Econometric Theory, **31**, 2015, 860–879.
doi:10.1017/S0266466614000693

SHRINKAGE EFFICIENCY BOUNDS

BRUCE E. HANSEN
University of Wisconsin

This paper is an extension of Magnus (2002, *Econometrics Journal* 5, 225–236) to multiple dimensions. 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 in the class of minimax orthogonally invariate estimators satisfying the sufficient conditions of Efron and Morris (1976, *Annals of Statistics* 4, 11–21). This allows us to compare the regret of existing orthogonally invariate 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, $X \sim \mathrm{N}(\theta, \mathbf{I}_p)$. The goal is to estimate θ. Consider the class of orthogonally invariate estimators which shrink X toward the zero vector and can be written as

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

where $\phi : [0, \infty) \to [0, \infty)$. Under quadratic loss, the risk of the estimator (1) is

$$R_p(\psi, \delta_\phi) = \mathrm{E}\left\|\delta_\phi(X) - \theta\right\|^2, \tag{2}$$

where $\psi = \|\theta\|^2$. We write the risk as a function of the scalar ψ as it is well known (e.g., equation (3)) that the risk only depends on θ through ψ and p.

For any given estimator $\delta_\phi(X)$ its maximum risk is $R_p^*\left(\delta_\phi\right) = \sup_\psi R_p\left(\psi, \delta_\phi\right)$. The maximum risk of the usual estimator $\delta_\phi(X) = X$ is $R_p^*(X) = p$ and is the minimax bound (e.g., Proposition 8.6 of van der Vaart, 1998), meaning that $R_p^*\left(\delta_\phi\right) \geq p$ for all estimators $\delta_\phi(X)$. An estimator such that $R_p^*\left(\delta_\phi\right) = p$ is called *minimax*. When $p > 2$ the usual estimator X is no longer the unique minimax estimator. Sufficient conditions on ϕ for the estimator (1) to be minimax have been developed by Baranchik (1970), Strawderman (1971), Alam (1973), Stein (1973, 1981), Berger (1976a, 1976b), Efron and Morris (1976), Faith (1978), DasGupta and Strawderman (1997), and Fourdrinier, Strawderman, and Wells (1998).

Research supported by the National Science Foundation. I thank a referee for a thorough reading of the paper and helpful comments. Address correspondence to Bruce Hansen, Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706, USA; e-mail: behansen@wisc.edu.

 © Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

Many minimax estimators in the class (1) have been proposed. The most famous are the James–Stein estimator (James and Stein, 1961) and its 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, 2004, 2007), and Kuriki and Takemura (2000). 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) 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 $R_p\left(\psi, \delta_\phi\right)$ across the class of minimax estimators defined by the sufficient conditions of 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.

In recent years, a new class of shrinkage estimators has been developed around the concept of the Lasso introduced by Tibshirani (1996). Both the Lasso and the estimator class (1) shrink unrestricted estimates toward zero, but with quite different effects. The orthogonally invariant estimators (1) shrink all estimates proportionately, while Lasso estimates shrink coefficients individually. Consequently, Lasso estimates can work much better in sparse settings, while the estimator class (1) can work much better when coefficients are of similar magnitude. Lasso-type estimators have received considerable attention in the recent statistics literature and have started to become quite popular in the econometrics literature as well. See, for example, Caner (2009), Belloni and Chernozhukov (2011), Belloni, Chen, Chernozhukov, and Hansen (2012), Belloni, Chernozhukov, and Fernandez-Val (2013), Liao (2013), Belloni, Chernozhukov, and Hansen (2014), and Liao and Phillips (2014).

Gauss code for the numerical work is available on the author's webpage http://www.ssc.wisc.edu/~bhansen/.

2. RISK

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

THEOREM 1. *If ϕ is Borel measurable,*

$$R_p(\psi, \delta_\phi) = R_p^0(\psi) + D_p(\psi, \phi), \tag{3}$$

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

where

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

$$D_p(\psi,\phi) = \int_0^\infty \left(\phi(q) - \phi_p^*(q,\psi)\right)^2 q^{-1} f_p(q,\psi)\, dq, \tag{4}$$

$$\phi_p^*(q,\psi) = q - \psi \frac{f_{p+2}(q,\psi)}{f_p(q,\psi)}, \tag{5}$$

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

Proof. For any ϕ,

$$\left\|\delta_\phi(X) - \theta\right\|^2 = \|X-\theta\|^2 + \frac{\phi^2(\|X\|^2)}{\|X\|^2} - 2\phi(\|X\|^2) + 2\theta' X \frac{\phi(\|X\|^2)}{\|X\|^2}. \tag{6}$$

Taking expectations of (6), using the fact that for any function $g(q)$, $\mathrm{E}(Xg(\|X\|^2)) = \theta \mathrm{E}(g(\chi_{p+2}^2(\psi)))$, (see, for example, Bock, 1975, Thm. A), we obtain

$$\begin{aligned} R_p(\psi,\delta_\phi) = p &+ \int_0^\infty \left(\phi^2(q) - 2q\phi(q)\right) q^{-1} f_p(q,\psi)\, dq \\ &+ 2\psi \int_0^\infty \phi(q) q^{-1} f_{p+2}(q,\psi)\, dq. \end{aligned} \tag{7}$$

Using definition (5) and completing the square, this equals

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

■

Theorem 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 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 $\lim_{q\to\infty} h_p^*(q,\psi) = 1$.*

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

See the Appendix for the proof.

On a side note, one useful implication of Theorem 1 and Lemma 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. *If* $P(\phi(\|X\|^2) > \|X\|^2) > 0$, *then* $R_p(\psi, \delta_\phi^+) < R_p(\psi, \delta_\phi)$.

Proof. Lemma 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 (4), we see that $D_p(\psi, \phi) > D_p(\psi, \phi^*)$, and the inequality is strict since $P(Q) > 0$ by assumption. From (3) we find $R_p(\psi, \delta_\phi^+) < R_p(\psi, \delta_\phi)$ as required. ■

Contrast Lemma 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 shows that this restriction is unnecessary.

3. EFFICIENCY BOUNDS

An efficiency bound is the infimum of $R_p\left(\psi, \delta_\phi\right)$ across ϕ in some function class. We focus on minimax estimators and thus restrict attention for the remainder of the paper to $p \geq 3$. Efron and Morris (1976, Thm. 3) established that the estimator (1) is minimax if

$$0 \leq \phi(q) \leq 2(p-2) \qquad \text{for all} \qquad q \geq 0, \tag{8}$$

if for all q with $\phi(q) < 2(p-2)$, then

$$\zeta(q) = \frac{q^{(p-2)/2}\phi(q)}{2(p-2) - \phi(q)} \tag{9}$$

is nondecreasing, and if q^* exists such that $\phi(q^*) = 2(p-2)$, then

$$\phi(q) = 2(p-2) \qquad \text{for all} \qquad q \geq q^*. \tag{10}$$

Conditions (8)–(10) are broader than those of Baranchik (1970), Strawderman (1971), and Alam (1973). Notice that condition (9) is satisfied if $\phi(q)$ is nondecreasing, but (9) is broader and allows $\phi(q)$ to be nonmonotonic. Conditions (8)–(10) appear to be close to necessary for minimaxity, as to my knowledge there is no known minimax estimator which does not satisfy (8)–(10). Furthermore, Efron and Morris (1976, Thm. 2) established that (8)–(10) are necessary for minimaxity assuming that the function $\phi(q)$ is absolutely continuous and that the Stein unbiased estimator of risk exists. Consequently, we will restrict attention to estimators which satisfy (8)–(10). This is a slight restriction on the class of minimax estimators, but a restriction which does not appear to be particularly important.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

Let Φ be the set of functions satisfying (8)–(10). The efficiency bound for shrinkage estimators in the class Φ is the infimum of $R_p\left(\psi, \delta_\phi\right)$ across $\phi \in \Phi$:

$$R_p(\psi) = \inf_{\phi \in \Phi} R_p\left(\psi, \delta_\phi\right). \tag{11}$$

From decomposition (3) we see that minimizing $D_p(\psi, \phi)$ is equivalent to minimizing $R_p\left(\psi, \delta_\phi\right)$. 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 (8). In this case the minimizer of $R_p\left(\psi, \delta_\phi\right)$ is the truncated function

$$\begin{aligned}\phi_p^{**}(q, \psi) = \phi_p^*(q, \psi)\mathbb{I}\left(0 \le \phi_p^*(q, \psi) \le 2(p-2)\right) \\ + 2(p-2)\mathbb{I}\left(\phi_p^*(q, \psi) > 2(p-2)\right).\end{aligned}$$

Since the function $h_p^*(q, \psi)$ is continuous in q (Lemma 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 1), then $\phi_p^{**}(q, \psi)$ is monotonically increasing in q on the set $\phi_p^{**}(q, \psi) \ge 0$. It follows that $\phi_p^{**}(q, \psi)$ is continuous and nondecreasing in $q \ge 0$, and therefore satisfies conditions (9) and (10) in addition to (8). It follows that $\phi_p^{**}(q, \psi)$ is the minimizer of $R_p\left(\psi, \delta_\phi\right)$ over $\phi \in \Phi$, and thus (1) with $\phi(q) = \phi_p^{**}(q, \psi)$ is the pointwise (in ψ) optimal minimax estimator for θ.

From (3) we see that the minimized value (the efficiency bound) is

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

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

THEOREM 2. *For $p \ge 3$, the efficiency bound for estimation of θ by (1) with $\phi \in \Phi$ is*

$$\begin{aligned}R_p(\psi) = p - 8(p-2) f_p\left(q_2(\psi), \psi\right) \\ - \int_{q_1(\psi)}^{q_2(\psi)} \phi_p^*(q, \psi)^2 q^{-1} f_p(q, \psi)\, dq,\end{aligned} \tag{12}$$

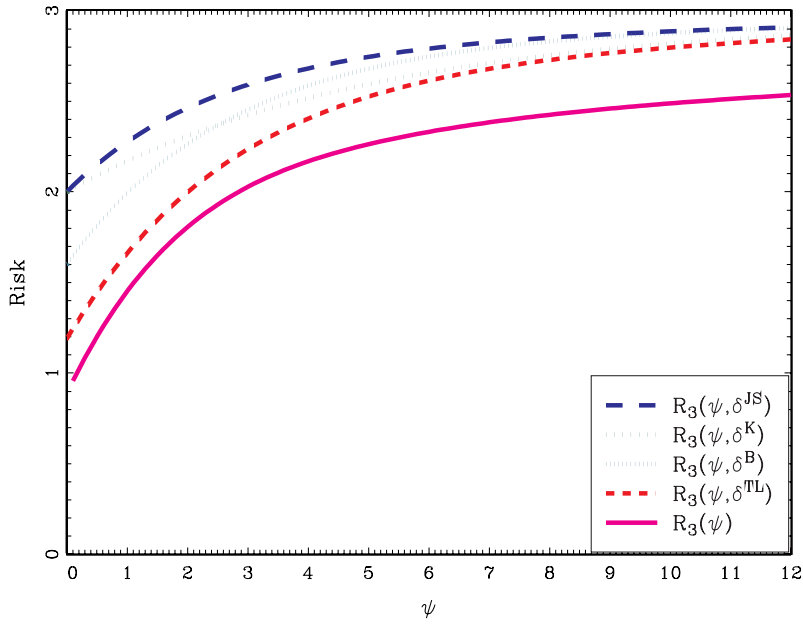
where $q_1(\psi)$ is the smallest $q \ge 0$ such that $h_p^*(q, \psi) \ge 0$ and $q_2(\psi)$ is the unique $q > 0$ where $\phi_p^*(q, \psi) = 2(p-2)$.

See the Appendix for the proof.

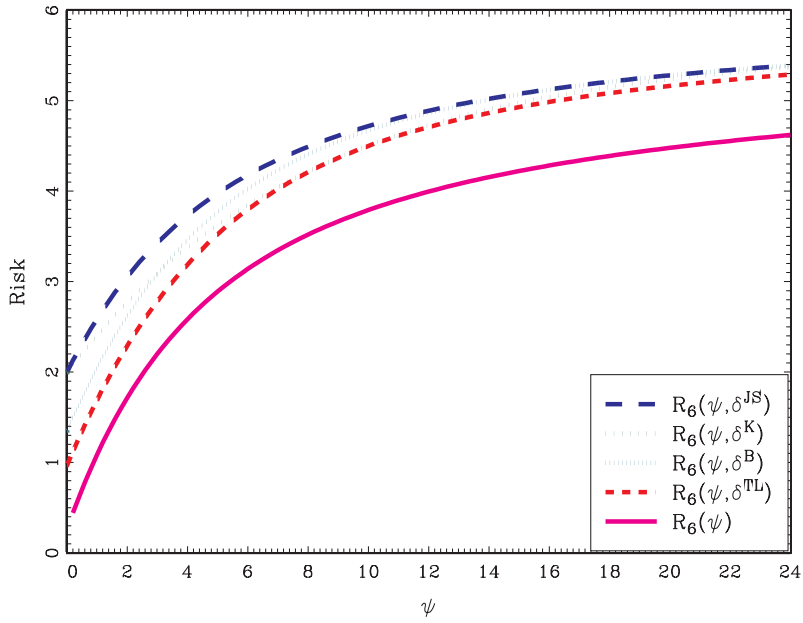
While the efficiency bound is the lower envelope of the risk functions of all minimax shrinkage estimators satisfying (8)–(10), there is no single estimator whose risk function equals the efficiency bound.

We display[1] 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

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press



(a)



(b)

FIGURE 1. Risk and efficiency bounds.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

increasing in ψ, asymptotes toward 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 ualitatively similar.

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:

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

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 minimaxity.

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)

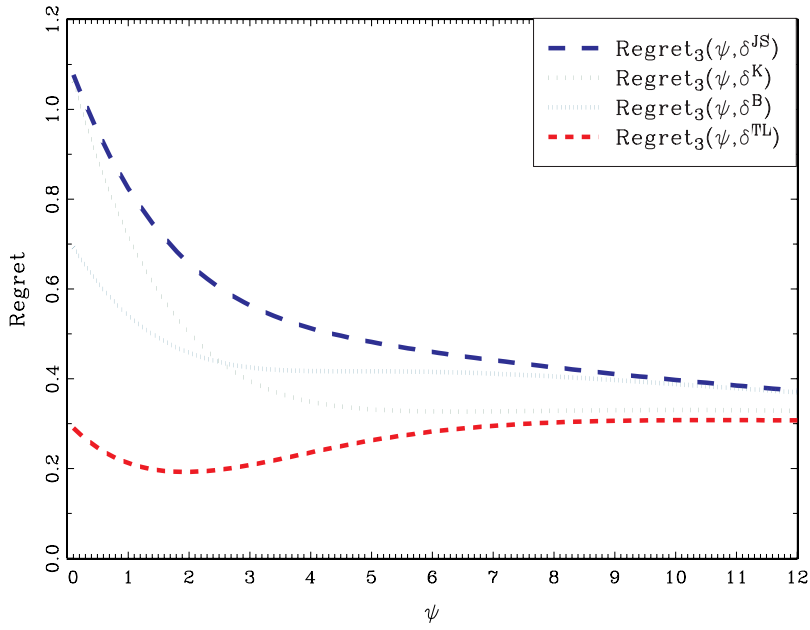
   $\phi^{LK}(q) = p - 2 - c_1 q^{-\alpha_1/2}$,

   where

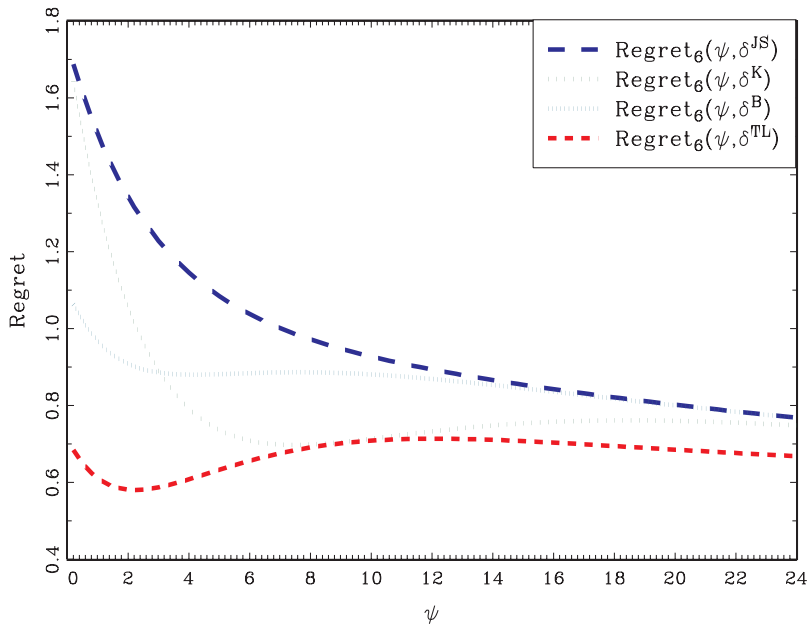
   $$c_1 = \alpha_1 2^{\alpha_1/2} \frac{\Gamma(p/2 - (1 + \alpha_1/2))}{\Gamma(p/2 - (1 + \alpha_1))}$$

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

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press



(a)



(b)

FIGURE 2. Regret.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

4. Guo and Pal (1992)

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

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

5. Shao and Strawderman (1994)

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

where

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

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)

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

with $r \ge 0$.

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

$$3 - p/2 \le a \le 1 + p/2,$$

$$b \ge -\frac{a + p/2 - 3}{3p/2 + 1 - a},$$

and

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

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

where

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

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 $p = 3, \ldots, 10$. 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 Kubokawa. 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) and Maruyama (1998, 2004), but these three estimators are essentially equivalent. Since the Strawderman estimator includes Kubokawa's as a special case, we can see the importance of optimal parameter selection. For example, for $p = 3$, Kubokawa's

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

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

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	0.297	0.476	0.600	0.693	0.765	0.824	0.873	0.907

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 satisfying (8)–(10):

$$\text{MinRegret}_p = \inf_{\phi \in \Phi} \text{MaxRegret}_p(\delta_\phi). \tag{13}$$

This is the lower bound for the MaxRegret among minimax shrinkage estimators satisfying (8)–(10).

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 $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_N < \infty$, let $\theta = \{a_1, b_1, \ldots, a_N, b_N, c\}$ denote a set of parameters. Our spline function takes the form

$$\phi_\theta(q) = \sum_{n=1}^{N} (a_n + b_n q)\mathbb{I}(\tau_{n-1} \le q \le \tau_n) + c\mathbb{I}(q > \tau_N) \tag{14}$$

subject to the continuity constraints

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

$$a_1 = 0 \tag{15}$$

$$a_n + b_n\tau_n = a_{n+1} + b_{n+1}\tau_n \qquad \text{for} \qquad 1 \le n \le N-1 \tag{16}$$

$$a_N + b_N\tau_N = c. \tag{17}$$

By picking N and τ_N sufficiently large, the class (14) can approximate any function in Φ (see, for example, Chapter 3 of de Boor, 2000).

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

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

Applying this formula to the continuous linear spline (14), it is straightforward to calculate the risk of δ_θ.

LEMMA 3. *The risk of the estimator (1) with the shrinkage function (14)–(17) is*

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

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 noncentral chi-square (see equation (A.1) in the Appendix) and integrating term-by-term, it is straightforward to obtain the following result.

LEMMA 4. *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 3 and 4, 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 (8) and (9), which is equivalent to the following conditions on the parameters in (14). For $n = 1, \ldots, N$,

$$2(p-2)a_n - a_n^2 + 2b_n(p - a_n)\tau_{n-1} - b_n^2\tau_{n-1}^2 \ge 0 \tag{18}$$

$$2(p-2)a_n - a_n^2 + 2b_n(p - a_n)\tau_n - b_n^2\tau_n^2 \ge 0 \tag{19}$$

$$0 \le a_n + b_n\tau_n \le 2(p-2). \tag{20}$$

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

Let Θ_N be the set of parameters θ which satisfy (15)–(17) and (18)–(20). The smallest MaxRegret among all minimax shrinkage estimators using the spline function (14) is then

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

Since the class of functions (14) 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 Broyden-Ftetcher-Goldfarb-Shanno (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) and Maruyama (1998, 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 an 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 (14) 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 3. It may be also convenient to observe that the estimator can be written as

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

By construction, $\phi^{TL}(q)$ satisfies the minimax conditions (8)–(10), 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 $p = 3, \ldots, 25$. 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.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

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

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

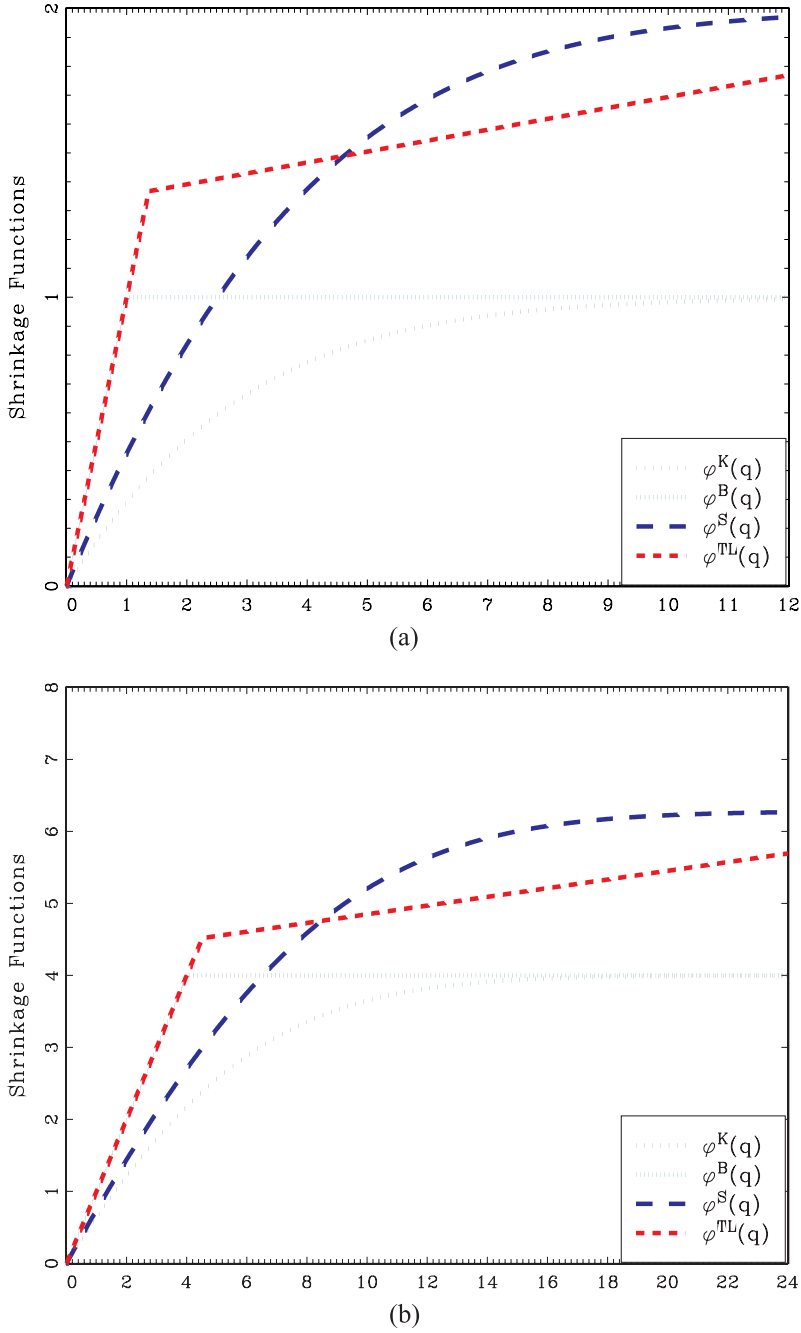


FIGURE 3. Shrinkage functions.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

7. CONCLUSION

This paper investigates orthogonally invariant shrinkage in the exact multivariate normal context. We develop a new efficiency bound for minimax shrinkage estimators and show how to rank existing shrinkage estimators using maximum regret. We use the maximum regret concept to motivate a new shrinkage estimator with better performance than the existing estimators. The gains are greatest in low-dimensional contexts which may be relevant for parametric econometric applications.

While our results are strictly for the exact multivariate normal context, they should apply asymptotically to any asymptotically normal estimator, as shown by Hansen (2014).

NOTE

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

REFERENCES

Alam, K. (1973) A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Annals of Statistics* 1, 517–525.

Baranchick, A.J. (1964) Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution. Technical report no. 51, Department of Statistics, Stanford University.

Baranchick, A.J. (1970) A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics* 41, 642–645.

Belloni, A., D. Chen, V. Chernozhukov, & C. Hansen (2012) Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80, 2369–2429.

Belloni, A. & V. Chernozhukov (2011) L1-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics* 39, 82–130.

Belloni, A., V. Chernozhukov, & I. Fernandez-Val (2013) Program Evaluation with High-dimensional Data. Working paper.

Belloni, A., V. Chernozhukov, & C. Hansen (2014) Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies* 81, 608–650.

Berger, J.O. (1976a) Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Annals of Statistics* 4, 223–226.

Berger, J.O. (1976b) Tail minimaxality in location vector problems and its applications. *Annals of Statistics* 4, 33–50.

Bock, M.E. (1975) Minimax estimators of the mean of a multivariate normal distribution. *Annals of Statistics* 3, 209–218.

Caner, M. (2009) Lasso-type GMM estimator. *Econometric Theory* 25, 1–23.

DasGupta, A. & W.E. Strawderman (1997) All estimates with a given risk, Riccati differential equations and a new proof of a theorem of Brown. *Annals of Statistics* 25, 1208–1221.

de Boor, C. (2000) *A Practical Guide to Splines*, rev. ed. Springer.

Efron, B.E. & C. Morris (1976) Families of minimax estimators of the mean of a multivariate normal distribution. *Annals of Statistics* 4, 11–21.

Faith, R.E. (1978) Minimax Bayes estimators of a multivariate normal mean. *Journal of Multivariate Analysis* 8, 372–379.

Fourdrinier, D., W.E. Strawderman, & M.T. Wells (1998) On the construction of Bayes minimax estimators. *Annals of Statistics* 26, 660–671.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

Guo, Y.Y. & N. Pal (1992) A sequence of improvements over the James-Stein estimator. *Journal of Multivariate Analysis* 42, 302–317.
Hansen, B.E. (2014) Efficient Shrinkage in Parametric Models. Working paper.
James, W. & C.M. Stein (1961) Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1, 361–380.
Kubokawa, T. (1991) An approach to improving the James-Stein estimator. *Journal of Multivariate Analysis* 36, 121–126.
Kuriki, S. & A. Takemura (2000) Shrinkage estimation towards a closed convex set with a smooth boundary. *Journal of Multivariate Analysis* 75, 79–111.
Lehmann, E.L. & G. Casella (1998) *Theory of Point Estimation*, 2nd ed. Springer.
Lehmann, E.L. & J.P. Romano (2005) *Testing Statistical Hypotheses*, 3rd ed. Springer.
Li, T.F. & K.H. Kuo (1982) Generalized James-Stein estimators. *Communications in Statistics - Theory and Methods* 11, 2249–2257.
Liao, Z. (2013) Adaptive GMM shrinkage estimation with consistent moment selection. *Econometric Theory* 25, 1–23.
Liao, Z. & P.C.B. Phillips (2014) Automated estimation of vector error correction models. *Econometric Theory*, forthcoming.
Magnus, J.R. (2002) Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal* 5, 225–236.
Magnus, W., F. Oberhettinger, & R.P. Soni (1966) *Formulas and Theorems for the Special Functions of Mathematical Physics*. Springer.
Maruyama, Y. (1998) A unified and broadened class of admissible minimax estimators of a multivariate mean. *Journal of Multivariate Analysis* 64, 196–205.
Maruyama, Y. (2004) Stein's idea and minimax admissible estimation of a multivariate normal mean. *Journal of Multivariate Analysis* 88, 320–334.
Maruyama, Y. (2007) Some notes on improving upon the James-Stein estimator. *Journal of Statistical Studies* 26, 77–84.
Phillips, P.C.B. (1984) The exact distribution of the Stein-rule estimator. *Journal of Econometrics* 25, 123–131.
Shao, P.Y.-S. & W.E. Strawderman (1994) Improving on the James-Stein positive-part estimator. *Annals of Statistics* 22, 1517–1538.
Stein, C.M. (1973) Estimation of the mean of a multivariate normal distribution. *Proceedings of the Prague Symposium on Asymptotic Statistics*, 345–381.
Stein, C.M. (1981) Estimation of the mean of a multivariate normal distribution. *Annals of Statistics* 9, 1135–1151.
Strawderman, W.E. (1971) Proper Bayes minimax estimators of the multivariate normal distribution. *Annals of Mathematical Statistics* 42, 385–388.
Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58, 267–288.
van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press.

APPENDIX

Proof of Lemma 1. The density of the noncentral chi-square can be written as

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

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

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

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

$a_j = 1/(j!\Gamma(1+p/2+j))$, and $b_j = 1/(j!\Gamma(p/2+j))$. 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 noncentral chi-square density is

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

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

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

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

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

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

Proof of Theorem 2. 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

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

The minimized risk is

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

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

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

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

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

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

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

Substituting, we obtain

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

as stated. ■

NUMERICAL APPENDIX

The risk of the estimators of James and Stein (1961), Baranchik (1964), Li and Kuo (1982), Guo and Pal (1992) and 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 4. For example,

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

To numerically calculate their risk, we used these formulas, truncating the series upon convergence. For the James–Stein estimator an alternative method would have been to use the moment expressions deduced from the finite-sample density derived by Phillips (1984).

The risk of the remaining estimators is not available in closed form. In these cases we computed the risk by numerical integration using expression (7). The integral was approximated using 20,000 equally spaced grid points 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 (12) was calculated numerically using 20,000 equally spaced grid points between $q_1(\psi)$ and $q_2(\psi)$.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press

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 grid points 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.

https://doi.org/10.1017/S0266466614000693 Published online by Cambridge University Press