

SHRINKAGE EFFICIENCY BOUNDS*

BY BRUCE E. HANSEN

University of Wisconsin

We consider estimation of a multivariate normal mean under sum of squared error loss. We construct the efficiency bound (the lowest achievable risk) for minimax shrinkage estimation. This allows us to compare the regret of existing shrinkage estimators. We also construct a new shrinkage estimator which achieves substantially lower maximum regret than existing estimators.

1. Introduction. Let $X \in \mathbb{R}^p$ be a single observation from a multivariate normal distribution with unknown mean vector $\theta \in \mathbb{R}^p$ and known covariance matrix \mathbf{I}_p , that is, $X \sim N(\theta, \mathbf{I}_p)$. The goal is to estimate θ . Consider the class of orthogonally invariate estimators which shrink X towards the zero vector and can be written as

$$(1.1) \quad \delta_\phi(X) = \left(1 - \frac{\phi(\|X\|^2)}{\|X\|^2}\right) X$$

where $\phi : [0, \infty) \rightarrow [0, \infty)$.

Under quadratic loss, the risk of an estimator $\delta_\phi(X)$ is

$$(1.2) \quad R_p(\psi, \delta_\phi) = E \|\delta_\phi(X) - \theta\|^2$$

where $\psi = \|\theta\|^2$. We write the risk as a function of the scalar ψ as it is well known (e.g., equation (2.1) below) that the risk only depends on θ through ψ and p .

When $p > 2$ the usual estimator X is no longer the unique minimax estimator. Sufficient conditions on ϕ for the estimator (1.1) to be minimax have been developed by Baranchik (1970), Strawderman (1971), Alam (1973), Stein (1973), Berger (1976a), Berger (1976b), Efron and Morris (1976), Faith (1978), Stein (1981), DasGupta and Strawderman (1997) and Fourdrinier, Strawderman and Wells (1998).

Many minimax estimators in the class (1.1) have been proposed. The most famous are the James-Stein estimator (James and Stein, 1961) and its

*October, 2009. Research partially supported by NSF SES-0550908.

AMS 2000 subject classifications: 62J07, 62C20, 62F10

Keywords and phrases: shrinkage, James-Stein, minimax, risk, regret, efficiency bound

positive-part version (Baranchik, 1964). Others include those of Strawderman (1971), Alam (1973), Berger (1976a), Li and Kuo (1982), Kubokawa (1991), Guo and Pal (1992), Shao and Strawderman (1994), Maruyama (1998), Kuriki and Takemura (2000), Maruyama (2004) and Maruyama (2007). It is difficult, however, to rank these estimators. For example, the only estimator known to dominate the positive-part estimator is that of Shao and Strawderman (1994).

While the risk of the estimator class (1.1) has been widely studied, there has been no investigation of *efficiency bounds* (the lowest possible risk) with the exception of Magnus (2002) who only considers the case $p = 1$, and thereby could not investigate minimax estimators. Our paper provides a sharp efficiency bound for $p \geq 3$, by taking the infimum of the risk $R_p(\psi, \delta_\phi)$ across the class of minimax estimators defined by the necessary and sufficient conditions of Stein (1973) and Efron and Morris (1976).

Given the efficiency bound, we define and evaluate the regret and maximum regret (MaxRegret) of common minimax shrinkage estimators. The regret of an estimator is the difference between its risk and the efficiency bound, and the MaxRegret is the maximum of the regret across the parameter space. For each estimator, we select its shrinkage parameter(s) to minimize the MaxRegret, a choice which can greatly improve estimator efficiency. We can also construct the smallest possible MaxRegret across the space of minimax shrinkage estimators, and we find that this value is substantially smaller than the MaxRegret of existing shrinkage estimators. Finally, we construct a new simple shrinkage estimator which has substantially smaller MaxRegret than existing shrinkage estimators.

2. Risk. We start with a simple yet insightful new representation for the risk (1.2).

THEOREM 2.1. *If ϕ is Borel measurable,*

$$(2.1) \quad R_p(\psi, \delta_\phi) = R_p^0(\psi) + D_p(\psi, \phi),$$

where

$$R_p^0(\psi) = p - \int_0^\infty \phi_p^*(q, \psi)^2 q^{-1} f_p(q, \psi) dq,$$

$$(2.2) \quad D_p(\psi, \phi) = \int_0^\infty (\phi(q) - \phi_p^*(q, \psi))^2 q^{-1} f_p(q, \psi) dq,$$

$$(2.3) \quad \phi_p^*(q, \psi) = q - \psi \frac{f_{p+2}(q, \psi)}{f_p(q, \psi)},$$

and $f_k(q, \psi)$ is the density of $\chi_k^2(\psi)$, a non-central chi-square random variable with k degrees of freedom and non-centrality parameter ψ .

PROOF. For any ϕ ,

$$(2.4) \quad \|\delta_\phi(X) - \theta\|^2 = \|X - \theta\|^2 + \frac{\phi^2(\|X\|^2)}{\|X\|^2} - 2\phi(\|X\|^2) + 2\theta'X \frac{\phi(\|X\|^2)}{\|X\|^2}.$$

Taking expectations of (2.4), using the fact that for any function $g(q)$, $E(Xg(\|X\|^2)) = \theta E(g(\chi_{p+2}^2(\psi)))$, (see, for example, Bock, 1975, Theorem A), we obtain

$$(2.5) \quad R_p(\psi, \delta_\phi) = p + \int_0^\infty (\phi^2(q) - 2q\phi(q)) q^{-1} f_p(q, \psi) dq \\ + 2\psi \int_0^\infty \phi(q) q^{-1} f_{p+2}(q, \psi) dq.$$

Using definition (2.3) and completing the square, this equals

$$p + \int_0^\infty (\phi^2(q) - 2\phi(q)\phi_p^*(q, \psi)) q^{-1} f_p(q, \psi) dq \\ = p - \int_0^\infty \phi_p^*(q, \psi)^2 q^{-1} f_p(q, \psi) dq \\ + \int_0^\infty (\phi(q) - \phi_p^*(q, \psi))^2 q^{-1} f_p(q, \psi) dq \\ = R_p^0(\psi) + D_p(\psi, \phi).$$

□

Theorem 2.1 decomposes the risk into two components: $R_p^0(\psi)$ which depends only on p and ψ , and $D_p(\psi, \phi)$, which can be written as a weighted distance between ϕ and ϕ_p^* . We now describe some critical features of the related function $h_p^*(q, \psi) = \phi_p^*(q, \psi)/q$.

LEMMA 2.1. $h_p^*(q, \psi) = \phi_p^*(q, \psi)/q$ is continuous and monotonically increasing in $q \geq 0$, $h_p^*(0, \psi) = 1 - \psi/p$ and $\lim_{q \rightarrow \infty} h_p^*(q, \psi) = 1$.

See the Appendix for the proof.

On a side note, one useful implication of Theorem 2.1 and Lemma 2.1 is a simple proof that positive-part trimming always reduces risk. Suppose that we have an estimator δ_ϕ where $\phi(q) > q$ for some $q > 0$. Then define a positive-part version $\phi^+ = \min(\phi(q), q)$ and estimator $\delta_\phi^+ = \delta_{\phi^+}$.

LEMMA 2.2. *If $P(\phi(\|X\|^2) > \|X\|^2) > 0$ then $R_p(\psi, \delta_\phi^+) < R_p(\psi, \delta_\phi)$.*

PROOF. Lemma 2.1 shows that $h^*(q, \psi) \leq 1$ so $\phi^*(q, \psi) \leq q$. It follows that on the set $Q = \{q : \phi(q) > q\}$, $(\phi(q) - \phi^*(q, \psi))^2 > (\phi^+(q) - \phi^*(q, \psi))^2$. Examining (2.2), we see that $D_p(\psi, \phi) > D_p(\psi, \phi^*)$, and the inequality is strict since $P(Q) > 0$ by assumption. From (2.1) we find $R_p(\psi, \delta_\phi^+) < R_p(\psi, \delta_\phi)$ as required. \square

Contrast Lemma 2.2 with Theorem 5.4 of Lehmann and Casella (1998) which provides the same inequality under the additional restriction that $h(q) = \phi(q)/q$ is strictly decreasing in q . Lemma 2.2 shows that this restriction is unnecessary.

3. Efficiency Bounds. An efficiency bound is the infimum of $R_p(\psi, \delta_\phi)$ across ϕ in some function class. We are focused on minimax estimators and thus restrict attention for the remainder of the paper to $p \geq 3$. For the minimaxity of $\delta_\phi(X)$, Stein (1973) and Efron and Morris (1976) developed the general conditions

$$(3.1) \quad 4q\phi'(q) + \phi(q)(2(p-2) - \phi(q)) \geq 0$$

and

$$(3.2) \quad 0 \leq \phi(q) \leq 2(p-2) \quad \text{for all } q \geq 0.$$

This condition is broader than those of Baranchik (1970), Strawderman (1971) and Alam (1973), and in particular allows $\phi(q)$ to be nonmonotonic.

Let Φ be the set of functions satisfying (3.1)-(3.2). The efficiency bound for minimax shrinkage estimators is the infimum of $R_p(\psi, \delta_\phi)$ across $\phi \in \Phi$:

$$(3.3) \quad R_p(\psi) = \inf_{\phi \in \Phi} R_p(\psi, \delta_\phi).$$

From decomposition (2.1) we see that minimizing $D_p(\psi, \phi)$ is equivalent to minimizing $R_p(\psi, \delta_\phi)$. As the former is a weighted average of the squared distance between $\phi(q)$ and $\phi_p^*(q, \psi)$, it is minimized by setting $\phi(q)$ as close as possible to $\phi_p^*(q, \psi)$. Suppose for the moment that we only impose condition (3.2). In this case the minimizer of $R_p(\psi, \delta_\phi)$ is the truncated function

$$(3.4) \quad \begin{aligned} \phi_p^{**}(q, \psi) &= \phi_p^*(q, \psi) \mathbb{I}(0 \leq \phi_p^*(q, \psi) \leq 2(p-2)) \\ &+ 2(p-2) \mathbb{I}(\phi_p^*(q, \psi) > 2(p-2)). \end{aligned}$$

Since the function $h_p^*(q, \psi)$ is continuous in q (Lemma 2.1), then so are $\phi_p^*(q, \psi)$ and $\phi_p^{**}(q, \psi)$. Furthermore, since $h_p^*(q, \psi)$ is monotonically increasing in q (Lemma 2.1) then $\phi_p^{**}(q, \psi)$ is monotonically increasing in q on the set $\phi_p^{**}(q, \psi) \geq 0$. It follows that $\phi_p^{**}(q, \psi)$ is continuous and non-decreasing in $q \geq 0$, and therefore satisfies condition (3.1) in addition to (3.2). It follows that $\phi_p^{**}(q, \psi)$ is the minimizer of $R_p(\psi, \delta_\phi)$ over $\phi \in \Phi$, and thus (1.1) with $\phi(q) = \phi_p^{**}(q, \psi)$ is the pointwise (in ψ) optimal minimax estimator for θ .

From (2.1) we see that the minimized value (the efficiency bound) is

$$(3.5) \quad R_p(\psi) = R_p^0(\psi) + D_p(\psi, \phi_p^{**}).$$

Our main result gives a simplified expression for the efficiency bound.

THEOREM 3.1. *For $p \geq 3$, the efficiency bound for estimation of θ by (1.1) with $\phi \in \Phi$ is*

$$(3.6) \quad \begin{aligned} R_p(\psi) = & p - 8(p-2) f_p(q_2(\psi), \psi) \\ & - \int_{q_1(\psi)}^{q_2(\psi)} \phi_p^*(q, \psi)^2 q^{-1} f_p(q, \psi) dq \end{aligned}$$

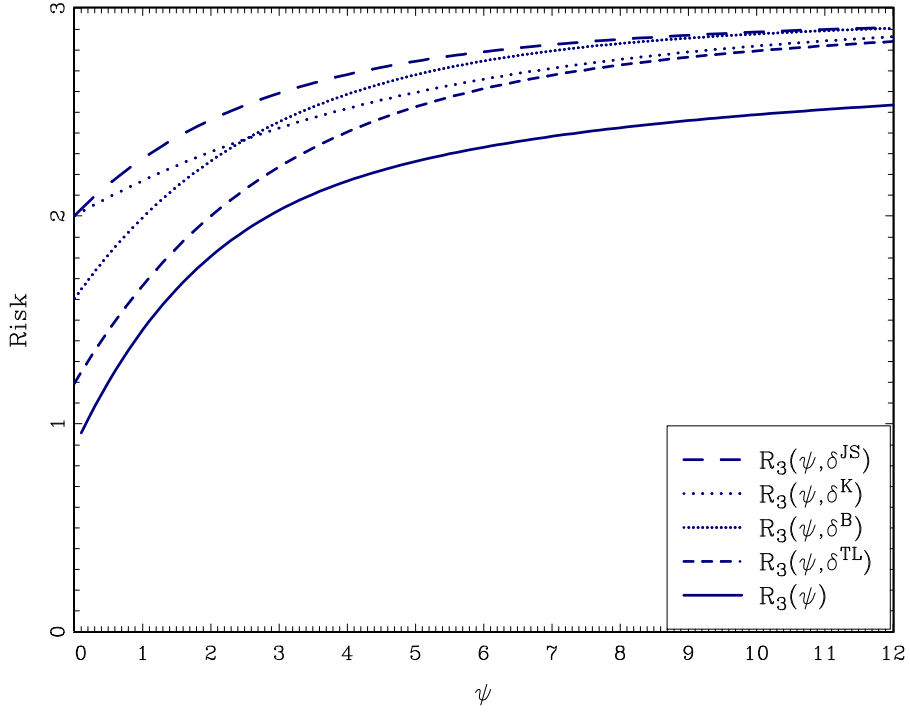
where $q_1(\psi)$ is the smallest $q \geq 0$ such that $h_p^*(q, \psi) \geq 0$ and $q_2(\psi)$ is the unique $q > 0$ where $\phi_p^*(q, \psi) = 2(p-2)$.

See the Appendix for the proof.

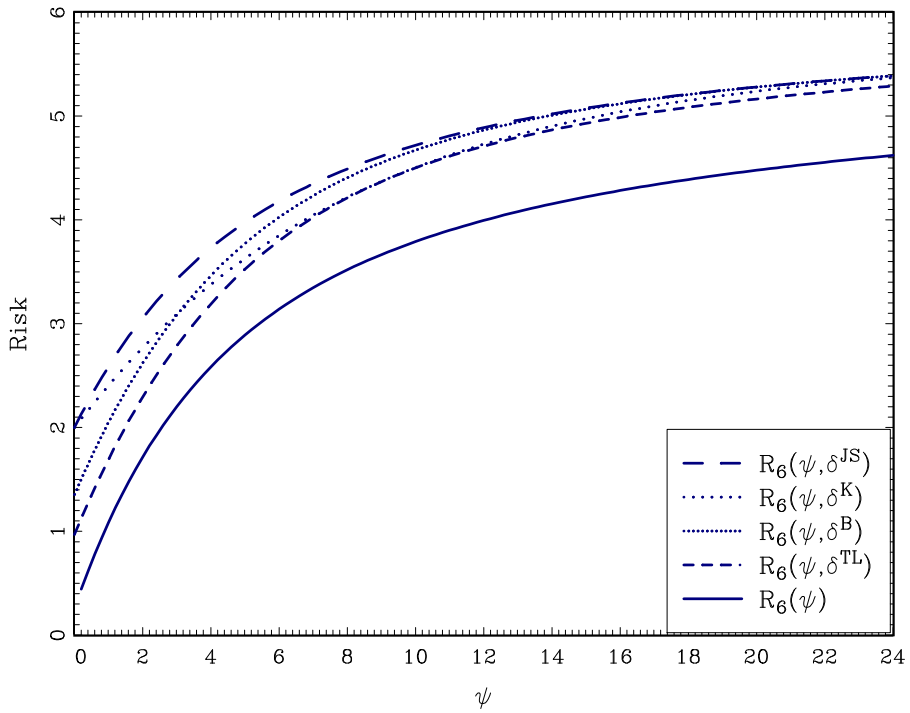
While the efficiency bound is the lower envelope of the risk functions of all minimax shrinkage estimators, there is no single estimator whose risk function equals the efficiency bound.

We display¹ the bound $R_p(\psi)$ for $p = 3$ and $p = 6$ in Figure 1. For reference we also plot the risk of the James-Stein estimator δ^{JS} , Kubokawa's estimator δ^K , Baranchik's positive-part estimator δ^B , and a new estimator δ^{TL} which will be introduced in Section 6. Observe that the efficiency bound $R_p(\psi)$ is monotonically increasing in ψ , asymptotes towards the minimax bound p , and is strictly less than the risk of any of the individual estimators displayed. We also see the well-known properties that Baranchik's estimator uniformly dominates the James-Stein estimator, that Kubokawa's estimator uniformly dominates the James-Stein estimator everywhere except at $\psi = 0$ (where their risks equal), and that neither Baranchik's nor Kubokawa's estimator uniformly dominates the other. Plots for other values of p are qualitatively similar.

¹For details on this and other numerical calculations see the numerical appendix.



(a) $p = 3$



(b) $p = 6$

FIG 1. Risk and Efficiency Bounds

4. Regret. The regret of an estimator $\delta_\phi(X)$ is the difference between its risk and the efficiency bound:

$$\text{Regret}_p(\psi, \delta_\phi) = R_p(\psi, \delta_\phi) - R_p(\psi).$$

The regret is the cost due to the use of the estimator δ_ϕ instead of the infeasible pointwise optimal estimator, and varies with ψ . For illustration, Figure 2 plots the Regret functions of the four estimators from Figure 1 for $p = 3$ and $p = 6$.

One measure of the uniform performance of an estimator is the MaxRegret – the maximum of the regret function over ψ :

$$\text{MaxRegret}_p(\delta_\phi) = \sup_{\psi \geq 0} \text{Regret}_p(\psi, \delta_\phi).$$

An estimator with low MaxRegret has the desirable property that its risk is uniformly close to the infeasible efficiency bound. As the MaxRegret is free of unknowns it can be (numerically) calculated, and used to compare and rank estimators.

We will now compare the MaxRegret of eleven minimax shrinkage estimators. For each, we give the shrinkage function ϕ and the known parameter restrictions which are sufficient for minimaxality.

1. James and Stein (1961)

$$\phi^{JS}(q) = p - 2$$

2. Baranchik (1964)

$$\phi^B(q) = \min(q, p - 2)$$

3. Li and Kuo (1982)

$$\phi^{LK}(q) = p - 2 - c_1 q^{-\alpha_1/2}$$

where

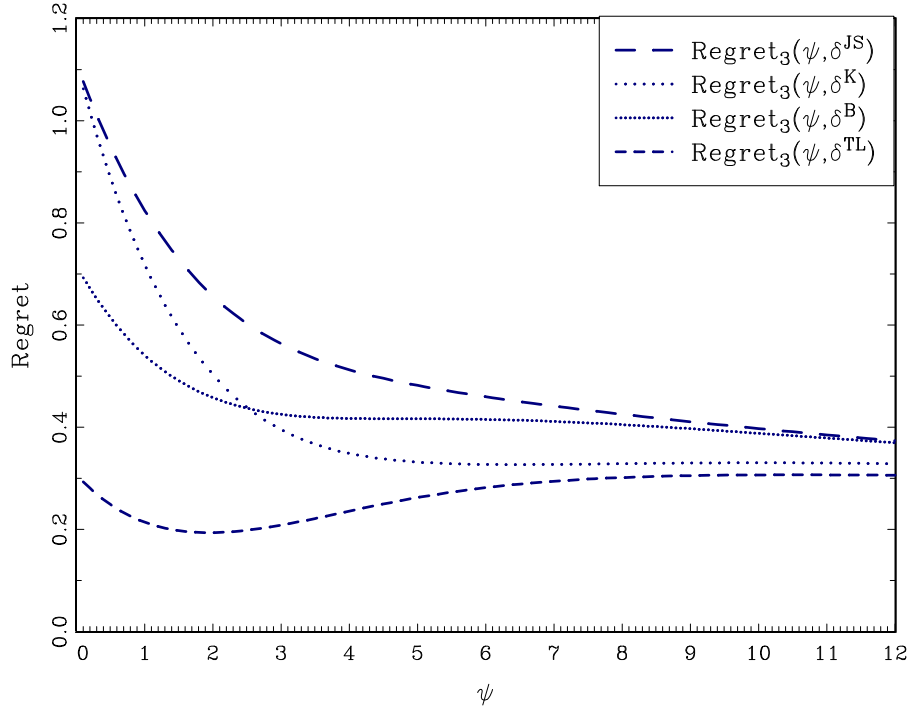
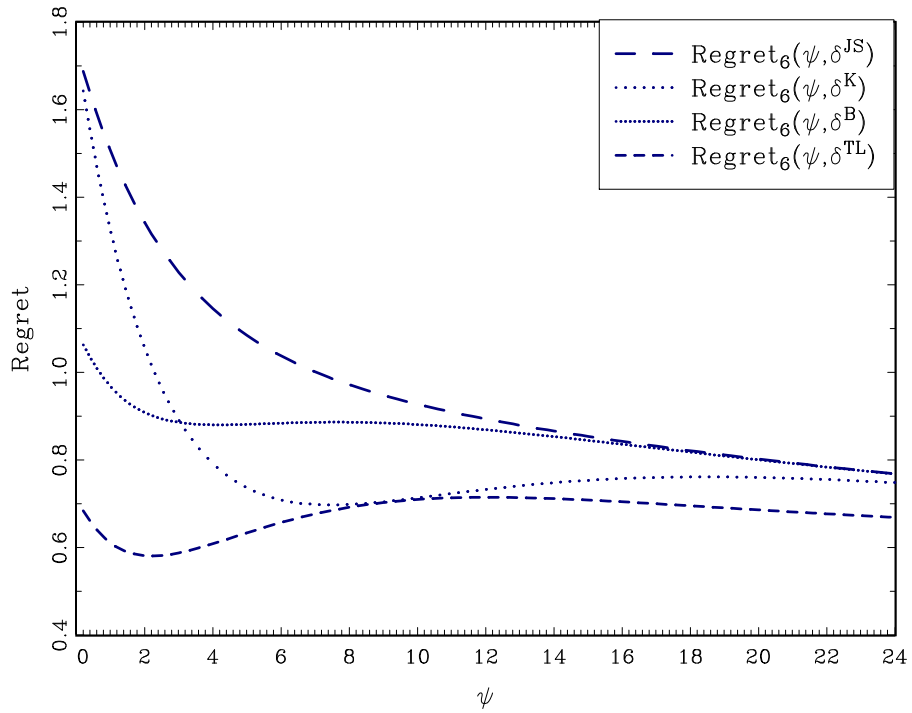
$$c_1 = \alpha_1 2^{\alpha_1/2} \frac{\Gamma(p/2 - (1 + \alpha_1/2))}{\Gamma(p/2 - (1 + \alpha_1))}$$

with $0 < \alpha_1 < p/2 - 1$.

4. Guo and Pal (1992)

$$\phi^{GP}(q) = p - 2 - \sum_{j=1}^n c_j q^{-\alpha_j/2}$$

with $0 < \alpha_1 < \dots < \alpha_n < p/2 - 1$. Guo and Pal (1992) give explicit expressions for the constants c_j .

(a) $p = 3$ (b) $p = 6$ FIG 2. *Regret*

5. Shao and Strawderman (1994)

$$\phi^{SS}(q) = \min(q, p-2) + ag(|q|)$$

where

$$g(t) = \begin{cases} 2\tau - 1 + t, & 0 \leq t \leq \tau \\ t - 1, & \tau < t \leq 1 \\ 0, & 1 < t < \infty. \end{cases}$$

Values for a and τ such that δ^{SS} dominates δ^B are given in Shao and Strawderman (1994, Table 1).

6. Kuriki and Takemura (2000)

$$\phi^{KT}(q) = \begin{cases} p-2 - \frac{r}{q^{1/2}-r}, & \text{if } q \geq \left(\frac{p-1}{p-2}\right)^2 r^2 \\ 0, & \text{if } q < \left(\frac{p-1}{p-2}\right)^2 r^2 \end{cases}$$

with $r \geq 0$.

7. Strawderman (1971), Alam (1973), Kubokawa (1991), Maruyama (1998) and Maruyama (2004)

$$\phi^M(q) = q \frac{B(b+1, p/2-a+2) {}_1F_1(b+1; p/2-a+b+3; q/2) + \beta}{B(b+1, p/2-a+1) {}_1F_1(b+1; p/2-a+b+2; q/2) + \beta}$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function and ${}_1F_1(a, b, x)$ is the confluent hypergeometric function, with

$$\begin{aligned} 3 - p/2 &\leq a \leq 1 + p/2, \\ b &\geq -\frac{a + p/2 - 3}{3p/2 + 1 - a}, \end{aligned}$$

and

$$\begin{aligned} 0 \leq \beta &\leq B(p/2 - a + 1, b - a + p/2 + 2) \\ &\times \left(-1 + \sqrt{1 + \frac{D(a, b)(p/2 - a + 1)(b + 1)}{(b - a + p/2 + 2)(b - a + p/2 + 3)}} \right), \end{aligned}$$

where

$$D(a, b) = \begin{cases} a + p/2 - 3, & \text{if } b \geq 0 \\ \frac{b(3p/2 + 1 - a) + (a + p/2 - 3)}{b + 1}, & \text{if } b < 0. \end{cases}$$

TABLE 1
Minimizing MaxRegret Parameters for Shrinkage Estimators

Estimator	Parameter	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
Strawderman (1971)	a	1.50	1.09	0.95	0.86	0.79	0.74	0.69	0.66
Alam (1973)	ν	0.93	0.89	0.87	0.85	0.84	0.83	0.82	0.81
Li and Kuo (1982)	a_1	0.38	0.74	1.09	1.41	1.72	2.00	2.27	2.52
Guo and Pal (1992)	a_1	0.38	0.75	1.11	1.44	1.75	2.04	2.32	2.58
	a_2	0.48	0.96	1.44	1.91	2.37	2.83	3.29	3.73
Kuriki and Takemura (2000)	r	0.08	0.27	0.50	0.72	0.93	1.06	1.12	1.14
Maruyama (1998)	a	1.50	1.10	1.04	1.02	1.02	1.04	1.06	1.05
	b	0.00	-0.01	-0.07	-0.11	-0.14	-0.17	-0.19	-0.19
Maruyama (2004)	a	1.50	1.10	0.98	0.95	0.99	0.98	1.02	1.01
	b	0.00	-0.01	0.00	-0.04	-0.11	-0.12	-0.16	-0.16
	β	0.000	0.000	0.015	0.015	0.009	0.012	0.008	0.010

Unrestricted, this is a shrinkage function due to Maruyama (2004). The others are special cases, and all set $\beta = 0$. Maruyama (1998) leaves a and b free. Kubokawa (1991) sets $a = 2$ and $b = 0$. Strawderman (1971) sets $b = 0$ and leaves a free. Alam (1973) sets $b = \nu - 1$ and $a = \nu + 1$ and leaves ν free.

For the estimators which depend on parameters (all of those described above except those of James and Stein (1961), Baranchik (1964) and Kubokawa (1991)), we select the estimator's parameters to minimize the MaxRegret, searching among the set of parameters satisfying the minimax condition. The minimizing parameter values are given in Table 1 for $p = 3, \dots, 10$. By using these parameter values, we obtain the smallest possible MaxRegret for each estimator.

MaxRegret $_p$ for the eleven estimators is reported in Table 2, with the estimators listed in rough order of descending MaxRegret. We can see that the least efficient estimators are those of James-Stein and Kubowawa. While Kubokawa's estimator weakly dominates the James-Stein estimator, the two estimators have equal MaxRegret since both regret functions are maximized at $\psi = 0$ where they have equal risk.

Based on the MaxRegret criteria, the estimators of Li and Kuo (1982), Guo and Pal (1992), Alam (1973), and Kuriki and Takemura (2000) are less efficient than the simple positive-part estimator of Baranchik (1964). The improvement of the Shao-Strawderman estimator over Baranchik's is quantitatively negligible. More substantial reductions in MaxRegret relative to Baranchik's estimator are obtained by the estimators of Strawderman (1971), Maruyama (1998), and Maruyama (2004), but these three estimators are essentially equivalent. Since the Strawderman estimator includes

TABLE 2
MaxRegret

Estimator	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
James and Stein (1961)	1.113	1.459	1.637	1.744	1.814	1.861	1.895	1.920
Kubokawa (1991)	1.113	1.459	1.637	1.744	1.814	1.861	1.895	1.920
Li and Kuo (1982)	1.026	1.299	1.418	1.479	1.511	1.528	1.537	1.540
Guo and Pal (1992)	1.026	1.298	1.416	1.476	1.507	1.523	1.531	1.533
Alam (1973)	0.973	1.239	1.377	1.444	1.494	1.521	1.535	1.540
Kuriki and Takemura (2000)	0.795	1.104	1.278	1.379	1.437	1.485	1.548	1.624
Baranchik (1964)	0.716	0.930	1.037	1.097	1.133	1.156	1.170	1.178
Shao and Strawderman (1994)	0.715	0.930	1.037	1.097	1.133	1.156	1.170	1.178
Strawderman (1971)	0.501	0.609	0.772	0.886	0.969	1.032	1.082	1.121
Maruyama (1998)	0.501	0.609	0.768	0.878	0.957	1.017	1.062	1.100
Maruyama (2004)	0.501	0.609	0.768	0.878	0.957	1.016	1.062	1.098
Trimmed Linear Shrinkage	0.308	0.483	0.614	0.714	0.790	0.852	0.903	0.945
MinRegret _{p}	0.297	0.476	0.600	0.693	0.765	0.824	0.873	0.907

Kubokawa's as a special case, we can see the importance of optimal parameter selection. For example, for $p = 3$, Kubokawa's estimator ($a = 2$) has $\text{MaxRegret}_p = 1.113$ while with the optimal choice of $a = 1.5$ then $\text{MaxRegret}_p = 0.501$.

5. Minimum MaxRegret. In the previous section we calculated the MaxRegret for a number of minimax shrinkage estimators. While the lowest MaxRegret was achieved by the estimator of Maruyama (2004), we might wonder if further substantive reductions are possible. To answer this question we define the smallest possible MaxRegret among minimax shrinkage estimators:

$$(5.1) \quad \text{MinRegret}_p = \inf_{\phi \in \Phi} \text{MaxRegret}_p(\delta_\phi).$$

This is the lower bound for the MaxRegret among minimax shrinkage estimators.

While a closed form expression for the MinRegret is not available, we can numerically approximate its value by minimizing the MaxRegret over a dense class of approximating shrinkage functions. It turns out to be convenient to use the class of continuous linear splines. We also tried the class of quadratic splines, but found no quantitative improvement.

For some positive integer N and knots $0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_N < \infty$, let $\theta = \{a_1, b_1, \dots, a_N, b_N, c\}$ denote a set of parameters. Our spline function

takes the form

$$(5.2) \quad \phi_\theta(q) = \sum_{n=1}^N (a_n + b_n q) \mathbb{I}(\tau_{n-1} \leq q \leq \tau_n) + c \mathbb{I}(q > \tau_N)$$

subject to the continuity constraints

$$(5.3) \quad a_1 = 0$$

$$(5.4) \quad a_n + b_n \tau_n = a_{n+1} + b_{n+1} \tau_n \quad \text{for } 1 \leq n \leq N-1$$

$$(5.5) \quad a_N + b_N \tau_N = c.$$

By picking N and τ_N sufficiently large, the class (5.2) can approximate any function in Φ .

Let $\delta_\theta = \delta_{\phi_\theta}$. To calculate the risk of this estimator, we recall the well-known result (e.g. Maruyama, 2004, equation (4.1)) that the risk of δ_ϕ for any absolutely continuous ϕ is

$$R_p(\psi, \delta_h) = p + \int_0^\infty (q^{-1} \phi(q)(\phi(q) - 2(p-2)) - 4\phi'(q)) f_p(q, \psi) dq.$$

Applying this formula to the continuous linear spline (5.2), it is straightforward to calculate the risk of δ_θ .

LEMMA 5.1. *The risk of the estimator (1.1) with the shrinkage function (5.2)-(5.5) is*

$$\begin{aligned} R_p(\psi, \delta_\theta) = & p + \sum_{n=1}^N \left\{ a_n (a_n - 2(p-2)) \int_{\tau_{n-1}}^{\tau_n} q^{-1} f_p(q, \psi) dq \right. \\ & \left. + 2b_n (a_n - p) \int_{\tau_{n-1}}^{\tau_n} f_p(q, \psi) dq + b_n^2 \int_{\tau_{n-1}}^{\tau_n} q f_p(q, \psi) dq \right\} \\ & - c(2(p-2) - c) \int_{\tau_N}^\infty q^{-1} f_p(q, \psi) dq. \end{aligned}$$

For numerical computation it is convenient to write the integrals in the above expression as convergent infinite sums. Using the infinite series definition of the non-central chi-square density (see equation (A.1) in the Appendix) and integrating term-by-term, it is straightforward to obtain the following result.

LEMMA 5.2. For $s > -p/2$,

$$\int_0^\tau q^s f_p(q, \psi) dq = e^{-\psi/2} 2^s \sum_{j=0}^{\infty} \frac{(\psi/2)^j}{j!} \frac{\gamma\left(s + \frac{p}{2} + j, \frac{\tau}{2}\right)}{\Gamma\left(\frac{p}{2} + j\right)}$$

where $\gamma(x, a) = \int_0^a e^{-q} q^{x-1} dq$ is the incomplete gamma function.

Combining Lemmas 5.1 and 5.2, we find that $R_p(\psi, \delta_\theta)$ is computationally simple to evaluate even when N is large

For ϕ_θ to lie in Φ it must satisfy (3.1)-(3.2), which is equivalent to the following conditions on the parameters in (5.2). For $n = 1, \dots, N$,

$$(5.6) \quad 2(p-2)a_n - a_n^2 + 2b_n(p-a_n)\tau_{n-1} - b_n^2\tau_{n-1}^2 \geq 0$$

$$(5.7) \quad 2(p-2)a_n - a_n^2 + 2b_n(p-a_n)\tau_n - b_n^2\tau_n^2 \geq 0$$

$$(5.8) \quad 0 \leq a_n + b_n\tau_n \leq 2(p-2).$$

Let Θ_N be the set of parameters θ which satisfy (5.3)-(5.5) and (5.6)-(5.8). The smallest MaxRegret among all minimax shrinkage estimators using the spline function (5.2) is then

$$\text{MinRegret}_p^N = \inf_{\theta \in \Theta_N} \sup_{\psi \geq 0} (R_p(\psi, \delta_\theta) - R_p(\psi)).$$

Since the class of functions (5.2) is dense in the class Φ , MinRegret_p^N is a good approximation to MinRegret_p when N and τ_N are large.

Setting $N = 50$ and $\tau_N = 30p$, we numerically calculated MinRegret_p^N by searching over θ using a constrained BFGS algorithm. The result is printed in the bottom row of Table 2. We find that MinRegret is indeed substantially smaller than the MaxRegret of the estimators thus far considered, especially for small p . For example, for $p = 3$, the estimators of Strawderman (1971), Maruyama (1998) and Maruyama (2004) achieve a MaxRegret of 0.501, while the MinRegret is 41% lower at 0.297.

While the 50-knot linear spline ϕ_θ just introduced could be used for a shrinkage estimator, this is a inelegant choice and thus not recommended. Instead, in the next section we recommend a simple approximation which achieves near-equivalent efficiency.

6. Trimmed Linear Shrinkage. In this section, we introduce a simple shrinkage function and estimator which is nearly equivalent to the 50-knot

MinRegret estimator of the previous section, and is recommended for practical application. For parameters a and b satisfying $0 \leq a \leq 2(p-2)$ and $0 \leq b \leq 1$, let

$$\phi^{TL}(q) = \min[q, a + bq, 2(p-2)].$$

This function is a special case of (5.2) with $N = 2$, $a_1 = 0$, $b_1 = 1$, $a_2 = a$, $b_2 = b$, $\tau_1 = a/(1-b)$, $\tau_2 = (c-a)/b$, and $c = 2(p-2)$, and its risk is given in Lemma 5.1. It may be also convenient to observe that the estimator can be written as

$$\delta^{TL}(X) = \begin{cases} 0, & \|X\|^2 \leq \tau_1 \\ \left(1 - b - \frac{a}{\|X\|^2}\right) X, & \tau_1 < \|X\|^2 \leq \tau_2 \\ \left(1 - \frac{2(p-2)}{\|X\|^2}\right) X, & \|X\|^2 > \tau_2. \end{cases}$$

By construction, $\phi^{TL}(q)$ satisfies the minimax conditions (3.1)-(3.2), so the estimator $\delta^{TL} = \delta_{\phi^{TL}}$ is minimax for all parameter values.

We selected the parameters a and b to minimize the estimator's MaxRegret, and report the optimal parameter values in Table 3 for $p = 3, \dots, 25$. The MaxRegret of the estimator $\delta^{TL} = \delta_{\phi^{TL}}$ is reported in Table 2 in the second line from the bottom. The Risk and Regret functions for the estimator are displayed in Figures 1 and 2, respectively.

In Table 2, we see that the MaxRegret of the estimator δ^{TL} is nearly equivalent to the MinRegret lower bound, and is substantially lower than that of the existing shrinkage estimators.

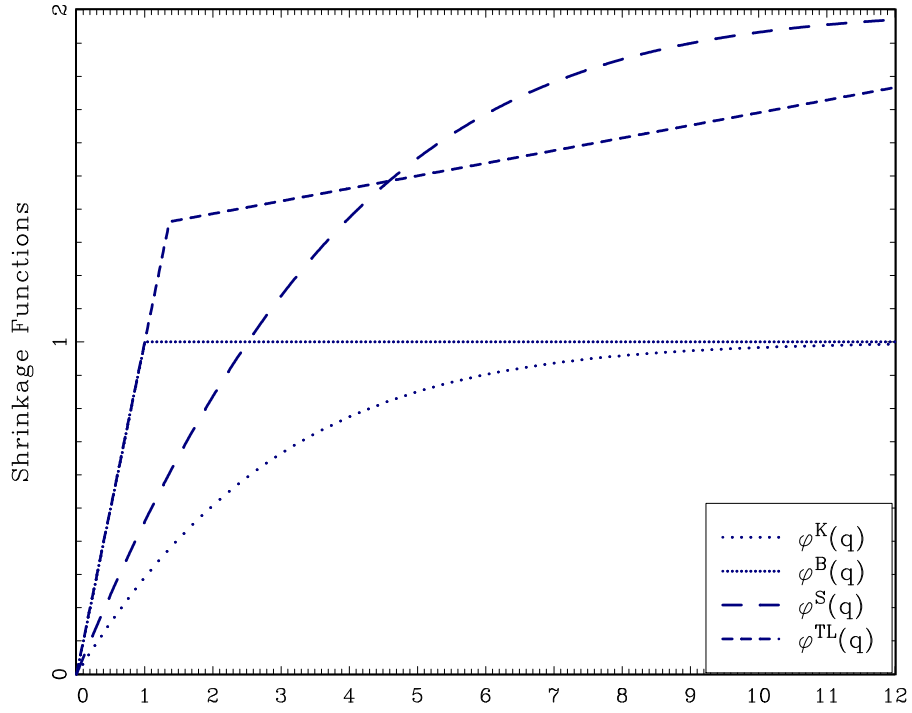
In Figures 1 and 2, we see that the risk and regret of δ^{TL} are uniformly smaller than the regret of the other estimators shown for the range of ψ displayed, and that for some values of ψ the difference is substantial. It is not the case, however, that δ^{TL} uniformly dominates these estimators for all ψ . For example the risk of δ^B is lower than that of δ^{TL} for $\psi \geq 21.6$ for $p = 3$ and $\psi \geq 43.2$ for $p = 6$. But these are extremely large values of ψ where the risk functions are extremely close to one another. Our interpretation is that for practical purposes δ^{TL} has substantially lower risk than δ^B and δ^K for most values of ψ .

Based on this analysis, for practical applications we recommend the estimator δ^{TL} with the parameter values from Table 3.

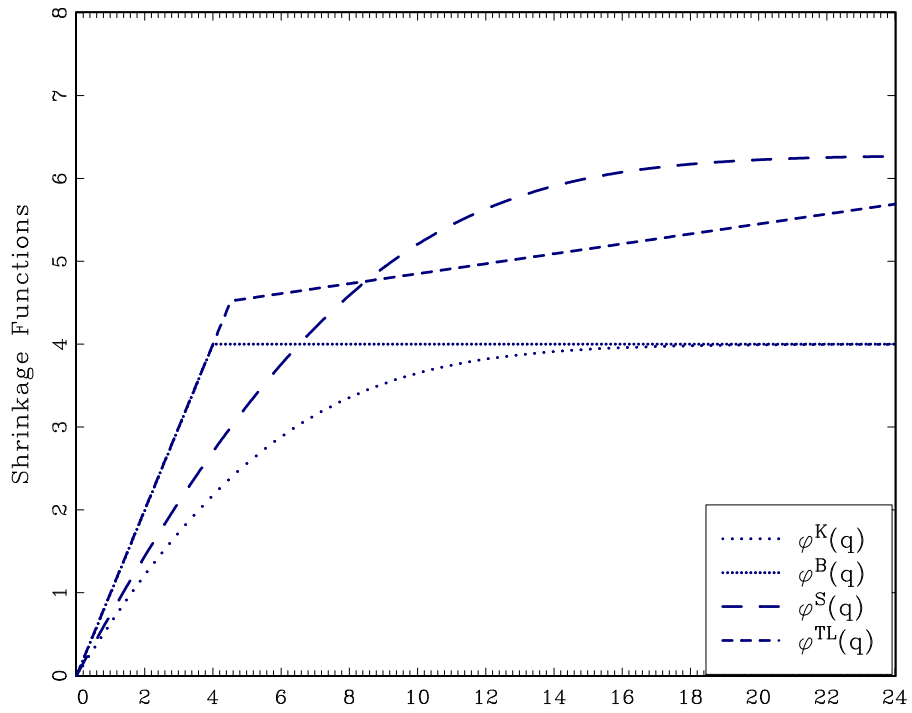
As a final comparison of the shrinkage estimators, in Figure 3 we display the Kubokawa, Baranchik, Strawderman, and Trimmed Linear shrinkage functions ϕ^K , ϕ^B , ϕ^S , and ϕ^{TL} , (using the optimal parameters) for $p = 3$ and $p = 6$.

TABLE 3
Trimmed Linear Shrinkage Parameters

	a	b	τ_1	τ_2
$p = 3$	1.315	0.038	1.36	18
$p = 4$	2.356	0.056	2.50	29
$p = 5$	3.315	0.061	3.52	44
$p = 6$	4.245	0.060	4.52	62
$p = 7$	5.170	0.058	5.49	83
$p = 8$	6.091	0.055	6.44	108
$p = 9$	7.010	0.052	7.39	135
$p = 10$	7.927	0.049	8.33	166
$p = 11$	8.844	0.046	9.27	199
$p = 12$	9.760	0.044	10.2	235
$p = 13$	10.677	0.041	11.1	274
$p = 14$	11.594	0.039	12.1	315
$p = 15$	12.518	0.037	13.0	360
$p = 16$	13.486	0.036	14.0	406
$p = 17$	14.447	0.034	15.0	455
$p = 18$	15.475	0.033	16.0	507
$p = 19$	16.496	0.031	17.0	562
$p = 20$	17.502	0.030	18.0	620
$p = 21$	18.470	0.029	19.0	681
$p = 22$	19.503	0.028	20.1	744
$p = 23$	20.492	0.027	21.1	811
$p = 24$	21.574	0.025	22.1	886
$p = 25$	22.667	0.024	23.2	964



(a) $p = 3$



(b) $p = 6$

FIG 3. *Shrinkage Functions*

APPENDIX

PROOF OF LEMMA 2.1. The density of the non-central chi-square can be written as

$$(A.1) \quad f_k(q, \psi) = e^{-(q+\psi)/2} 2^{-k/2} q^{k/2-1} \sum_{j=0}^{\infty} \frac{\left(\frac{\psi q}{4}\right)^j}{j! \Gamma\left(\frac{k}{2} + j\right)}$$

We can thus write $h_p^*(q, \psi) = 1 - (\psi/2)g(\psi q/4)$ where

$$(A.2) \quad g(x) = \frac{\sum_{j=0}^{\infty} a_j x^j}{\sum_{j=0}^{\infty} b_j x^j},$$

$a_j = 1/(j! \Gamma(1 + p/2))$ and $b_j = 1/(j! \Gamma(p/2))$. Since a_j/b_j is monotonically decreasing in j , then as discussed by Lehmann and Romano (2005, p. 308), $g(x)$ is monotonically decreasing in x , so $h_p^*(q, \psi)$ is monotonically increasing in q .

It is also easy to see that $g(0) = a_0/b_0 = 2/p$ so $h^*(0, \psi) = 1 - \psi/p$.

An alternative expression for the non-central chi-square density is

$$f_k(q, \psi) = \frac{1}{2} e^{-(q+\psi)/2} \left(\frac{q}{\psi}\right)^{k/4-1/2} I_{k/2-1}(\sqrt{q\psi})$$

where $I_m(x)$ is the modified Bessel function. Thus we can write

$$\begin{aligned} h_p^*(q, \psi) &= 1 - \psi \frac{f_{p+2}(q, \psi)}{q f_p(q, \psi)} \\ &= 1 - \sqrt{\frac{\psi}{q}} \frac{I_{p/2}(\sqrt{q\psi})}{I_{p/2-1}(\sqrt{q\psi})}. \end{aligned}$$

The Bessel function satisfies the large-argument expansion (see Magnus, Oberhettinger and Soni, 1966, p. 139)

$$I_\alpha(x) \sim (2\pi x)^{-1/2} e^x (1 + O(x^{-1})),$$

implying $I_{p/2}(x)/I_{p/2-1}(x) = 1 + O(x^{-1})$ and thus $h_p^*(q, \psi) \rightarrow 1$ as $q \rightarrow \infty$. \square

PROOF OF THEOREM 3.1. To simplify notation we omit the ψ argument from the functions and write $\phi_p^*(q) = \phi^*(q, \psi)$, $q_1 = q_1(\psi)$, $q_2 = q_2(\psi)$, $f_p(q) = f_p(q, \psi)$, and so on.

It is convenient to observe that we can write (3.4) as

$$\phi_p^{**}(q) = \begin{cases} 0, & q < q_1 \\ \phi_p^*(q), & q_1 \leq q \leq q_2 \\ 2(p-2), & q > q_2. \end{cases}$$

The minimized risk is

$$\begin{aligned} R_p(\psi) &= R_p^0(\psi) + D_p(\psi, \phi_p^{**}) \\ &= p - \int_0^\infty \phi_p^*(q)^2 q^{-1} f_p(q) dq + \int_0^\infty (\phi_p^{**}(q) - \phi_p^*(q))^2 q^{-1} f_p(q) dq \\ &= p - \int_{q_1}^\infty \phi_p^*(q)^2 q^{-1} f_p(q) dq \\ &\quad + \int_{q_2}^\infty (2(p-2) - \phi_p^*(q))^2 q^{-1} f_p(q) dq \\ &= p - \int_{q_1}^{q_2} \phi_p^*(q)^2 q^{-1} f_p(q) dq + 4(p-2)^2 \int_{q_2}^\infty q^{-1} f_p(q) dq \\ &\quad - 4(p-2) \int_{q_2}^\infty \phi_p^*(q) q^{-1} f_p(q) dq \\ &= p - \int_{q_1}^{q_2} \phi_p^*(q)^2 q^{-1} f_p(q) dq \\ &\quad + 4(p-2) \left((p-2) \int_{q_2}^\infty q^{-1} f_p(q) dq + \psi \int_{q_2}^\infty q^{-1} f_{p+2}(q) dq - \int_{q_2}^\infty f_p(q) dq \right). \end{aligned}$$

Using (A.1), it is straightforward to check that

$$f_p'(q) = -\frac{1}{2} f_p(q) + \left(\frac{p-2}{2} \right) q^{-1} f_p(q) + \frac{\psi}{2} q^{-1} f_{p+2}(q).$$

Integrating from q_2 to ∞ and multiplying by 2, it follows that

$$(p-2) \int_{q_2}^\infty q^{-1} f_p(q) dq + \psi \int_{q_2}^\infty q^{-1} f_{p+2}(q) dq - \int_{q_2}^\infty f_p(q) dq = -2f_p(q_2).$$

Substituting, we obtain

$$R_p(\psi) = p - \int_{q_1}^{q_2} \phi_p^*(q)^2 q^{-1} f_p(q) dq - 8(p-2)f_p(q_2)$$

as stated. \square

NUMERICAL APPENDIX

The risk of the estimators of James and Stein (1961), Baranchik (1964), Li and Kuo (1982), Guo and Pal (1992), Shao and Strawderman (1994), and the Trimmed Linear estimator of Section 6 can be expressed as linear functions of moments of the $\chi_k^2(\psi)$ distribution, and the latter can be written as convergent series using Lemma 5.2. For example,

$$\begin{aligned} R_p(\psi, \delta^{JS}) &= p - (p-2)^2 \int_0^\infty q^{-1} f_p(q, \psi) dq \\ &= p - (p-2)^2 e^{-\psi/2} 2^{-1} \sum_{j=0}^\infty \frac{(\psi/2)^j}{j! \left(\frac{p}{2} - 1 + j\right)}. \end{aligned}$$

To numerically calculate the risk of these estimators, we used these formulas, truncating the series upon convergence.

The risk of the remaining estimators are not available in closed form. In these cases we computed the risk by numerical integration using expression (2.5). The integral was approximated using 20,000 equally-spaced gridpoints between 0 and an upper bound set to exceed the 99.99% quantile of the $\chi_{p+2}^2(\psi)$ distribution.

Similarly, to compute the efficiency bound $R_p(\psi)$ the integral in (3.6) was calculated numerically using 20,000 equally-spaced gridpoints between $q_1(\psi)$ and $q_2(\psi)$.

The MaxRegret of the estimators was approximated by taking the maximum of the Regret computed for each ψ on a grid of 2,000 equally-spaced values between 0 and $20p$. (For δ^{TL} we used a 4,000 gridpoints up to $40p$ as the Regret function is much flatter in ψ .)

The optimal parameters reported in Table 1 were calculated by minimizing the MaxRegret over a grid of values. The resolution of the grid is indicated by the number of reported digits. The parameters reported in Table 3 for the Trimmed Linear estimator were obtained by minimizing the MaxRegret using a constrained BFGS algorithm.

REFERENCES

- ALAM, K. (1973). A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **1** 517-525.
- BARANCHIK, A. J. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Stanford Univ. Technical Report No. 50.
- BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642-645.

- BERGER, J. O. (1976a). Admissible minimax estimation of a multivariate mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223-226.
- BERGER, J. O. (1976b). Tail minimaxality in location vector problems and its applications. *Ann. Statist.* **4** 33-50.
- BOCK, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **3** 209-218.
- DASGUPTA, A. and STRAWDERMAN, W. E. (1997). All estimates with a given risk, Riccati differential equations and a new proof of a theorem of Brown. *Ann. Statist.* **25** 1208-1221.
- EFRON, B. E. and MORRIS, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11-21.
- FAITH, R. E. (1978). Minimax Bayes estimators of a multivariate normal mean. *J. Multivariate Anal.* **8** 372-379.
- FOURDRINIER, D., STRAWDERMAN, W. E. and WELLS, M. T. (1998). On the construction of Bayes minimax estimators. *Ann. Statist.* **26** 660-671.
- GUO, Y. Y. and PAL, N. (1992). A sequence of improvements over the James-Stein estimator. *J. Multivariate Anal.* **42** 302-317.
- JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361-379.
- KUBOKAWA, T. (1991). An approach to improving the James-Stein estimator. *J. Multivariate Anal.* **36** 121-126.
- KURIKI, S. and TAKEMURA, A. (2000). Shrinkage estimation towards a closed convex set with a smooth boundary. *J. Multivariate Anal.* **75** 79-111.
- LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation, Second Edition*. Springer, New York.
- LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses, Third Edition*. Springer, New York.
- LI, T. F. and KUO, W. H. (1982). Generalized James-Stein estimators. *Comm. Statist. Theory Methods* **11** 2249-2257.
- MAGNUS, W., OBERHETTINGER, F. and SONI, R. P. (1966). *Formulas and Theorems for the Special Functions of Mathematical Physics*. Springer, New York.
- MARUYAMA, Y. (1998). A unified and broadened class of admissible minimax estimators of a multivariate normal mean. *J. Multivariate Anal.* **64** 196-205.
- MARUYAMA, Y. (2004). Stein's idea and minimax admissible estimation of a multivariate normal mean. *J. Multivariate Anal.* **88** 320-334.
- MARUYAMA, Y. (2007). Some notes on improving upon the James-Stein estimator. *J. Stat. Stud.* **26** 77-84.
- SHAO, P. Y.-S. and STRAWDERMAN, W. E. (1994). Improving on the James-Stein positive-part estimator. *Ann. Statist.* **22** 1517-1538.
- STEIN, C. (1973). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symposium Asymptotic Statistics* 345-381.
- STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135-1151.
- STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385-388.

DEPARTMENT OF ECONOMICS
1180 OBSERVATORY DRIVE
MADISON, WI 53706
E-MAIL: behansen@wisc.edu
URL: <http://www.ssc.wisc.edu/~bhansen>