The Annals of Statistics
1991, Vol. 19, No. 2, 952–960

SHRINKAGE DOMINATION IN A MULTIVARIATE COMMON MEAN PROBLEM

BY EDWARD I. GEORGE

University of Chicago

Consider the problem of estimating the $p \times 1$ mean vector θ under expected squared error loss, based on the observation of two independent multivariate normal vectors $Y_1 \sim N_p(\theta, \sigma^2 I)$ and $Y_2 \sim N_p(\theta, \lambda\sigma^2 I)$ when λ and σ^2 are *unknown*. For $p \geq 3$, estimators of the form $\delta_\eta = \eta Y_1 + (1 - \eta)Y_2$ where η is a fixed number in $(0, 1)$, are shown to be uniformly dominated in risk by Stein estimators in spite of the fact that independent estimates of scale are *unavailable*. A consequence of this result is that when λ is assumed known, shrinkage domination is robust to incorrect specification of λ.

1. Introduction. This paper considers the following multivariate common mean problem. Suppose we observe two independent $p \times 1$ multivariate normal vectors, $p \geq 3$:

$$Y_1 \sim N_p(\theta, \sigma^2 I) \quad \text{and} \quad Y_2 \sim N_p(\theta, \lambda\sigma^2 I), \tag{1.1}$$

where $Y_j = (Y_{1j}, \dots, Y_{pj})'$, $\theta = (\theta_1, \dots, \theta_p)'$ and λ and σ^2 are positive scalars. The problem is to find an estimator $\delta \equiv \delta(Y_1, Y_2)$ of θ when λ and σ^2 are unknown, under the risk criterion of expected squared error loss,

$$R(\psi, \delta) = E_\psi\|\delta - \theta\|^2, \tag{1.2}$$

where the expectation is taken over the sample space under the distribution determined by $\psi = (\theta, \lambda, \sigma^2)$. It is desirable to keep the risk small over the entire parameter space

$$\Psi \equiv \{\psi : \theta \in \mathbb{R}^p,\ \lambda \in (0, \infty),\ \sigma^2 \in (0, \infty)\}. \tag{1.3}$$

Note that by linear transformation, this problem is identical to estimating θ based on observing $Y_1 \sim N_p(\theta, \sigma^2\Sigma)$ and $Y_2 \sim N_p(\theta, \lambda\sigma^2\Sigma)$ under the criterion $R_\Sigma(\psi, \delta) = E_\psi(\delta - \theta)'\Sigma^{-1}(\delta - \theta)$, where Σ is a known covariance matrix.

This common mean problem might arise when p different items are measured by each of two unbiased measuring devices which have unknown and possibly different measurement precisions. For example, suppose the value of each of p parcels of real estate was assessed by two independent assessors. Letting Y_{ij} be the assessment of parcel i by assessor j and letting θ_i be the "true" value of parcel i, the setup (1.1) might be appropriate if one assessor was better (less variable) than the other. Note that it might be necessary to use transformed units to obtain constant variance for each assessor.

Received February 1989; revised March 1990.
AMS 1980 *subject classifications*. Primary 62H12; secondary 62C99, 62J07.
Key words and phrases. Risk, robustness, shrinkage estimation, Stein estimators.

The problem of estimating θ in the setting (1.1) is also discussed in detail by Zellner (1971). Numerous authors have studied related versions of the above problem in which independent variance estimates are available. References for this work include Bhattacharya (1980), Brown and Cohen (1974) and Cohen and Sackrowitz (1974) who consider the univariate common mean problem ($p = 1$), Chiou and Cohen (1985), Loh (1991) and Krishnamoorty (1989) who consider the multivariate common mean problem and Box and Tiao (1973) and Yancey, Judge and Miyazaki (1984) who consider the problem of making inference about a common vector of regression coefficients based on two independent regressions with unknown and possibly different residual variances. The proposed solutions to each of these problems rely on the availability of independent variance estimates. As will be seen, our multivariate common mean problem is distinguished by the feature that independent variance estimates are unavailable.

The main point of this paper is to show that any estimator of the form

$$\delta_\eta = \eta Y_1 + (1 - \eta) Y_2, \tag{1.4}$$

where η is a fixed number in (0, 1), can be uniformly dominated by a Stein-like shrinkage estimator with respect to the risk in (1.2) for *all* values of $\psi \in \Psi$. Note that the risk of δ_η is

$$R(\psi, \delta_\eta) = p\sigma_\eta^2 \quad \text{where } \sigma_\eta^2 \equiv \left[\eta^2 + (1 - \eta)^2 \lambda\right]\sigma^2 \tag{1.5}$$

is the variance of δ_η.

When the variance ratio λ is known, δ_η for which $\eta = \lambda/(1 + \lambda)$ is the minimax estimator, the MLE, the BLUE and the best translation-equivariant estimator for θ. Indeed, in this case the setup (1.1) can be reduced to observing the sufficient statistics

$$Y_\lambda \equiv \frac{\lambda}{1 + \lambda} Y_1 + \frac{1}{1 + \lambda} Y_2 \quad \text{and} \quad S \equiv \|Y_1 - Y_2\|^2, \tag{1.6a}$$

which are independently distributed as

$$Y_\lambda \sim N_p(\theta, \sigma_\lambda^2 I), \quad \text{where } \sigma_\lambda^2 = \frac{\lambda}{1 + \lambda}\sigma^2 \tag{1.6b}$$

and

$$S \sim \frac{(1 + \lambda)^2}{\lambda} \sigma_\lambda^2 \chi_p^2. \tag{1.6c}$$

Thus, if the parameter space can be restricted to

$$\Psi_\eta \equiv \{\psi \colon \psi \in \Psi, \lambda/(1 + \lambda) = \eta\} \tag{1.7}$$

on which $\delta_\eta \equiv Y_\lambda$, then our problem can be reduced to that of estimating a multivariate normal mean based on a single observation in which case the estimator $\delta_\eta \equiv Y_\lambda$ has the aforementioned properties.

Of course, a consequence of such a reduction is that $\delta_\eta \equiv Y_\lambda$ is inadmissible (on Ψ_η), with respect to the risk (1.2), and can be dominated a Stein estimator of the form

$$\delta_\eta^S = \left[1 - \frac{(p-2)\hat{\sigma}_\eta^2}{\|\delta_\eta\|^2}\right]\delta_\eta, \tag{1.8}$$

where $\hat{\sigma}_\eta^2 = \eta(1-\eta)S/(p+2)$ is an estimator of σ_η^2. (Note that on Ψ_η, $\sigma_\eta^2 = \sigma_\lambda^2$.) The known proofs of this dominance [see Stein (1981) or Lehmann (1983)] exploit the independence of $\delta_\eta \equiv Y_\lambda$ and S on Ψ_η [a consequence of $\text{Cov}(\delta_\eta, Y_1 - Y_2) = (\eta - (1-\eta)\lambda)\sigma^2 = 0$ when $\eta = \lambda(1+\lambda)$], by first conditioning on S, and then making use of the same results used to dominate $\delta_\eta \equiv Y_\lambda$ when σ_η^2 is known.

What distinguishes our multivariate common mean problem from the setting above is that λ is unknown, and hence Y_λ is unavailable. The MLE does not exist. There is no sufficiency reduction; Y_1 and Y_2 are the minimal sufficient statistics. Furthermore, the statistics δ_η and S will be dependent for any $\psi \notin \Psi_\eta$. However, in spite of this dependence, it is shown in the next section that any estimator of the form δ_η will still be uniformly dominated by a shrinkage estimator similar to δ_η^S for all $\psi \in \Psi$.

Aside from the surprising fact that uniform domination by a shrinkage estimator is possible even when independent estimates of scale are unavailable, this result establishes the following robustness property of the Stein estimator. In some situations, it will be useful and plausible to assume λ is known, rather than treat the general case. For example, an equal variance assumption where $\lambda = 1$ might be justified by symmetry considerations. In this case, one might want to use the Stein estimator $\delta_{1/2}^S$ which is known to dominate the minimax estimator $\delta_{1/2}$. However, if λ is incorrectly specified, one would (or at least should) want to know the consequences. The main results of this paper show that shrinkage estimators continue to dominate δ_η for all λ so that shrinkage domination is robust to incorrect specification of λ.

It is important to emphasize that this paper does not settle the question of how to select a good estimator of θ when λ is indeed unknown. In this case, it is probably more reasonable to consider adaptive alternatives to δ_η of the form $\delta_\gamma = \gamma Y_1 + (1-\gamma)Y_2$, where γ is an adaptive estimator of η. We have obtained such adaptive δ_γ which dominate δ_η everywhere except for a small neighborhood around $\eta = \lambda/(1+\lambda)$ where they are practically as good. The main results of this paper suggest that analogous shrinkage domination results may also apply to these adaptive δ_γ. We plan to report separately on this problem

2. Shrinkage domination. In this section, we prove that any estimator of the form $\delta_\eta = \eta Y_1 + (1-\eta)Y_2$ from (1.4) will, for certain choices of c, be uniformly dominated by a shrinkage estimator of the form

$$\delta_\eta^c = \left[1 - c\frac{\|Y_1 - Y_2\|^2}{\|\delta_\eta\|^2}\right]\delta_\eta. \tag{2.1}$$

It may be of interest to note that our proof is initially directed toward deriving an unbiased estimator of the risk; see Berger (1985). However, this unbiased estimator of the risk turns out to be unbounded, and so must be averaged to obtain a bound. The details regarding the bounding of this average, which may also be of independent interest, have been presented as lemmas in Section 3.

THEOREM 2.1. *Let* $C(\eta, 3) = 2\kappa/13$ *and* $C(\eta, p) = 2(p-2)\kappa/(p+8)$ *for* $p \geq 4$, *where* $\kappa \equiv \min\{\eta^2, (1-\eta)^2\}$. *Then for* $\eta \in (0,1)$, $p \geq 3$ *and* $0 < c < C(\eta, p)$, δ_η^c *uniformly dominates* δ_η *in risk, that is,* $R(\psi, \delta_\eta^c) < R(\psi, \delta_\eta)$ *for all* $\psi \in \Psi$.

PROOF. Letting $S = \|Y_1 - Y_2\|^2$ as in (1.5), rewrite δ_η^c in (2.1) as

$$\delta_\eta^c = \delta_\eta + g_\eta^c \quad \text{where } g_\eta^c \equiv g_\eta^c(Y_1, Y_2) = -\frac{cS}{\|\delta_\eta\|^2}\delta_\eta. \tag{2.2}$$

The risk of δ_η^c can thus be expressed as

$$R(\psi, \delta_\eta^c) = E_\psi\|\delta_\eta + g_\eta^c - \theta\|^2 = p\sigma_\eta^2 + E_\psi\left[2(\delta_\eta - \theta)' g_\eta^c + \|g_\eta^c\|^2\right]. \tag{2.3}$$

Using the fact due to Stein (1981), that $E(X-\mu)h(X) = \sigma^2 E h'(X)$ for $X \sim N(\mu, \sigma^2)$ and any differentiable function h satisfying $E|h'(X)| < \infty$, it follows that

$$\begin{aligned} &E_\psi\left[(\delta_\eta - \theta)' g_\eta^c\right] \\ &\quad = E_\psi\left[\left(\eta(Y_1 - \theta) + (1-\eta)(Y_2 - \theta)\right)' g_\eta^c\right] \\ &\quad = E_\psi\left[\left(\eta\sigma^2\nabla_1 \cdot + (1-\eta)\lambda\sigma^2\nabla_2 \cdot\right) g_\eta^c\right] \\ &\quad = E_\psi\left[-c\sigma_\eta^2(p-2)\frac{S}{\|\delta_\eta\|^2} - 2c\sigma^2(\eta - (1-\eta)\lambda)\frac{(Y_1 - Y_2)'\delta_\eta}{\|\delta_\eta\|^2}\right], \end{aligned} \tag{2.4}$$

where $\nabla_t \cdot g_\eta^c = \sum_{i=1}^p (\partial g_\eta^c / \partial Y_{it})$. Thus, the risk in (2.3) may be expressed as

$$R(\psi, \delta_\eta^c) = p\sigma_\eta^2 + (c^2 A_1 - 2c(p-2)A_2)\sigma_\eta^2 - 4cA_3, \tag{2.5}$$

where

$$A_1 = \sigma_\eta^{-2} E_\psi\left[S^2/\|\delta_\eta\|^2\right], \qquad A_2 = E_\psi\left[S/\|\delta_\eta\|^2\right], \tag{2.6}$$

$$A_3 = (\eta - (1-\eta)\lambda)\sigma^2 E_\psi\left[\frac{(Y_1 - Y_2)'\delta_\eta}{\|\delta_\eta\|^2}\right]. \tag{2.7}$$

By Lemma 3.1, $A_1/A_2 < 13\kappa^{-1}$ for $p = 3$, and $A_1/A_2 \leq (p+8)\kappa^{-1}$ for $p \leq 4$. Furthermore, $A_3 > 0$ by Lemma 3.5. Thus, for $0 < c < C(\eta, p)$,

$$\begin{aligned} R(\psi, \delta_\eta^c) &\leq p\sigma_\eta^2 + (c^2 A_1 - 2c(p-2)A_2)\sigma_\eta^2 \\ &\leq p\sigma_\eta^2 + (c(p+8)\kappa^{-1} - 2(p-2))cA_2\sigma_\eta^2 < p\sigma_\eta^2. \end{aligned}$$ □

Although there is no minimax estimator for the general problem where λ is unknown, it may be useful to regard $\delta_{1/2} = (1/2)(Y_1 + Y_2)$ as a benchmark competitor for the general problem. Note that $\delta_{1/2}$ is the "safest" estimator of the form δ_η, in the sense that $RR(\delta_{1/2}, \lambda) \le RR(\delta_{\lambda/(1+\lambda)}, 1)$, where we define $RR(\delta_\eta, \lambda) = R(\psi, \delta_\eta)/R(\psi, Y_\lambda)$, the risk efficiency of δ_η at λ. The following is just Theorem 2.1 applied to $\delta_{1/2}$.

COROLLARY 2.1. *Let $C^*(3) = 1/26$ and $C^*(p) = (p-2)/2(p+8)$ for $p \ge 4$. Then for $p \ge 3$ and $0 < c < C^*(p)$, $\delta_{1/2}^c$ uniformly dominates $\delta_{1/2}$ in risk.*

Evidently, the bounds on c in Theorem 2.1 and Corollary 2.1 are not the largest possible. It appears from strong simulation evidence that the bound on A_1/A_2 given in Lemma 3.1 can be lowered to at least $A_1/A_2 < 6\kappa^{-1}$ for $p = 3$, and to $A_1/A_2 \le (p+2)\kappa^{-1}$ for $p \ge 4$. If this is the case, the bound for c can be increased to $C(\eta, 3) = \kappa/3$ and $C(\eta, p) = 2(p-2)\kappa/(p+2)$ for $p \ge 4$ in Theorem 2.1, and to $C^*(3) = 1/12$ and $C^*(p) = (p-2)/2(p+2)$ for $p \ge 4$ in Corollary 2.1. However, even these bounds are probably conservative since the amount of risk reduction captured by the term A_3 in (2.7) has not been exploited by the proof of Theorem 2.1.

Simulation evidence also suggests that for $\eta = 1/2$, A_1/A_2 is bounded below by $4p$ when $\theta = 0$ and $\lambda \to 0$ or ∞, and is close to its apparent maximum of about $4(p+2)$ in the rest of the parameter space. Because of this relatively small variation of A_1/A_2 over Ψ, it follows from (2.5) that a good choice for the constant in $\delta_{1/2}^c$ would be $c = (p-2)/4(p+2)$. This is exactly the choice in (1.8) recommended by Stein (1966) for the situation when it is known that $\lambda = 1$. Unfortunately, the choice of c for other δ_η^c is less clear, especially when η is close to 0 or 1, where A_1/A_2 can have large variation. Of course, in all these cases, a positive-part version would probably be better.

To get some sense of the amount of improvement available by δ_η^c over δ_η, we simulated the risk of δ_η^S in (1.8) for $\eta = 0.5, 0.75, 0.9$ when $|\theta|^2 = 0, 10, 1000$ and $\lambda = 1, 3, 9$. The results are presented in Table 1. The risks of the corresponding δ_η [which can be computed exactly from (1.5)] are also given for comparison. Of these δ_η, $\delta_{0.5}$ is best when $\lambda = 1$, $\delta_{0.75}$ is best when $\lambda = 3$ and $\delta_{0.9}$ is best when $\lambda = 9$. What is most striking in the table is that δ_η^S dominates the corresponding δ_η in every setting, sometimes substantially. This underscores the shrinkage domination robustness of the Stein estimator discussed in the previous section. Finally, note that when $\eta = 0.75$ or 0.9, the choice $c = \eta(1-\eta)(p-2)/(p+2)$ for δ_η^S is larger than $C(\eta, p)$, the bound for domination given in Theorem 2.1. The observed shrinkage domination in these cases further supports the conclusion that the bounds on c given by Theorem 2.1 are conservative.

3. Lemmas for proving Theorem 2.1. In this final section, we present the details required for obtaining the risk bounds in Theorem 2.1. These are organized as a series of interconnected lemmas which may also be of indepen-

TABLE 1
Simulation estimates of risk $R(\psi, \delta)$
$p = 10, \sigma^2 = 1$

δ	$\|\theta\|^2$	$\eta = 0.5$	$\eta = 0.75$	$\eta = 0.9$
		$\lambda = 1$		
	0.	1.66	2.64	6.03
δ_η^S	10.	3.99	4.95	7.25
	1000.	4.99	6.23	8.19
δ_η		5.00	6.25	8.20
		$\lambda = 3$		
	0.	2.69	2.50	4.59
δ_η^S	10.	6.52	5.51	6.76
	1000.	9.94	7.47	8.38
δ_η		10.00	7.50	8.40
		$\lambda = 9$		
	0.	4.52	5.47	3.01
δ_η^S	10.	10.42	8.48	6.33
	1000.	24.58	11.20	8.96
δ_η		25.00	11.25	9.00

Note: All standard errors are less than or equal to 0.004 for $\lambda = 1$ and $\lambda = 3$, and they are less than or equal to 0.012 for $\lambda = 9$. The risk entries for δ_η are computed exactly using (1.5). The simulation was based on 1,000,000 repetitions.

dent interest. Related results can be found in Casella and Hwang (1982). Note that many of the bounds provided by the lemmas here are not the tightest possible.

LEMMA 3.1. *For A_1 and A_2 in (2.6) and $\kappa \equiv \min\{\eta^2, (1-\eta)^2\}$,*

(i) $$A_1/A_2 < 13\kappa^{-1} \qquad \text{for } p = 3,$$

(ii) $$A_1/A_2 \le (p+8)\kappa^{-1} \qquad \text{for } p \ge 4.$$

PROOF. A_1 and A_2 may be expressed as

$$A_1 = (1+\lambda)^2\sigma^4\sigma_\eta^{-4}E(\|U\|^4/\|V\|^2), \qquad A_2 = (1+\lambda)\sigma^2\sigma_\eta^{-2}E(\|U\|^2/\|V\|^2),$$

where U and V are two $p \times 1$ normal vectors such that

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \begin{pmatrix} I & \alpha I \\ \alpha I & I \end{pmatrix}\right),$$

with $\mu = \theta/\sigma_\eta$ and $\alpha = (\eta - (1-\eta)\lambda)\sigma/(1+\lambda)^{1/2}\sigma_\eta$. Thus

$$A_1/A_2 = (1+\lambda)\sigma^2\sigma_\eta^{-2}h(\mu, \alpha),$$

where

$$h(\mu, \alpha) \equiv \frac{E_{\mu,\alpha}(\|U\|^4/\|V\|^2)}{E_{\mu,\alpha}(\|U\|^2/\|V\|^2)}.$$

Since $(1+\lambda)\sigma^2\sigma_\eta^{-2} = (1+\lambda)[\eta^2 + (1-\eta)^2\lambda]^{-1} \le \kappa^{-1}$, it suffices to show that $h(\mu, \alpha) < 13$ for $p = 3$, and $h(\mu, \alpha) \le p + 8$ for $p \ge 4$. Conditioning on V yields

$$E(\|U\|^2|V) = \alpha^2\|V-\mu\|^2 + (1-\alpha^2)p,$$
$$E(\|U\|^4|V) = \alpha^4\|V-\mu\|^4 + 2\alpha^2(1-\alpha^2)(p+2)\|V-\mu\|^2 + (1-\alpha^2)^2 p(p+2).$$

Thus, letting

$$x_1 = E_\mu(1/\|V\|^2), \qquad x_2 = E_\mu(\|V-\mu\|^2/\|V\|^2), \qquad x_3 = E_\mu(\|V-\mu\|^4/\|V\|^2), \tag{3.1}$$

$h(\mu, \alpha)$ may be rewritten as

$$h(\mu,\alpha) = \frac{\alpha^4 x_3 + 2\alpha^2(1-\alpha^2)(p+2)x_2 + (1-\alpha^2)^2 p(p+2)x_1}{\alpha^2 x_2 + (1-\alpha^2)px_1}$$
$$= \alpha^2\left[\beta\frac{x_3}{x_2} + (1-\beta)\frac{p+2}{p}\frac{x_2}{x_1}\right] + (1+\alpha^2)(p+2),$$

where $\beta = \alpha^2 x_2/(\alpha^2 x_2 + (1-\alpha^2)px_1)$. The desired bounds on $h(\mu, \alpha)$ now follow directly from this last equality and Lemma 3.2. □

LEMMA 3.2. *For $V \sim N_p(\mu, I)$, let x_1, x_2 and x_3 be as in* (3.1). *Then*

(i) *for $p = 3$, $x_2/x_1 < 5$ and $x_3/x_2 < 13$,*

(ii) *for $p \ge 4$, $x_2/x_1 \le p$ and $x_3/x_2 \le p + 8$.*

PROOF. As in Lemma 3.3, define $\varphi(q) \equiv E[q + 2Z]^{-1}$ for $q \ne 0, -2, -4, \dots$, where Z is a Poisson random variable with mean $\lambda = \|\mu\|^2/2$. Making use of Lemma 3.4, x_1, x_2 and x_3 may be expressed as

$$x_1 = \varphi(p-2), \tag{3.2a}$$

$$x_2 = E_\mu[1 - 2V'\mu/\|V\|^2 + \|\mu\|^2/\|V\|^2] = [2(p-2) + 2\lambda]\varphi(p-2) - 1, \tag{3.2b}$$

$$x_3 = E_\mu[\|V\|^2 + 4(V'\mu)^2/\|V\|^2 + \|\mu\|^4/\|V\|^2 - 4V'\mu - 4\|\mu\|^2 V'\mu/\|V\|^2 + 2\|\mu\|^2] = [4(p-1)(p-2) + 8\lambda(p-2) + 4\lambda^2]\varphi(p-2) + 4 - 3p - 2\lambda. \tag{3.2c}$$

For $p = 3$, applying (3.3b) to (3.2a)–(3.2c) yields.

$$x_1 = \varphi(1), \qquad x_2 = 2\varphi(1) + \varphi(-1), \qquad x_3 = 8\varphi(1) + 4\varphi(-1) + 3\varphi(-3).$$

Applying (3.3d) to x_2/x_1 here yields $x_2/x_1 < 5$. From x_3/x_2 here, we obtain

$$\frac{x_3}{x_2} \le 4 + \frac{3\varphi(-3)}{\varphi(1)} < 13,$$

where the first inequality follows from $-1 \le \varphi(-1)/\varphi(1)$, and the second inequality follows from (3.3e).

For $p \ge 4$, applying (3.3c) to (3.2a) and (3.2b) yields $x_2 \le p x_1$. We also obtain

$$\begin{aligned} x_3 &\le [12(p-2) - 2\lambda(p-4)]\varphi(p-2) + (p-4) \\ &\le [12(p-2) - 2\lambda(p-4) + (p-4)(p-2+2\lambda)]\varphi(p-2) \qquad (3.2\text{d}) \\ &= (p-2)(p+8)x_1, \end{aligned}$$

where the first inequality follows by applying (3.3c) twice, and the second inequality follows by applying (3.3a). Coupled with the fact that $(p-2)x_1 \le x_2$ which follows from (3.3a), (3.2d) yields $x_3 \le (p+8)x_2$. □

The proofs of Lemmas 3.3, 3.4 and 3.5 have been omitted for the sake of brevity. These proofs can be found in George (1990).

LEMMA 3.3. *For a Poisson random variable Z with mean λ, define $\varphi(q) \equiv E[q + 2Z]^{-1}$ for $q \neq 0, -2, -4, \ldots$. Then*

(3.3a)	(i)	$1 \le (q + 2\lambda)\varphi(q)$	*for* $q \ge 1$,
(3.3b)	(ii)	$2\lambda\varphi(q+2) = 1 - q\varphi(q)$	*for* $q \neq 0, -2, -4, \ldots$,
(3.3c)	(iii)	$2\lambda\varphi(q+2) \le 1 - q\varphi(q+2)$	*for* $q \ge 0$,
(3.3d)	(iv)	$\varphi(-1) < 2\varphi(1)$,	
(3.3e)	(v)	$\varphi(-3) < 3\varphi(1)$.	

LEMMA 3.4. *For $V \sim N_p(\mu, I)$ and a Poisson random variable Z with mean $\lambda = \|\mu\|^2/2$,*

(i) $$E\|V\|^{-2} = E\left[\frac{1}{p - 2 + 2Z}\right],$$

(ii) $$E(V'\mu)\|V\|^{-2} = E\left[\frac{2Z}{p - 2 + 2Z}\right],$$

(iii) $$E(V'\mu)^2\|V\|^{-2} = 2\lambda - (p-1)E\left[\frac{2Z}{p - 2 + 2Z}\right].$$

LEMMA 3.5. *For A_3 given in* (2.7) *and a Poisson random variable Z with mean $\lambda = \|\theta\|^2/2\sigma_\eta^2$,*

$$A_3 = \frac{(\eta - (1-\eta)\lambda)^2\sigma^4}{\sigma_\eta^2} E\left[\frac{p-2}{p-2+2Z}\right].$$

Acknowledgments. The author would like to thank George Casella, Dean Foster, Jiunn Hwang and anonymous referees for helpful suggestions, and Jayaram Muthuswamy for programming assistance. This research was supported by the Graduate School of Business at the University of Chicago.

REFERENCES

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

BHATTACHARYA, A. G. (1980). Estimation of a common mean and recovery of interblock information. *Ann. Statist.* **8** 205–211.

BOX, G.E. and TIAO G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Mass.

BROWN, L. D. and COHEN, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2** 963–976.

CASELLA, G. and HWANG, J. T. (1982). Limit expressions for the risk of the James–Stein estimators. *Canad. J. Statist.* **10** 305–309.

CHIOU, W.-J. and COHEN, A. (1985). On estimating a common multivariate normal mean vector. *Ann. Inst. Statist. Math.* **37** 499–506.

COHEN, A. and SACKROWITZ, H. B. (1974). On estimating the common mean of two normal populations. *Ann. Statist.* **2** 1274–1282.

GEORGE, E. I. (1990). Shrinkage domination in a multivariate common mean problem. Technical Report 72, Graduate School of Business, Univ. Chicago.

KRISHNAMOORTY, K. (1989). Estimation of a common multivariate normal mean vector. Technical Report, Dept. Statistics, Temple Univ.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

LOH, W.-L. (1990). Estimating the common mean of two multivariate normal distributions. *Ann. Statist.* **19** 297–313.

STEIN, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics: Festschrift for Jerzy Neyman* (F. N. David, ed.) 351–366. Wiley, New York.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

YANCEY, T. A., JUDGE, G. G. and MIYAZAKI, S. (1984). Some improved estimators in the case of possible heteroscedasticity. *J. Econometrics* **25** 133–150.

ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637