ESTIMATORS FOR GAUSSIAN MODELS HAVING A BLOCK-WISE STRUCTURE

Lawrence D. Brown and Linda H. Zhao

University of Pennsylvania

Abstract: Many multivariate Gaussian models can conveniently be split into independent, block-wise problems. Common settings where this situation arises are balanced ANOVA models, balanced longitudinal models, and certain block-wise shrinkage estimators in nonparametric regression estimation involving orthogonal bases such as Fourier or wavelet bases.

It is well known that the standard, least squares estimate in multidimensional Gaussian models can often be improved through the use of minimax shrinkage estimators or related Bayes estimators. In the following we show that the traditional estimators constructed via independent shrinkage can be improved in terms of their squared-error risk, and we provide improved minimax estimators. An alternate class of block-wise shrinkage estimators is also considered, and fairly precise conditions are given that characterize when these estimators are admissible or quasi-admissible.

These results can also be applied to the classical Stein-Lindley estimator that shrinks toward an overall mean. It is shown how this estimator can be improved by introducing additional shrinkage.

Key words and phrases: ANOVA models, James-Stein estimators, harmonic priors, nonparametric estimation, quasi-admissibility, quasi-Bayes.

1. Introduction

Many multivariate Gaussian models can conveniently be split into disjoint, block-wise problems. Others can be written in such a form after a linear transformation. We examine here the situation in which the observations on each block are independent of those in the other blocks. Common settings where this situation arises are balanced ANOVA models, balanced longitudinal models, and certain block-wise shrinkage estimators in nonparametric regression estimation involving orthogonal bases such as Fourier or wavelet bases. We will describe a few such situations in Section 2.1.

It is well known that the standard, least squares estimate in multidimensional Gaussian models can often be improved through the use of minimax shrinkage estimators or related Bayes estimators that lead to well-motivated shrinkage estimation. In a situation involving statistically independent blocks it is natural to

apply this shrinkage separately within each block. From the Bayesian perspective this results from placing independent priors on the statistically independent blocks.

In the following we show that most common such estimators constructed via independent shrinkage are inadmissible in the original problem in terms of their squared-error risk, and we provide improved minimax estimators. In some situations blockwise estimators involving additional shrinkage may be admissible, and fairly precise conditions are given that characterize which of these estimators are admissible.

Many recent papers have produced partial characterizations of admissible Bayes procedures for structured multivariate problems. Consult Berger and Strawderman (1996), Berger and Robert (1990), and Berger, Strawderman and Tang (2005) for related studies and further references. For the block-wise case studied here, our results extend results from these previous studies. (Inadmissibility in our block-wise setting can be deduced using results in Berger and Strawderman (1996), but these do not describe suitable improved estimators.)

The Bayesian results are parallel to results for frequentistically motivated estimators related to the well-known James-Stein estimator. It is known that the original James-Stein estimator (James and Stein (1961)) and its positive-part version are not admissible. However in the canonical multivariate normal means problem involving a single block the positive part version is close to being admissible in a sense made precise in Brown (1988). Here, we use the term quasi-admissibility to refer to this property. We then show that in the block-wise case independent James-Stein type estimators on the individual blocks are not quasi-admissible, and we describe shrinkage estimators that are quasi-admissible, and improve on these in the sense of squared-error risk.

For the classical multivariate normal means estimation problem, Stein (1962) and Lindley (1962) proposed the use of a shrinkage estimator that shrinks toward the vector of overall means. In Example 2.1, below, we explain how such an estimator can be rewritten in the block-wise form. Theorem 3.1 thus applies to such an estimator, and shows how it can be improved by the introduction of additional shrinkage.

Section 2 contains background, motivations and examples. Commonly used shrinkage estimators in our setup can be improved by including an additional shrinkage factor, as is shown in Section 3. Section 4 includes other estimators which are admissible and minimax and may be considered as practical alternatives. We conclude the paper with comments.

Konno (1991) and Tsukuma and Kubokawa (2007) have demonstrated the advantage of additional shrinkage in a closely related setting. Their setting can be written in a blockwise form like ours but with equal size blocks. (Their problem

involves an unknown covariance matrix, in contrast with ours.) We allow unequal size blocks. Our Theorem 3.1 shows that additional shrinkage is advantageous even with such a block structure so long as there are at least two blocks and the overall dimension satisfies $p \geq 3$. In our Theorem 3.2 we characterize the desirable values for the additional shrinkage constant. This characterization involves the concepts of quasi-admissibility. In Section 4 we investigate estimators which shrink only within their respective blocks. In Theorems 4.1 and 4.2 we provide sufficient and nearly necessary conditions for such within block estimators to be both minimax and quasi-admissible.

2. Definitions and Preliminary Results

2.1. Canonical problem and estimators

Let Y be a p-variate homoscedastic normal random variable. We write

$$Y \sim N_p(\boldsymbol{\theta}, \sigma^2 \Sigma_Y). \tag{2.1}$$

The matrix Σ_Y is assumed known, with $\Sigma_Y = I$. There is some loss of generality in making this assumption, but generalizations appear to require additional research and are not pursued here. The scalar σ^2 is also assumed to be known, except in Remark 4.4.

We consider situations in which the vector $\boldsymbol{\theta} \in \mathfrak{R}^p$ can be partitioned as

$$\boldsymbol{\theta}^\tau = (\boldsymbol{\theta}_1^\tau, \ldots, \boldsymbol{\theta}_m^\tau)^\tau \quad \text{with} \quad \boldsymbol{\theta}_j \in \mathfrak{R}^{p_j}, \quad \sum_1^m p_j = p. \tag{2.2}$$

Let $\boldsymbol{Y}_{(1)}, \ldots, \boldsymbol{Y}_{(m)}$ denote the corresponding component sub-vectors of Y.

We do not exclude the case where *some* of the $p_j = 1$. Here are some particular examples that can be reduced to this canonical form.

Example 2.1. (Shrinkage to a common mean)

We begin with a setting formally different from that in (2.1) and reduce it to that form. In order to clarify this reduction, let us begin with a different notation. Thus, let

$$Z \sim N_p(\zeta, \sigma^2 I). \tag{2.3}$$

Assume $p \geq 4$.

Considerations in the papers of Stein (1962), Lindley (1962), and Efron and Morris (1972, 1973) suggest the following empirical Bayes estimator in the setting (2.3):

$$\tilde{\zeta}(z) = \overline{z}\mathbf{1} + \left(1 - \frac{(p-3)\sigma^2}{\|z - \overline{z}\mathbf{1}\|^2}\right)_+ (z - \overline{z}\mathbf{1}) \quad \text{where} \quad \overline{z} \stackrel{\triangle}{=} p^{-1}\sum z_i. \tag{2.4}$$

This estimator is essentially that given in Efron and Morris (1973, formula (7.1)). For more recent presentations of this estimator, see textbooks such as Berger (1985) or Lehmann and Casella (1998).

As an alternative presentation of (2.4) in blockwise form, let Q be an orthogonal matrix whose first row is $\mathbf{1}'/\sqrt{p}$. Then let

$$Y = QZ, \quad \theta = Q\xi.$$

Let $m = 2$, $p_1 = 1$, $p_2 = -1$. Then (2.4) can be re-written as (2.1)−(2.2) with $\theta_1 = \sqrt{p}\bar{\zeta}$. This corresponds to the following estimator $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2)^\tau$ of $\boldsymbol{\theta}$ in the setting (2.1)−(2.2):

$$\tilde{\boldsymbol{\theta}}_1 = \boldsymbol{y}_1, \quad \tilde{\boldsymbol{\theta}}_2 = \left(1 - \frac{(p-3)\sigma^2}{\|\boldsymbol{y}_2\|^2}\right)_+ \boldsymbol{y}_2. \tag{2.5}$$

Note that the estimator $\tilde{\boldsymbol{\theta}}_2$ is the James-Stein positive part estimator (2.4) within the space spanned by $\boldsymbol{Y}_2$. See (2.13), below. (Efron and Morris do not add the positive-part restriction to the formula corresponding to (2.5) in their paper, but they do discuss elsewhere the desirability of such a restriction. Lindley and Smith discuss the desirability of using an estimator of a general form like (2.5), but do not explicitly propose the constant $p - 3$ or the positive-part restriction.) See Remark 4.2 for further discussion of this example.

Example 2.2. (Balanced one way and two way additive ANOVA)

A convenient notation for a two way ANOVA model has

$$Z_{ij} \sim N_{IJK}(\zeta_{ij}, \sigma^2), \quad i = 1, \ldots, I, \; j = 1, \ldots, J, \; k = 1, \ldots, K, \quad \text{where}$$
$$\zeta_{ij} = \mu + \alpha_i\sqrt{JK} + \beta_j\sqrt{IK} \text{ and } \sum \alpha_i = \sum \beta_j = 0. \tag{2.6}$$

(The normalization of α_i, β_j here is not standard, but is adopted to simplify certain expressions to follow.) Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)'$, and similarly for $\boldsymbol{\beta}$. Then we may identify three blocks in the model (2.1), with $\theta_1 = \mu$, $\theta_2 = M_2\alpha$, $\theta_3 = M_3\beta$ and $\boldsymbol{Y}_1 = \hat{\mu}$, $\boldsymbol{Y}_2 = M_2\hat{\boldsymbol{\alpha}}$, $\boldsymbol{Y}_3 = M_3\hat{\boldsymbol{\beta}}$ where M_2, M_3 are $(I-1) \times I$ and $(J-1) \times J$ matrices, respectively, and the rows of these matrices are orthonormal and are orthogonal to $\mathbf{1} = (1, \ldots, 1)'$. $\hat{\mu}$, $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$ are the usual estimators. Here $p = I + J - 1$. Assume $I \geq 4$, $J \geq 4$ in what follows.

Blockwise shrinkage estimator can be applied within this setting. When the blockwise shrinkage estimator is transformed back to the original coordinates it corresponds to

$$\tilde{\theta}_1 = \tilde{\mu} = \boldsymbol{Y}_1, \quad \tilde{\boldsymbol{\alpha}} = \left(1 - \frac{I-3}{\|\hat{\boldsymbol{\alpha}}\|^2}\sigma^2\right)_+ \hat{\boldsymbol{\alpha}}, \quad \tilde{\boldsymbol{\beta}} = \left(1 - \frac{J-3}{\|\hat{\boldsymbol{\beta}}\|}\sigma^2\right)_+ \tilde{\boldsymbol{\beta}}. \tag{2.7}$$

Again, see Remark 4.2 for further discussion of the example.

We also have a similar set up for a one way ANOVA.

Example 2.3. (Blockwise models for nonparametric regression)

Blockwise James-Stein estimators for nonparametric regression have been studied by many authors. See for example Donoho and Johnstone (1995) Cai (1999), Cavalier and Tsybakov (2001), Cai, Low and Zhao (2000), and Brown, Zhao and Mao (2004). In this approach a standard nonparametric regression problem involving n observations is transformed by an orthogonal transformation to a problem of the form (2.1) with $p = n$. (Wavelet transformations or Fourier transforms are both suitable possibilities.) A blockwise structure is then imposed where the dimensions of the blocks increase monotonically. The James-Stein estimator, or some similar shrinkage estimator, is then computed and back-transformed in order to yield an estimator in the original nonparametric regression problem. See the above references and others mentioned there for further details.

2.2. Bayes estimators

Consider the setting (2.1) under ordinary quadratic loss, $L(\theta, d) = \|\theta - d\|^2$, and corresponding risk, $R(\theta, \delta) = E_\theta(L(\theta, \delta(Y)))$. Of particular interest to us are minimax estimators; that is, those satisfying

$$R(\theta, \delta_m) \leq p\sigma^2 = \inf_\delta \sup_\theta R(\theta, \delta) = R(\theta, \delta_0) \quad \text{where} \quad \delta_0(y) = y.$$

Brown (1971) shows that every admissible estimator is generalized Bayes. When $p \geq 3$ many admissible minimax estimators are known. Among these the harmonic Bayes estimator is particularly simple and appealing. It corresponds to the prior density

$$g_{harm;p}(\theta) \stackrel{\triangle}{=} g_{harm}(\theta) \stackrel{\triangle}{=} \frac{1}{\|\theta\|^{(p-2)}}, \tag{2.8}$$

and has a simple, explicit functional form. This is the marginal density (over θ) of a hierarchical prior under which $\theta | \tau \sim N_p(0, \tau^2 I)$ and the hyperparameter $\tau > 0$ has density τ. The functional form for the Bayes procedure can be derived from this representation. See Strawderman (1971), Lin and Tsai (1973), Kubokawa (1991), and Brown, Zhao and Mao (2004).

2.3. Quasi-admissible and quasi-Bayes estimators

Let δ be a given estimator. Throughout this paper we use the symbol γ to denote

$$\gamma(y) = \delta(y) - y. \tag{2.9}$$

Hence an estimator can be defined through the formula for δ or for γ.

An estimator is termed "regular" if each coordinate function γ_i is absolutely continuous in the corresponding coordinate y_i of y, and

$$E_\theta\left(\|\nabla \cdot \gamma(Y)\|\right) < \infty \ \text{ and } \ E_\theta\left(\|\gamma(Y)\|^2\right) < \infty,$$

where $\nabla \cdot \gamma$ is the gradient function over γ. (It is a consequence of results in Brown (1971) that one need only consider regular estimators since they form a statistically complete class. That is, no non-regular estimator can be admissible, and any non-regular estimator is dominated in risk by a regular one.) All estimators that appear in the following are assumed to be regular.

Stein's unbiased estimate of risk (Stein (1973, 1981)) applies to any regular estimator. It says that

$$R(\theta, \delta) = E_\theta\left(p\sigma^2 + 2\sigma^2 \nabla \cdot \gamma(Y) + \|\gamma(Y)\|^2\right) \doteq E_\theta(\hat{R}_\delta(Y)), \qquad (2.10)$$

where $\hat{R}_\delta$ denotes the unbiased estimate of risk. This leads to a heuristic theory of "quasi-admissibility". The heuristic justification for this theory lies in the observation that Y is near θ, and hence the expectand on the right of (2.10) at $Y = \theta$ is an approximation to the true risk at θ. Thus,

$$R(\theta, \delta) \approx \hat{R}_\delta = p\sigma^2 + 2\sigma^2 \nabla \cdot \gamma(\theta) + \|\gamma(\theta)\|^2. \qquad (2.11)$$

Correspondingly, we define an estimator $\delta_{(1)}$ to be "quasi-admissible" if there is no estimator $\delta_{(2)}$ such that

$$2\sigma^2 \nabla \cdot \gamma_{(2)}(y) + \|\gamma_{(2)}(y)\|^2 \leq 2\sigma^2 \nabla \cdot \gamma_{(1)}(y) + \|\gamma_{(1)}(y)\|^2 \qquad (2.12)$$

for all y, with strict inequality for at least one value of y. Because of (2.10) there is a relationship between quasi-admissibility and ordinary admissibility; namely, any admissible estimator is quasi-admissible. (Similarly, any quasi-inadmissible estimator is inadmissible.) But the converse assertion is not, in general, true.

Experience has shown that estimators that are quasi-admissible are also numerically close to being admissible, although the definitions do not formally guarantee such a numerical property. The positive-part James and Stein (1961) estimator is the canonical example for the preceding assertions. It is defined for $p \geq 3$ as

$$\delta_{JS+}(y) = \left(1 - \frac{p-2}{\|y\|^2}\sigma^2\right)_+ y. \qquad (2.13)$$

See Maruyama (1999, 2004) and also Brown, Zhao and Mao (2004) for results showing that this estimator is indeed numerically very close to being admissible.

This estimator is not admissible but it is quasi-admissible, as will be further discussed below.

Brown (1971) shows that the (generalized) Bayes estimators are a complete class with respect to ordinary admissibility, and that every such procedure can be written in the form (2.9) with

$$\gamma^* = \sigma^2 \frac{\nabla g^*}{g^*},\tag{2.14}$$

where $g^*(y) = \int_{\Re^p} \varphi(y-\theta)G(d\theta)$. (Note that g^* is the marginal density of Y.) There is a corresponding characterization in relation to quasi-admissibility. Any regular estimator will be termed "quasi-Bayes" if it can be written in the form

$$\gamma^\circ = \sigma^2 \frac{\nabla g^\circ}{g^\circ}\tag{2.15}$$

for some absolutely continuous function g°. Note that the estimator δ_{JS+} is quasi-Bayes with the corresponding marginal quasi-density being

$$g_{JS+;p}(y) = \frac{e^{((p-2)\sigma^2 - \|y\|^2)/2\sigma^2}}{((p-2)\sigma^2)^{(p-2)/2}} I_{\{\|y\|^2 > (p-2)\sigma^2\}}(y) \frac{1}{\|y\|^{p-2}}.\tag{2.16}$$

(It is observed in Brown (1971) that the estimator δ_{JS+} cannot be (generalized) Bayes since the function $g_{JS+;p}$ is not analytic, and hence cannot be the marginal density corresponding to any prior measure.)

Brown (1988) developed a general theory of quasi-admissibility. (In that paper he did not use a term such as "quasi-admissibility" for the relevant concept. Later authors have used various terms for it. For example Rukhin (1995) uses the term permissible, and Bock (1988) uses the term pseudo-Bayes for a related construct. We feel that quasi-admissibility is a more expressive term for this concept, and avoids some confusion with other uses of possible alternate terminology.)

Brown (1988) shows that any quasi-admissible estimator must be quasi-Bayes. Furthermore, Brown (1988, formula (7.2)) gives a condition that implies a quasi-Bayes procedure corresponding to g is quasi-admissible. Here is a formal statement of that result.

Lemma 2.1. *Let $O_r(\cdot)$ denote the uniform measure on the sphere in $\Re^p$ of radius r. Then let*

$$\overline{m}_g(r) \triangleq \int g(y)O_r(dy).$$

If the corresponding γ given by (2.15) is bounded and if

$$\int_2^\infty \left(r^{p-1}\overline{m}_g(r)\right)^{-1} dr = \infty,\tag{2.17}$$

then it is quasi-admissible.

It is shown in Brown (1971) that if a Bayes estimator satisfies (2.17) then it is admissible. It follows from (2.16)−(2.17) that the positive-part James-Stein estimator is quasi-admissible.

2.4. Blockwise estimators when $m \geq 2$

When $m \geq 2$ it is natural to propose an estimator composed of the preceding estimators within each block. In order to provide a unified statement of this proposal, let us by convention define the harmonic prior when $p \leq 2$ as having the uniform density, and make a similar convention for the James-Stein estimator, so that we write $\delta_{Harm}(y) = \delta_{JS+}(y) = y$ when $p \leq 2$.

Then, in the setting (2.1)−(2.2), the blockwise $JS+$ estimator is $\delta_{JS+}^{block}(\theta) = (\tilde{\boldsymbol{\theta}}_1^\tau, \ldots, \tilde{\boldsymbol{\theta}}_m^\tau)^\tau$ where

$$\tilde{\boldsymbol{\theta}}_j = \delta_{JS+}(\boldsymbol{Y}_j), \quad j = 1, \ldots, m. \tag{2.18}$$

The estimator δ_{Harm}^{block} is defined analogously, with each blockwise component being the Bayes estimator with respect to the harmonic prior restricted to that block. Consequently, this estimator can also be described as the Bayes estimator with respect to the product across blocks of harmonic priors. The corresponding combined density can be written

$$g_{harm}^{block}(\theta) = \prod g_{Harm}(\theta_j). \tag{2.19}$$

These estimators are also very natural within the context of examples described above. For example, within the setting of Example 2.1, the estimator δ_{JS+}^{block} is exactly (2.19) and is hence equivalent to the Efron-Morris estimator described in (2.4).

3. Minimaxity and (quasi)-admissibility

3.1. Minimaxity and inadmissibility of standard blockwise estimators

The blockwise estimators described in Section 2.4 are minimax, as formally stated in the following proposition.

Proposition 3.1. *Both δ_{JS+}^{block} and δ_{harm}^{block} are minimax.*

Proof. This follows directly from the fact that the blockwise components of these estimators are minimax within their respective blocks.

These blockwise estimators are intuitively natural. They are also minimax, as shown above. Hence it may be surprising to some that when $m \geq 2$ they are not admissible in appropriate senses. However, the following theorem and

corollaries show that when $m \geq 2$ these estimators are not even quasi admissible, and describe simple estimators that dominate each of them.

Theorem 3.1. *Let δ_B be a blockwise estimator of the form*

$$\delta_B(Y) = \tilde{\theta}_B = (\tilde{\boldsymbol{\theta}}_{B1}^{\tau}, \ldots, \tilde{\boldsymbol{\theta}}_{Bm}^{\tau})^{\tau} \text{ where } \tilde{\boldsymbol{\theta}}_{Bj} = \left(1 - \frac{r_j(\|\boldsymbol{Y}_j\|)}{\|\boldsymbol{Y}_j\|^2} \sigma^2 \right) \boldsymbol{Y}_j,$$

and assume $0 \leq r_j(v) \leq R_j$.

Assume $p \geq 3$ and $R \overset{\triangle}{=} \sum_{j=1}^{m} R_j < p - 2$. Then δ_B is not quasi-admissible and a better estimator is

$$\delta(y) = \delta_b(y) - \sigma^2 \frac{c(\|y\|^2)}{\|y\|^2} y, \tag{3.1}$$

where c is a non-decreasing absolutely continuous function that satisfies $0 < c < 2(p - 2 - R)$. In particular, a better estimator is given by

$$\delta^c(y) = \delta_B(y) - \left(\frac{c\sigma^2}{\|y^2\|} \wedge 1 \right) y, \tag{3.2}$$

where $0 < c < 2(p - 2 - R)$.

Proof. If $\|y\|^2 > c\sigma^2$, the difference in the unbiased estimators of risk is

$$\frac{1}{\sigma^4} \left[\hat{R}_{\delta_B} - \hat{R}_{\delta} \right]$$

$$= \frac{1}{\delta^4} \left\{ 2\sigma^2 (\nabla \cdot \gamma_B(y)) + \|\gamma_B(y)\|^2 - \left[2\sigma^2 (\nabla \cdot \gamma(y)) + \|\gamma(y)\|^2 \right] \right\}$$

$$= 2 \left(\nabla \cdot \frac{c}{\|y\|^2} y \right) + \left\| \sum \left(\frac{-r_j(\|\boldsymbol{y}_j\|)}{\|\boldsymbol{y}_j\|^2} \boldsymbol{y}_j \right) \right\|^2 - \left\| \sum \left(\frac{-r_j(\|\boldsymbol{y}_j\|)}{\|\boldsymbol{y}_j\|^2} \boldsymbol{y}_j - \frac{c}{\|y\|^2} \boldsymbol{y}_j \right) \right\|^2$$

$$= \frac{1}{\|y\|^2} \left\{ 2c(p - 2) + 4c'(\|y\|^2)\|y\|^2 - 2c \sum r_j(\|\boldsymbol{y}_j\|) - c^2 \right\}$$

$$\geq \frac{c}{\|y\|^2} \left\{ 2(p - 2) - 2R - c \right\} > 0. \tag{3.3}$$

The positivity in (3.3) shows that δ is not quasi-admissible. Formula (3.2) is an immediate consequence of (3.1).

Corollary 3.1. *Let $m \geq 2$. Then neither δ_{JS+}^{block} nor δ_{harm}^{block} are quasi-admissible. Let δ_B denote either δ_{JS+}^{block} or δ_{harm}^{block} in Section 2.4, respectively. Let $m \geq 2$, $p \geq 3$, and $p^{\#} \overset{\triangle}{=} \sum (p_j - 2)_+$. Then $p^{\#} < p - 2$ and δ^c in (3.1) dominates δ_{JS+}^{block} or δ_{harm}^{block}, respectively, in the sense of quasi-admissibility, where*

$$0 < c < 2(p - 2 - p^{\#}). \tag{3.4}$$

(*See Remark* 3.1, *for further comments about the choice of c.*)

Proof. For $\delta JS+^{block}$ it is trivially true that $R_j = (p-2)_+$. References of Strawderman and of Kubokawa cited above each show that the Harmonic prior estimator satisfies the hypothesis of the theorem with $R_j = (p-2)_+$. The corollary then follows directly from the theorem.

3.2. Quasi-admissible minimax estimators

For suitable choices of c the respective estimators δ^c are quasi-admissible, in addition to being minimax and dominating δ_{JS+}^{block} and δ_{harm}^{block}.

Theorem 3.2. *Let $m \geq 2$. Let δ_B denote either δ_{JS+}^{block} or δ_{harm}^{block} in Section 2.4, respectively. Then the corresponding δ^c in (3.2) is quasi-admissible if*

$$c \geq p - 2 - p^{\#}. \tag{3.5}$$

Proof. Consider first the case of $JS+$. In order to provide a unified notation, if $p = 1, 2$ let $g_{JS+;p}(y) \overset{\triangle}{=} 1$. Then δ_{JS+}^{block} is quasi-Bayes with respect to the quasi-prior $g(y) = \prod g_{JS+;p_j}(\boldsymbol{y}_j)$ defined in (2.16). The estimator δ^c is then quasi-Bayes with respect to the quasi-prior

$$g_c(y) = h_c(y) \prod g_{JS+;p_j}(\boldsymbol{y}_j), \tag{3.6}$$

where

$$h_c(y) = \frac{e^{c/2}}{\|c\sigma^2\|^{c/2}} I_{\{\|y\|^2 \leq c\sigma^2\}}(y) e^{-\|y\|^2/2\sigma^2} + I_{\{\|y\|^2 > c\sigma^2\}}(y) \frac{1}{\|y\|^c}. \tag{3.7}$$

The components of (3.6) satisfy

$$g_{JS+;p_j}(\boldsymbol{y}_j) = O\left(\frac{1}{1 + \|\boldsymbol{y}_j\|^{(p_j-2)_+}}\right), \quad h_c(y) = O\left(\frac{1}{1 + \|y\|^c}\right). \tag{3.8}$$

Hence

$$\overline{m}_{g_c}(r) = O\left(\int \frac{1}{1 + \|y\|^c} \prod \frac{1}{1 + \|y_j\|^{(p_j-2)_+}} O_r(dy)\right). \tag{3.9}$$

It follows from Lemma A.1 in the appendix that

$$\overline{m}_g(r) = O\left(\frac{1}{1 + r^{c + \sum(p_j-2)_+}}\right).$$

Then, (2.17) shows that δ^c is quasi-admissible.

The verification in the case of δ_{Harm}^{block} is similar. The references of Strawderman and of Kubokawa cited above also contain expressions for the value of

g_{Harm}^*; see also Brown and Zhao (2006, equation 2−3(41)). These show that g_{Harm}^* also satisfies (3.8); that is,

$$g_{Harm}^*(\boldsymbol{y}_j) = O\left(\frac{1}{1 + \|\boldsymbol{y}_j\|^{(p_j-2)_+}}\right).$$

The remainder of the proof for δ_{Harm}^{block} follows exactly the pattern of that for δ_{JS+}^{block}.

Remark 3.1(Choice of c). The preceding results provide a range of values for the constant c. Thus any choice within the range

$$p - 2 - p^\# \leq c < 2(p - 2 - p^\#) \tag{3.10}$$

will yield a quasi-admissible estimator dominating δ_{JS+}^{block} or δ_{harm}^{block}, respectively. The choice $c = p - 2 - p^\#$ would be the traditionally motivated choice. It has the property that among all estimators satisfying (3.10) it provides the maximum improvement for large values of $\|\theta\|$. Larger choices of c could provide more improvement for small $\|\theta\|$ at the cost of doing slightly worse as $\|\theta\| \to \infty$.

Remark 3.2. The preceding results suggest a *possible* form for an admissible minimax estimator dominating δ_{Harm}^{block}. It must be generalized Bayes, and the form of (3.6) then suggests **trying** a prior of exactly or approximately the form

$$g(\boldsymbol{\theta}) = \frac{1}{\|\boldsymbol{\theta}\|^c} \prod g_{Harm;p_j}(\boldsymbol{\theta}_j). \tag{3.11}$$

(This prior results from a hierarchical Bayes construction in which the hyperparameters are $\omega_0, \ldots, \omega_m$ and, conditional on these hyperparameters, $\boldsymbol{\theta}_j \sim N_{p_j}(0, 1/(\omega_0 + \omega_j I_{\{p_j \geq 3\}}(p_j)))$, independent, and the hyperparameters have a suitable joint density.

   The form (3.11) automatically yields admissibility as a consequence of (2.17). The density (3.11) is superharmonic, and hence its Bayes procedure is a minimax procedure. (See Stein (1981).) We have not been able to determine whether this Bayes procedure dominates δ_{Harm}^{block}.

4. Other Minimax Forms and Additional Remarks

4.1. A different, blockwise estimator

   The preceding results show that the natural blockwise estimators are not (quasi) admissible and propose alternate, improved estimators. These improved estimators are not of a blockwise form, but rather "borrow strength" by combining information across blocks. Indeed, no strictly blockwise estimator can improve on the estimators of Section 2.4 in the appropriate sense of quasi-admissibility for δ_{JS+}^{block} and of admissibility for δ_{harm}^{block}. This is because these estimators have the relevant admissibility property within blocks. However, as long

as $p_j \geq 3$, $j = 1, \ldots, m$, there are blockwise estimators that are (quasi) admissible and minimax. In some cases such estimators also exist when there are one or two less than or equal to 2. These estimators have risk functions moderately close to those of δ_{JS+}^{block} or δ_{harm}^{block}, although they do not dominate them. Throughout the following assume for notational convenience, and without loss of generality, that

$$p_1 \leq \cdots \leq p_m \tag{4.1}$$

Here are appropriate blockwise minimax estimators. Let c_j, $j = 1, \ldots, m$ be non-negative constants such that

$$(p_j - 2)_+ \leq c_j \leq 2(p_j - 2)_+. \tag{4.2}$$

Define the estimator on the jth block as

$$\delta_+^*(\boldsymbol{y}_j) = \left(1 - \frac{c_j}{\|\boldsymbol{y}_j\|^2}\sigma^2\right)_+ \boldsymbol{y}_j. \tag{4.3}$$

Then define the blockwise estimator δ_{JS+}^{*block} in the natural way via

$$\tilde{\boldsymbol{\theta}}_j^* = \delta_+^*(Y_j), \quad j = 1, \ldots, m. \tag{4.4}$$

As an alternative to δ_{Harm}^{block}, begin by defining a prior on each block. Let

$$g_{Str}(\boldsymbol{\theta}_j) = \int_0^\infty \left(\frac{w}{1+w}\right)^{c_j - (p_j - 2)} w^{(p_j - 4)/2} e^{-w\|\boldsymbol{\theta}_j\|^2/2} dw \;\; \text{if} \;\; p_j \geq 3,$$
$$= 1 \;\; \text{if} \;\; p_j = 1, 2. \tag{4.5}$$

The notation recalls that this prior, along with a formula for the estimator, is given in Strawderman (1971). (The parameter a in Strawderman's formula (9) is related to our c_j via $a = (p_j + 2 - 2c_j)/2$; some additional calculation is needed in order to verify that (4.5) is the density of $\boldsymbol{\theta}_j$ corresponding to Strawderman's prior (9).) Now, let δ_{Bayes}^{*block} denote the Bayes estimator with respect to $\prod g_{Str}(\boldsymbol{\theta}_j)$.

The estimators δ_{JS+}^{*block} and δ_{Bayes}^{*block} are minimax because they are minimax within each block. (In the case of δ_{Bayes}^{*block} one needs to note that the minimaxity assertion in Strawderman (1971, "Theorem") is valid so long as

$$\frac{p_j}{2} + 1 > a \geq 3 - \frac{p_j}{2}. \tag{4.6}$$

This follows from noting that Strawderman's "Lemma" applies to the expression in his $(11)-(12)$ so long as (4.6) is satisfied. See also Strawderman (1973, Remark 4). The following theorem gives conditions under which the estimators in (4.3) are quasi-admissible or admissible.

Theorem 4.1. *Let $\{c_j\}$ be as in (4.2). Assume*

$$\sum(c_j \wedge p_j) > p - 2 \ or$$
$$\sum(c_j \wedge p_j) = p - 2 \ and \ at \ most \ one \ value \ of \ c_j = p_j. \qquad (4.7)$$

Then the estimators δ_{JS+}^{*block} and δ_{Bayes}^{*block} defined in (4.3)−(4.4) and below (4.5) are each minimax, and they are quasi-admissible and admissible, respectively.

Proof. As already noted, minimaxity follows from the fact that each blockwise component of the estimators is minimax.

For the quasi-admissibility of δ_{JS+}^{*block} note that, as in (3.8), the quasi-prior corresponding to (4.1) satisfies

$$g_{*block}(\boldsymbol{y}_j) = O\left(\frac{1}{1 + \|\boldsymbol{y}_j\|^{c_j}}\right).$$

Lemma A.2 then applies and shows that the conditions on $\{c_j\}$ imply that the quasi-prior corresponding to δ_{JS+}^{*block} satisfies

$$\overline{m}_{g_{*block}}(r) = O\left(\int \prod \frac{1}{1 + \|\boldsymbol{y}_j\|^{c_j}} O_r(dy)\right) = O\left(\frac{\log(2 + r)}{1 + r^{p-2}}\right). \qquad (4.8)$$

As in (3.9) this implies quasi-admissibility according to (2.17).

The verification of admissibility for δ_{Bayes}^{*block} is similar, again based on (2.17). The key fact is contained in Strawderman (1971) where it is shown that the marginal distribution g^* corresponding to this prior also satisfies the first equality in (4.8).

The following is a partial converse to this theorem. As noted in Remark 4.1 below, there is only a thin gap between the theorem and the following converse.

Theorem 4.2. *Let $\{c_j\}$ be as in (4.2). Assume*

$$\sum(c_j \wedge p_j) < p - 2. \qquad (4.9)$$

Then the estimators δ_{JS+}^{*block} and δ_{Bayes}^{*block} defined in (4.3)−(4.4) and below (4.5) are each minimax and they are quasi-inadmissible. (Hence they are each also inadmissible in the ordinary sense.)

Proof. This statement is really a corollary of Theorem 3.1. Let $U \overset{\triangle}{=} \{j : c_j < p_j\}$. Without loss of generality we may assume for the following that $U = \{1, \ldots, J_U\}$. Consider the problem involving only the blocks with $j \in U$ (ie, with $j = 1, \ldots, J_U$). It follows from (4.9) that $\sum_{j \in U} c_j < \sum_{j \in U}(p_j - 2)_+$. Then

Theorem 3.1 implies that the estimators are not quasi-admissible when restricted to the blocks with $j \in U$. Hence they are not quasi-admissible.

4.2. General remarks

Remark 4.1. When the values of $\{p_j\}$ allow choices of $\{c_j\}$ for which the conditions of Theorem 4.1 are satisfied, then there exist blockwise estimators that are minimax and (quasi)admissible. This happens for many configurations of the $\{p_j\}$, and the condition (4.7) of the theorem is easy to check in any specific case.

Conversely, there are situations for which Theorem 4.2 does not allow choices of the $\{c_j\}$ under which the corresponding estimators δ_{JS+}^{*block} and δ_{Bayes}^{*block} are minimax and (quasi)admissible. It seems quite plausible to conjecture that in such cases there does not exist a blockwise estimator that is minimax and (quasi)admissible. However we do not know whether this conjecture is valid.

There are only a few configurations of $\{p_j\}$ for which Theorem 4.1 does not establish the existence of a blockwise minimax and (quasi)admissible estimator of the form of δ_{JS+}^{*block} or δ_{Bayes}^{*block} and, at the same time, Theorem 4.2 does not show that no such estimator is both minimax and (quasi)admissible. Under the ordering constraint (4.1), these configurations are those beginning with $p_1, p_2 = 2, 4$, or with $p_1, p_2, p_3 = 1, 1, 4$ or $1, 3, 4$ and containing at least two values of $p_j = 4$. A natural proposal for such a situation would be to use δ_{JS+}^{*block} or δ_{Bayes}^{*block} with $c_j = 2(p_j - 2)_+$. Such an estimator would be minimax. Theorem 4.1 does not guarantee that it is (quasi)admissible, nor does Theorem 4.2 show that it is not. However, even if it is not (quasi)admissible it is arbitrarily close to being so. This is because any such estimator with $c_j = 2(p_j - 2)_+(1 + \varepsilon)$ would be (quasi)admissible and have risk function uniformly close to that of the estimator with $c_j = 2(p_j - 2)_+$.

Remark 4.2. (about Examples 2.1 and 2.2) In the context of Example 2.1 with $p \geq 4$, the choice $c_1 = 0$, $p - 2 \leq c_2 \leq 2(p - 3)$ in δ_{JS+}^{*block} leads to

$$\tilde{\zeta}^*(z) = \overline{z}\mathbf{1} + \left(1 - \frac{c_2}{\|z - \overline{z}\mathbf{1}\|^2}\sigma^2\right)_+ (z - \overline{z}\mathbf{1}). \tag{4.10}$$

Note that when $p = 4$, then the only suitable choice yielding both minimaxity and quasi-admissibility is $c_2 = 2$. Otherwise there is a range of possible values of c_2 for which both (4.2) and (4.7) are valid, namely, $p - 2 \leq c_2 \leq 2(p - 3)$.

In the balanced two way additive model of Example 2.2, the conditions (4.2) and (4.7) can be satisfied whenever $I \geq 5$, $J \geq 5$, so that $p_1 = 1$, $p_2 = I - 1 \geq 4$, $p_3 = I - 1 \geq 4$. In such a case one may choose $c_1 = 0$, $c_2 = I - 1$, $c_3 = J - 1$. The

resulting estimator δ_{JS+}^{*block} in terms of the parameters of the model corresponds to

$$\tilde{\mu} = \hat{\mu}, \quad \tilde{\boldsymbol{\alpha}} = \left(1 - \frac{c_2}{\|\hat{\boldsymbol{\alpha}}\|^2}\sigma^2\right)_+ \hat{\boldsymbol{\alpha}}, \quad \tilde{\boldsymbol{\beta}} = \left(1 - \frac{c_3}{\|\hat{\boldsymbol{\beta}}\|^2}\sigma^2\right)_+ \hat{\boldsymbol{\beta}}. \qquad (4.11)$$

(Recall that the row and column effects satisfy the side condition of adding to zero.) If $I \leq J$ (without loss of generality) the conditions can also be satisfied if $I \geq 4$, $J \geq 5$, so that $p_2 = I - 1 \geq 3$, $p_3 = J - 1 \geq 4$. In the case $I = 4$, $J = 5$, the only suitable choice is $c_2 = 2$, $c_3 = 4$; otherwise there is some flexibility in the choice. In situations not included in the above comments the only available choices for $\{c_j\}$ satisfy (4.9), and hence the resulting estimators cannot be both minimax and (quasi)admissible.

Remark 4.3(choice of $\{c_j\}$). As in the previous settings, there is usually a variety of possible choices for $\{c_j\}$ satisfying (4.2). When it is desired to use these blockwise estimators, we recommend the choice $c_j = p_j$ whenever this satisfies (4.2) and (4.7). This choice yields estimators whose risk is not far from that of the usual James-Stein+ choice of $c_j = (p_j - 2)_+$ within each block, and yet can yield an estimator that is minimax and quasi-admissible overall.

Remark 4.4.(unknown, estimable error variance) In all the above it has been assumed that σ^2 is a known constant. It is much more common in applications for σ^2 to be an unknown parameter, and to observe an independent variable $V \ni V/\sigma^2 \sim \chi_d^2$. In that case the value of V/d estimates σ^2. This estimate can be plugged into the preceding formulas in place of the value of σ^2. It is known that such a plug-in procedure yields minimaxity in the original $JS+$ estimator so long as either d is not too small or suitable minor changes are incorporated in the shrinkage constant. See James and Stein (1961). This same type of result will hold for all the modified James-Stein estimators considered above. It appears that a similar plug-in rule will generally also be minimax for the generalized Bayes estimators considered above, though again one should expect that some not too stringent conditions on d will be needed. It is much more problematic to satisfactorily generalize the results about admissibility and (quasi)admissibility to the situation with unknown but estimable σ^2. Since we are as yet unable to do so we do not present further treatment of this problem here. Strawderman (1973) describes some Bayes and generalized Bayes minimax estimators for this setting, but does not establish admissibility for the natural extensions of the harmonic Bayes estimator. More recent work is also relevant to this problem as well as to the situation involving heteroscedasticity in (2.1). See for example Maruyama and Strawderman (2005) and other references cited therein.

Acknowledgement

Appendix. Technical Lemmas

Lemma A.1. *Let $y \in \mathfrak{R}^p$ denote a general vector $y = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_m)'$, $m \geq 2$, partitioned as in (2.2) and Theorem 3.2. Let q_j, $j = 1, \ldots, m$, be non-negative constants satisfying $q_j < p_j$, $j = 1, \ldots, m$. Then*

$$\frac{1}{1 + r^c} \int \prod \frac{1}{1 + \|\boldsymbol{y}_j\|^{q_j}} O_r(dy) = O\left(\frac{1}{r^{c + \sum q_j}}\right). \tag{A.1}$$

Proof. If $y \in \mathfrak{R}^p$ is distributed according to the uniform distribution, O_r, on the sphere of radius r, then $\{\|\boldsymbol{y}_1\|^2/r^2, \ldots, \|\boldsymbol{y}_m\|^2/r^2\}$ has the Dirichlet $(p_1/2, \ldots, p_m/2)$ distribution. Let $D(\cdot)$ denote this distribution concentrated on the set $H_1 \overset{\triangle}{=} \{\xi \in \mathfrak{R}^m : \xi_j \geq 0, j = 1, \ldots, m; \sum \xi_j = 1\}$. Then

$$\int \prod \frac{1}{1 + \|\boldsymbol{y}_j\|^{q_j}} O_r(dy) = r^{-\sum q_j} \int \prod \frac{1}{r^{-q_j} + \xi_j^{q_j/2}} D(d\xi_1 \cdots d\xi_m)$$

$$= r^{-\sum q_j} O\left(\int_{\xi \in H_1} \prod \frac{1}{\xi_j^{q_j/2}} \prod \xi_j^{(p_j - 2)/2} d\xi\right)$$

$$= r^{-\sum q_j} O\left(\int_{\xi \in H_1} \prod \frac{1}{\xi_j^{q_j/2}} \prod \xi_j^{(p_j - 2)/2} d\xi\right)$$

$$= O\left(r^{-\sum q_j} \int_{\xi \in H_1} \prod \xi_j^{(p_j - q_j - 2)/2} d\xi\right) = O\left(r^{-\sum q_j}\right). \tag{A.2}$$

In the preceding we have used the fact that $\int_{\xi \in H_1} \prod \xi_j^{(p_j - q_j - 2)/2} d\xi < \infty$ since $q_j < p_j$, $j = 1, \ldots, m$. The conclusion (A.1) now follows directly from (A.2).

Lemma A.2. *Let $y \in \mathfrak{R}^p$ denote a general vector $y = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_m)'$, $m \geq 2$, partitioned as in (2.2) and Theorem 3.2. Let q_j, $j = 1, \ldots, m$, be non-negative constants. Let $r \geq 2$ and*

$$\Upsilon = \sum (q_j \wedge p_j), \quad \Lambda = \#\{j : q_j = p_j\}. \tag{A.3}$$

Then

$$\int \prod \frac{1}{1 + \|\boldsymbol{y}_j\|^{q_j}} O_r(dy) = O\left(\frac{(\log r)^\Lambda}{r^\Upsilon}\right). \tag{A.4}$$

Proof. As in the previous lemma,

$$\int H \overset{\triangle}{=} \prod \frac{1}{1 + \|\boldsymbol{y}_j\|^{q_j}} O_r(dy)$$

$$= r^{-\sum q_j} \int \prod \frac{1}{r^{-q_j} + \xi_j^{q_j/2}} D(d\xi_1 \cdots d\xi_m)$$

$$= O\left(\int_{\xi \in H_1} \prod \frac{r^{-q_j}}{r^{-q_j} + \xi_j^{q_j/2}} \prod \xi_j^{(p_j-2)/2} d\xi \right)$$

$$= O\left(\int_{\xi \in H_1} \prod \frac{r^{-q_j}}{(r^{-2} \vee \xi_j)^{q_j/2}} \prod \xi_j^{(p_j-2)/2} d\xi \right). \tag{A.5}$$

A supplementary argument, given at the end of the proof, verifies that

$$H = O\left(\int_0^1 \cdots \int_0^1 \prod \frac{r^{-q_j}}{(r^{-2} \vee \xi_j)^{q_j/2}} \xi_j^{(p_j-2)/2} d\xi_j \right)$$

$$= O\left(\prod \left(\int_0^1 \frac{r^{-q_j}}{(r^{-2} \vee \xi_j)^{q_j/2}} \xi_j^{(p_j-2)/2} d\xi \right) \right). \tag{A.6}$$

Now,

$$\int_0^1 \frac{r^{-q_j}}{(r^{-2} \vee \xi_j)^{q_j/2}} \xi_j^{(p_j-2)/2} d\xi \le \int_0^{r^{-2}} \xi_j^{(p_j-2)/2} d\xi_j + \int_{r^{-2}}^1 r^{-q_j} \xi^{(p_j-q_j-2)/2} d\xi$$

$$= \begin{cases} O(r^{-q_j}) & \text{if } q_j < p_j, \\ O(r^{-p_j} \log r) & \text{if } q_j = p_j, \\ O(r^{-p_j}) & \text{if } q_j > p_j. \end{cases} \tag{A.7}$$

The desired conclusion follows from (A.6) and (A.7).

It remains to verify the first equality in (A.6). For this purpose, let $T_1, \ldots, T_m$ be independent random variables with respective densities

$$f_j(t_j) = t_j^{(p_j-2)/2} I_{(0,1)}(t_j). \tag{A.8}$$

Let

$$g_j(t_1, \ldots, t_m) = \frac{1}{(\sum t_k)^{q_j/2} (r^{-2} + t_j / \sum t_k)^{q_j/2}}. \tag{A.9}$$

Note that

$$\int \prod \frac{1}{r^{-q_j} + \xi_j^{q_j/2}} D(d\xi_1 \cdots d\xi_m) = E\left(g_j(T_1, \ldots T_m) \Big| \sum T_k = 1 \right). \tag{A.10}$$

Also note that

$$E\left(g_j(T_1, \ldots T_m) \Big| \sum T_k = s \right) = s^{q_j/2} E\left(g_j(T_1, \ldots T_m) \Big| \sum T_k = 1 \right). \tag{A.11}$$

It follows that

$$E(g_j(T_1, \ldots T_m)) = E\left(\left(\sum T_k \right)^{q_j/2} E\left(g_j(T_1, \ldots, T_m) \Big| \sum T_k = 1 \right) \right)$$

$$= E\left(\left(\sum T_k\right)^{q_j/2}\right) E\left(g_j(T_1,\ldots,T_m)\Big| \sum T_k = 1\right). \text{(A.12)}$$

(A.10) and (A.12) combine to show that the first equality of (A.6) is valid, as claimed.

Remark A.1. It can be shown that when $p_j = q_j$, $j = 1,\ldots,m$, then the value of Λ in (A.4) can be taken as $m - 1$ instead of m. This minor improvement plays no role in the earlier results of our paper, and hence we omit the formal statement and proof.

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Second edition. Springer, New York.

Berger, J. O. and Robert, C. P. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.* **18**, 617-651.

Berger, J. O. and Strawderman, W. E. (1996). Choice of hierarchical priors: admissibility in estimation of normal means. *Ann. Statist.* **24**, 931-951.

Berger, J. O., Strawderman, W. E. and Tang, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Ann. Statist.* **33**, 606-646.

Bock, M. E. (1988). Shrinkage estimators: Pseudo-Bayes rules for normal mean vectors. *Statistical Decision Theory and Related Topics IV* (Edited by S. S. Gupta and J. Berger) **1**, 281-297. Springer, New York.

Brown, L. D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Ann. Math. Statist.* **42**, 855-903.

Brown, L. D. (1988). The differential inequality of a statistical estimation problem. *Proc. 4th Purdue Symp: Statist. Dec. Theory and Related Topics* **1**, 299-324. Springer, New York.

Brown, L. D. and Zhao, L. H. (2006). Shrinkage estimation, an expanding statistical theme. (Presented as a Medallion Lecture at ENAR/IMS meeting, Spring 2006, available at `www.stat.wharton.upenn.edu/~lbrown`).

Brown, L. D., Zhao, L. H. and Mao, W. (2004). A pseudo-Bayesian procedure for nonparametric regression estimation and pointwise confidence bands. Available at `www.stat.wharton.upenn.edu/~lbrown`.

Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 898-924.

Cai, T., Low, M. and Zhao, L. H. (2000). Sharp adaptive estimation by a blockwise method. Available at `www.stat.wharton.upenn.edu/~lzhao`.

Cavalier, L. and Tsybakov, A. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods Statist.* **10**, 247-282.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 1200-1224.

Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika* **59**, 335-347.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68**, 117-130.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. 4th Berk. Symp. Math. Statist. and Prob.* **1**, 361-378, Univ. of Calif. Press, Berkeley, CA.

Konno, Y. (1991). On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.* **36**, 44-55.

Kubokawa, T. (1991). An approach to improving the James-Stein estimator. *J. Multivariate Anal.* **36**, 121-126.

Lin, P-E. and Tsai, H-L. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Ann. Statist.* **1**, 142-145.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation, Second Edition.* Springer, New York.

Lindley, D. V. (1962). Confidence sets for the mean of a multivariate normal distribution (discussion). *J. Roy. Statist. Soc. Ser. B* **24**, 285-287.

Maruyama, Y. (1999). Improving on the James-Stein estimator. *Statist. Decisions* **14**, 137-140.

Maruyama, Y. (2004). Stein's idea and minimax admissible estimation of a multivariate normal mean. *J. Multivariate Anal.* **88**, 320-334.

Maruyama, Y. and Strawderman, W. E. (2005). A new class of generalized Bayes minimax ridge regression estimators. *Ann. Statist.* **33**, 1753-1770.

Rukhin, A. L. (1995). Admissibility: Survey of a concept in progress. *Internat. Statist. Rev.* **63**, 95-115.

Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution (with discussion). *J. Roy. Statist. Soc. Ser. B* **24**, 265-296.

Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. on Asymptotic Statist.* 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 estimators of the multivariate normal mean. *Ann. Math. Statist.* **45**, 382-388.

Strawderman, W. E. (1973). Proper Bayes estimators of the multivariate normal mean for the case of common unknown variances. *Ann. Statist.* **1**, 1189-1194.

Tsukuma, H. and Kubokawa, T. (2007). Methods for improvement in estimation of a normal mean matrix. *J. Multivariate Anal.* **98**, 1592-1610.

Statistics Department, University of Pennsylvania 3730 Walnut Street Philadelphia, PA 19104, U.S.A.

E-mail: lbrown@wharton.upenn.edu

Statistics Department, University of Pennsylvania 3730 Walnut Street Philadelphia, PA 19104, U.S.A.

E-mail: lzhao@wharton.upenn.edu