}$, with $\gamma(t) = t\frac{|f'(t)|}{f(t)}$. Taken together, the following hence form a sufficient condition for (33) to hold:

(i) $E\left[\gamma\left(\alpha\left\{(W-a)^2+\epsilon\right\}\right)\middle| W > \frac{1}{k(y)}\right] \geq E\left[\gamma\left(\alpha\left\{(W-a)^2+\epsilon\right\}\right)\middle| W < a\right]$, and

(ii) $E\left[\gamma\left(\alpha\left\{(W-a)^2+\epsilon\right\}\right)\middle| W > \frac{1}{k(y)}\right] \geq E\left[\gamma\left(\alpha\left\{(W-a)^2+\epsilon\right\}\right)\middle| a \leq W < \frac{1}{k(y)}\right]$.

Condition (ii) is immediate, since $\gamma(\alpha\{(W-a)^2+\epsilon\})$ increases in $(W-a)^2$ on (a,∞) by assumption (3). For (i), set $Z = |W-a|$ so that $Z|W > 1/k(y)$ has pdf proportional to $|(a+z)k(y)-1|^{p-1}\,(a+z)^n\,f(\alpha(z^2+\epsilon))\,\mathbb{I}_{(1/k(y)-a,\infty)}(z)$, while $Z|W < a$ has pdf proportional to $|1-(a-z)k(y)|^{p-1}\,(a-z)^n\,f(\alpha(z^2+\epsilon))\,\mathbb{I}_{(0,a)}(z)$. We thus have the ratio

$$\frac{f_{Z|W<a}(z)}{f_{Z|W>\frac{1}{k(y)}}(z)} \propto \begin{cases} \infty & \text{if } z < a,\, z < \frac{1}{k(y)} - a; \\ 0 & \text{if } z \geq a,\, z > \frac{1}{k(y)} - a; \\ \left(\frac{a-z}{a+z}\right)^n \left(\frac{zk(y)+(1-ak(y))}{zk(y)-(1-ak(y))}\right)^{p-1} & \text{if } 0 < z < a,\, z > \frac{1}{k(y)} - a. \end{cases}$$

Since both $\frac{a-z}{a+z}$ and $\frac{zk(y)+b}{zk(y)-b}$ decrease in z for $z < a$ and $z > 1/k(y) - a$, with $b = 1 - ak(y) > 0$ (Lemma 7), we have a decreasing monotone likelihood ratio.

Finally, with $\gamma(\alpha(z^2+\epsilon))$ increasing in $z>0$ by (3), condition (i) follows and our proof of (32) is complete. $\square$

Acknowledgments This work was partially supported by a grant from the Simons Foundation (#209035) to William Strawderman. Mohammad Jafari Jozani and Éric Marchand gratefully acknowledge the research support of the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank referees for careful reading of the paper and useful suggestions.

References

Farrell, R. H. (1964). Estimators of a location parameter in the absolutely continuous case. *Annals of Mathematical Statistics*, *35*, 949–998.

Fourdrinier, D., Strawderman, W. E. (2010). *Robust generalized Bayes minimax estimators of location vectors for spherically symmetric distribution with unknown scale. Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown* (pp. 249–262). Beachwood: Institute of Mathematical Statistics.

Katz, M. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Annals of Mathematical Statistics*, *32*, 136–142.

Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics*, *28*, 573–601.

Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Annals of Statistics*, *22*, 290–299.

Kubokawa, T. (2004). Minimaxity in estimation of restricted parameters. *Journal of the Japanese Statistical Society*, *34*, 1–19.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses*. Springer, 2nd edn. Wiley, New York. Transferred to Wadsworth & Brooks/Cole (1991).

Marchand, É., Strawderman, W.E. (2005). On improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, *57*, 129–143.

Marchand, É., Strawderman, W.E. (2006). *On the behaviour of Bayesian credible intervals for some restricted parameter space problems. Recent developments in nonparametric inference and probability: a Festschrift for Michael Woodroofe, IMS Lecture Notes-Monograph Series, 50* (pp. 112–126). Beachwood: Institute of Mathematical Statistics.

Marchand, É., Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, *18*, 635–643.

Marchand, É., Strawderman, W.E. (2013). On Bayesian credible sets in restricted parameter space problems and lower bounds for frequentist coverage. *Electronic Journal of Statistics*, *7*, 1419–1431.

Marchand, É., Strawderman, W.E., Bosa, K., Lmoudden, A. (2008). On the frequentist coverage of Bayesian credible intervals for lower bounded means. *Electronic Journal of Statistics*, *2*, 1028–1042.

Marchand, É., Jafari Jozani, M., Tripathi, Y. (2012). *Inadmissible estimators of normal quantiles and two-sample problems with additional information. Contemporary developments in Bayesian analysis and statistical decision theory: a Festschrift for William E. Strawderman, IMS Collections, 8, 104–116*. Beachwood: Institute of Mathematical Statistics.

Maruyama, Y. (2003). A robust generalized Bayes estimator improving on the James–Stein estimator for spherically symmetric distributions. *Statistics & Decisions*, *21*, 69–77.

Maruyama, Y., Iwasaki, K. (2005). Sensitivity of minimaxity and admissibility in the estimation of a positive normal mean. *Annals of the Institute of Statistical Mathematics*, *57*, 145–156.

Maruyama, Y., Strawderman, W.E. (2005). A new class of generalized Bayes minimax ridge regression estimators. *Annals of Statistics*, *33*, 1753–1770.

Sacks, J. (1963). Generalized Bayes solutions in estimation problems. *Annals of Mathematical Statistics*, *34*, 751–768.

Zhang, T., Woodroofe, M. (2003). Credible and confidence sets for restricted parameter spaces. *Journal of Statistical Planning and Inference*, *115*, 479–490.