



2011

Optimal Estimation of Multidimensional Normal Means With an Unknown Variance

Lawrence Brown
University of Pennsylvania

Xu Han
Princeton University

Follow this and additional works at: https://repository.upenn.edu/statistics_papers

 Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Brown, L., & Han, X. (2011). Optimal Estimation of Multidimensional Normal Means With an Unknown Variance. *The Annals of Statistics*, 1-31. Retrieved from https://repository.upenn.edu/statistics_papers/43

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/statistics_papers/43
For more information, please contact repository@pobox.upenn.edu.

Optimal Estimation of Multidimensional Normal Means With an Unknown Variance

Abstract

Let $X \sim N_p(\boldsymbol{\theta}, \sigma^2 I_p)$ and $W \sim \sigma^2 \chi_m^2$, where both $\boldsymbol{\theta}$ and σ^2 are unknown, and X is independent of W .

Optimal estimation of $\boldsymbol{\theta}$ with unknown σ^2 is a fundamental issue in applications but basic theoretical issues remain open. We consider estimation of $\boldsymbol{\theta}$ under squared error loss. We develop sufficient conditions for prior density functions such that the corresponding generalized Bayes estimators for $\boldsymbol{\theta}$ are admissible. This paper has a two-fold purpose: 1. Provide a benchmark for the evaluation of shrinkage estimation for a multivariate normal mean with an unknown variance; 2. Use admissibility as a criterion to select priors for hierarchical Bayes models. To illustrate how to select hierarchical priors, we apply these sufficient conditions to a widely used hierarchical Bayes model proposed by Maruyama & Strawderman [M-S] (2005), and obtain a class of admissible and minimax generalized Bayes estimators for the normal mean $\boldsymbol{\theta}$. We also develop necessary conditions for admissibility of generalized Bayes estimators in the M-S (2005) hierarchical Bayes model. All the results in this paper can be directly applied in the familiar setting of Gaussian linear regression.

Keywords

normal mean problem, admissibility, generalized Bayes estimator, unknown variance, shrinkage estimator, minimaxity

Disciplines

Physical Sciences and Mathematics

OPTIMAL ESTIMATION OF MULTIDIMENSIONAL NORMAL MEANS WITH AN UNKNOWN VARIANCE

BY LAWRENCE D. BROWN* AND XU HAN

University of Pennsylvania and Princeton University

Let $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \sigma^2 \mathbf{I}_p)$ and $W \sim \sigma^2 \chi_m^2$, where both $\boldsymbol{\theta}$ and σ^2 are unknown, and \mathbf{X} is independent of W . Optimal estimation of $\boldsymbol{\theta}$ with unknown σ^2 is a fundamental issue in applications but basic theoretical issues remain open. We consider estimation of $\boldsymbol{\theta}$ under squared error loss. We develop sufficient conditions for prior density functions such that the corresponding generalized Bayes estimators for $\boldsymbol{\theta}$ are admissible. This paper has a two-fold purpose: 1. Provide a benchmark for the evaluation of shrinkage estimation for a multivariate normal mean with an unknown variance; 2. Use admissibility as a criterion to select priors for hierarchical Bayes models. To illustrate how to select hierarchical priors, we apply these sufficient conditions to a widely used hierarchical Bayes model proposed by Maruyama & Strawderman [M-S] (2005), and obtain a class of admissible and minimax generalized Bayes estimators for the normal mean $\boldsymbol{\theta}$. We also develop necessary conditions for admissibility of generalized Bayes estimators in the M-S (2005) hierarchical Bayes model. All the results in this paper can be directly applied in the familiar setting of Gaussian linear regression.

1. Introduction. Estimation of the mean for a multivariate normal distribution has received enormous attention, and is also of substantial importance in contemporary statistical theory and application. Suppose $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \sigma^2 \mathbf{I}_p)$, and we use squared error loss. Stein discovered that when $p \geq 3$, the usual maximum likelihood estimator, \mathbf{x} , is not the best estimator of the unknown normal mean $\boldsymbol{\theta}$, when σ^2 is known and unknown but estimable. James and Stein (1961) provided an explicit class of dominating estimators. They provided results for both the cases of known σ^2 and of unknown σ^2 . A variety of modifications and generalizations of it are often referred to as statistical shrinkage estimators because of the way they act to improve on the usual estimators.

Our focus is on the case of unknown but estimable σ^2 . Various authors have proposed hierarchical Bayes priors for this setting. See especially Straw-

*Supported in part by NSF Grant DMS-07-07033

AMS 2000 subject classifications: 62C15, 62C10

Keywords and phrases: normal mean problem, admissibility, generalized Bayes estimator, unknown variance, shrinkage estimator, minimaxity

derman (1973), Maruyama & Strawderman (2005) and Wells & Zhou (2008). Some of these priors are “proper” (i.e. have mass one). But a broad class are generalized priors (i.e. have infinite mass but finite formal posterior distributions as discussed in Berger (1985)). The proper priors of course lead to admissible estimators, but only some of generalized priors do so. As discussed below this admissibility issue has been previously investigated for the case of known variance, σ^2 . Our focus is on the case of unknown variance. Maruyama & Strawderman (2005) and Wells & Zhou (2008) have investigated minimaxity properties of both (proper or generalized) Bayes estimators in this setting. Our focus is on the admissibility of these (generalized) Bayes estimators. We develop a general result in Theorem 1 and then apply it to the Maruyama & Strawderman (2005) estimators in section 3.2.

1.1. *Background.* Assume that the variance scalar σ^2 is known. For admissibility, Brown (1971) characterized all admissible estimators as being generalized Bayes (but not conversely). By Brown’s theorem, the James-Stein estimator is not admissible because it is not of analytic form, and thus it could not be generalized Bayes. Brown (1971) also gives sufficient and nearly necessary conditions for admissibility of generalized Bayes estimators. These results aid in the construction of admissible estimators. Strawderman (1971) used a hierarchical Bayes model to obtain a class of proper Bayes minimax estimators (which are of course admissible). Based on applications of Brown’s (1971) conditions, Berger & Strawderman (1996), and Berger, Strawderman & Tang (2005) examined the admissibility and inadmissibility of several commonly used generalized Bayes estimators.

For minimaxity, the usual maximum likelihood estimator \mathbf{x} is minimax, and any improvement is also minimax. Therefore James-Stein estimators are minimax. A substantial number of papers have applied Stein’s unbiased estimate of risk, and have obtained a wide class of minimax Bayes estimators. It is impossible for us to list all the references here, but interested readers are referred to the celebrated paper by Stein (1981) and applications shown by Efron & Morris (1971, 1972, 1975). Additional references can be found in Berger (1985) and Lehmann & Casella (1998).

However, most of the time, in practical problems the variance scalar σ^2 is unknown. To deal with this important case one approach is to use an estimate of σ^2 and then plug it in to treat the problem as if it were a known variance case. A well known example is the James-Stein positive part plug in estimator provided by James & Stein (1961). Another approach is to use a hierarchical Bayes model to put an objective prior on σ^2 , and then consider the corresponding Bayes estimators.

For admissibility, to the best of our knowledge, very few results have been obtained because of the technical difficulty. Strawderman (1973) used a hierarchical Bayes model to construct a class of proper Bayes estimators for the unknown variance case. But admissibility of generalized Bayes estimators is still an open and important problem.

1.2. *Outline and Contributions.* In the current paper, we address the problem of optimal estimation for the multivariate normal mean with an unknown variance. Our theoretical results and technical contributions can be summarized as the following three parts:

- We develop sufficient conditions for prior density functions such that the corresponding generalized Bayes estimators are admissible. Compared with the existing results for the known variance case (e.g., Brown & Hwang 1982), the technical difficulties in the unknown variance case come from (a) the improper prior on the variance and (b) the dependence of the normal mean parameter on the variance in the hierarchical Bayes model. The fundamental tool to prove admissibility is Blyth's method. In the current paper, we first provide a new sequence of priors for application of Blyth's method. These enable us to develop new uniform upper bound for the sequence of differences between the two Bayes risks that occur in Blyth's method.
- We apply our sufficient conditions to a widely used hierarchical Bayes model studied in M-S (2005), and obtain a class of admissible and minimax generalized Bayes estimators. Application of general admissibility theorems to a particular hierarchical Bayes model can be very technical even in the known variance case, see Berger & Strawderman (1996) and Berger, Strawderman & Tang (2005). In the present case appropriate and careful calculation is needed to establish specific admissibility conditions for the Maruyama & Strawderman (2005) priors. In the current paper, we develop a technique based on the order of magnitude for marginal densities and derivative of marginal densities. The technique itself is of independent interest, and has applications in hierarchical Bayes modeling.
- We develop necessary conditions for admissibility of generalized Bayes estimators in the M-S (2005) hierarchical Bayes model. The results here are used to investigate the sharpness of sufficient conditions developed in the current paper. The proof is based on a general theorem in Brown (1980), but technical issues of applying Brown (1980) to the unknown variance case are new. Our results can be considered as an important step to establish more general and sharper

necessary conditions in the unknown variance case.

Admissibility together with minimaxity provides a benchmark for the evaluation of shrinkage estimators. For example, the James-Stein positive part plug in estimator, various empirical Bayes estimators and generalized Bayes estimators have been shown to lead to satisfactory results in statistical data analysis. A wide range of applications can be found in Blattberg & George (1991), DuMouchel & Harris (1983), Fay & Herriot (1979), Gelfand, Hills, Racine-Poon & Smith (1990), Hui & Berger (1983), Jorion (1986), Zellner & Hong (1989), etc. However, there has not been any theoretical work to evaluate whether these shrinkage estimators are admissible or close to being admissible, and these estimators can be sub-optimal. Several of these applications involve situations that generalize the formulation in our paper, but our work in admissibility is an important step in the study of these more general situations.

Several authors have described admissibility as a powerful tool to select satisfactory hierarchical generalized Bayesian priors. In particular, Berger & Strawderman (1996), and Berger, Strawderman & Tang (2005) considered the estimation of the normal mean when the variance is known. They discussed several hierarchical Bayesian models for the normal mean parameter θ and used admissibility as a criterion to select hierarchical priors. Berger, Strawderman & Tang (2005) pointed out, “Use of objective improper priors in hierarchical modeling is of enormous practical importance, yet little is known about which such priors are good or bad. The most successful approach to evaluation of objective improper priors has been to study the frequentist properties of the ensuing Bayes procedures. In particular, it is important that the prior distribution not be too diffuse, and study of *admissibility* is the most powerful tool known for detecting an over-diffuse prior.” Also Berger (1985) has pointed out, “Prior distributions which are *on the boundary of admissibility* are particularly attractive noninformative priors.” Berger & Bernardo (1992), “Comparison and choice of noninformative priors must involve some type of frequentist computation and that consideration of *admissibility* of resulting estimators is an often enlightening approach.” [Our italics.]

The rest of our paper is organized as follows: section 2 includes concepts, notation and formulation of the problem, with relation to linear regression models; section 3 provides theorems giving sufficient and necessary conditions for admissibility of generalized Bayes estimators, with related lemmas; section 4 provides numerical analysis and investigation of the performance of particular generalized Bayes estimators, including those from M-S (2005); section 5 provides all the technical proofs.

2. Definition and Notation. Let $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \sigma^2 \mathbf{I}_p)$ be a p -dimensional multivariate normal random vector, with unknown mean vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ and unknown variance scalar σ^2 . Here \mathbf{I}_p is the $p \times p$ identity matrix. Let $W \sim \sigma^2 \chi_m^2$, where W and \mathbf{X} are independent. For convenience, from now on, we will reparameterize σ^2 in terms of the precision $\eta^2 = 1/\sigma^2$. Correspondingly we have

$$(1) \quad \mathbf{X} \sim N_p(\boldsymbol{\theta}, \eta^{-2} \mathbf{I}_p) \quad W \sim \eta^{-2} \chi_m^2.$$

With observations $\mathbf{x} = (x_1, \dots, x_p)^T$ and w , we use a procedure $\boldsymbol{\delta}(\mathbf{x}, w) = (\delta_1(\mathbf{x}, w), \dots, \delta_p(\mathbf{x}, w))^T : \mathbb{R}^p \times \mathbb{R} \rightarrow \mathbb{R}^p$ to estimate the unknown normal mean $\boldsymbol{\theta}$.

Consider the normalized squared error loss function

$$(2) \quad L(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}(\mathbf{x}, w)) = \|\boldsymbol{\delta}(\mathbf{x}, w) - \boldsymbol{\theta}\|^2 \eta^2 = \sum_{i=1}^p (\delta_i(\mathbf{x}, w) - \theta_i)^2 \eta^2.$$

Then the corresponding risk function will be

$$(3) \quad R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}(\mathbf{x}, w)) = E_{\boldsymbol{\theta}, \eta^2} L(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}(\mathbf{x}, w)).$$

DEFINITION 1. *An estimator $\boldsymbol{\delta}_1$ is inadmissible if there exists another estimator $\boldsymbol{\delta}_2$ such that $R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_2) \leq R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_1)$ for all $\boldsymbol{\theta}, \eta^2$ and $R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_2) < R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_1)$ for some $\boldsymbol{\theta}, \eta^2$. An estimator is admissible if it is not inadmissible.*

Our goal is to construct sufficient conditions for admissibility of estimators of the normal mean $\boldsymbol{\theta}$.

Focus first on the ordinary Bayes estimators. So assume for now that G is a probability distribution on $\mathbb{R}^p \times \mathbb{R}$. By definition a Bayes estimator minimizes the posterior risk. Thus the Bayes estimator, $\boldsymbol{\delta}_G$, solves the minimization problem:

$$(4) \quad \begin{aligned} B(G) &\equiv \int_0^\infty \int_{\mathbb{R}^p} R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_G) G(d\boldsymbol{\theta}, d\eta^2) \\ &\equiv \inf_{\boldsymbol{\delta}: \mathbb{R}^p \times \mathbb{R} \rightarrow \mathbb{R}^p} \int_0^\infty \int_{\mathbb{R}^p} R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}) G(d\boldsymbol{\theta}, d\eta^2), \end{aligned}$$

where the preceding also defines the usual symbol, $B(G)$, for the Bayes risk. And we have $\boldsymbol{\delta}_G = \operatorname{argmin}_{\mathbf{d} \in \mathbb{R}^p} E(L(\boldsymbol{\theta}, \eta^2, \mathbf{d}) | \mathbf{x}, w)$. Take the derivative of $E(L(\boldsymbol{\theta}, \eta^2, \mathbf{d}) | \mathbf{x}, w)$ with respect to \mathbf{d} and set it equal 0. This leads to the expression:

$$(5) \quad \boldsymbol{\delta}_G(\mathbf{x}, w) = \frac{\int_0^\infty \int_{\mathbb{R}^p} \boldsymbol{\theta} \eta^2 f(\mathbf{x} | \boldsymbol{\theta}, \eta^2) f(w | \eta^2) G(d\boldsymbol{\theta}, d\eta^2)}{\int_0^\infty \int_{\mathbb{R}^p} \eta^2 f(\mathbf{x} | \boldsymbol{\theta}, \eta^2) f(w | \eta^2) G(d\boldsymbol{\theta}, d\eta^2)},$$

where

$$(6) \quad f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) \propto (\eta^2)^{p/2} \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\theta}\|^2}{2} \eta^2\right)$$

$$(7) \quad f(w|\eta^2) \propto w^{(m-2)/2} (\eta^2)^{m/2} \exp\left(-\frac{w\eta^2}{2}\right).$$

If G is a finite (non-negative) measure, then $\int_0^\infty \int_{\mathbb{R}^p} G(d\boldsymbol{\theta}, d\eta^2) < \infty$, and $\frac{G(d\boldsymbol{\theta}, d\eta^2)}{\int_0^\infty \int_{\mathbb{R}^p} G(d\boldsymbol{\theta}, d\eta^2)}$ is a probability measure. The definition for Bayes risk in (4) and Bayes estimator in (5) of probability measure can be extended to a finite measure. The definition in (5) can be further extended to hold for general (non-negative) measures G so long as the integrals in the numerator and denominator exist for all x, w . In such a case $\boldsymbol{\delta}_G$ is called a generalized Bayes estimator.

We pay special attention to the following hierarchical Bayes model:

$$(8) \quad \begin{aligned} \mathbf{X} &\sim N_p(\boldsymbol{\theta}, \eta^{-2} \mathbf{I}_p) & W &\sim \eta^{-2} \chi_m^2 \\ \boldsymbol{\theta}|\eta^2 &\sim g(\boldsymbol{\theta}; \eta^2) & \eta^2 &\sim \pi(\eta^2). \end{aligned}$$

The prior density for normal mean $\boldsymbol{\theta}$ depends on η^2 , which results in a two-level Bayes model. The prior density functions $g(\boldsymbol{\theta}; \eta^2)$ and $\pi(\eta^2)$ could be improper (i.e., could correspond to infinite measures). The model of M-S (2005) described in (20) is a special case.

The corresponding generalized Bayes estimator for the normal mean $\boldsymbol{\theta}$ is

$$(9) \quad \boldsymbol{\delta}_G(\mathbf{x}, w) = \frac{\int_0^\infty \int_{\mathbb{R}^p} \boldsymbol{\theta} \eta^2 f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2) d\boldsymbol{\theta} d\eta^2}{\int_0^\infty \int_{\mathbb{R}^p} \eta^2 f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2) d\boldsymbol{\theta} d\eta^2}.$$

Note that η^2 is considered as the variable of integration. We assume the existence of the integrals in the numerator and denominator of (9). The goal is to construct sufficient conditions for the prior density functions $g(\boldsymbol{\theta}; \eta^2)$ and $\pi(\eta^2)$ such that the generalized Bayes estimator $\boldsymbol{\delta}_G(\mathbf{x}, w)$ is admissible for estimating the unknown normal mean $\boldsymbol{\theta}$ with respect to the squared error loss function (2).

The following remarks sketch some familiar features of our formulation in order to emphasize its range of applicability. The following remarks taken from earlier literature show that all the results in this paper related to set up (1)-(3) can be directly applied in the conventional linear regression setting.

Remark 1 (Relation to General Covariance Matrix Case): Consider a generalization in which $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \eta^{-2} \boldsymbol{\Sigma})$ with loss function $(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1} \eta^2 (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$, where $\boldsymbol{\Sigma}$ is a known nonsingular symmetric $p \times p$ matrix.

For the present remark we refer to this formulation as “Problem 1”. There exists a matrix \mathbf{Q} such that $\mathbf{Q}^T \Sigma \mathbf{Q} = \mathbf{I}_p$. Thus $\mathbf{Q}^T \mathbf{X} \sim N_p(\mathbf{Q}^T \boldsymbol{\theta}, \eta^{-2} \mathbf{I}_p)$ with loss function $(\boldsymbol{\delta}(\mathbf{Q}^T \mathbf{X}) - \mathbf{Q}^T \boldsymbol{\theta})' \eta^2 (\boldsymbol{\delta}(\mathbf{Q}^T \mathbf{X}) - \mathbf{Q}^T \boldsymbol{\theta})$, is equivalent to our original setting (1)-(3). We refer to this formulation as “Problem 2”. By Definition 1, if $\boldsymbol{\delta}(\mathbf{Q}^T \mathbf{X})$ is admissible for Problem 2, then $(\mathbf{Q}^T)^{-1} \boldsymbol{\delta}(\mathbf{Q}^T \mathbf{X})$ is admissible for Problem 1. Thus our results of admissibility for the original setting (1)-(3) can be directly applied to Problem 1.

Remark 2 (Relation to Linear Regression Model): Consider the linear regression setting, $\mathbf{y} = \mathbf{M}\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$, where \mathbf{y} is a n -dimensional observed vector, \mathbf{M} is a $n \times p$ known design matrix, $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ is random error. Assume \mathbf{M} has full rank p . Both the p -dimensional coefficient vector $\boldsymbol{\beta}$ and the variance σ^2 are unknown. We want to estimate $\boldsymbol{\beta}$. The conventional least squares estimator for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{y}$. Thus $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2 (\mathbf{M}^T \mathbf{M})^{-1})$ with the loss function $(\boldsymbol{\delta} - \boldsymbol{\beta})' \frac{(\mathbf{M}^T \mathbf{M})}{\sigma^2} (\boldsymbol{\delta} - \boldsymbol{\beta})$ is equivalent to the setting in Remark 1 if we let $\Sigma = (\mathbf{M}^T \mathbf{M})^{-1}$. Let W be the sum of squares for residuals. Therefore our results for the original setting (1)-(3) can be directly applied to estimating the coefficient vector $\boldsymbol{\beta}$ in the linear regression model.

3. Main Results.

3.1. *Admissibility Theorem.* In the setting of (8), let

$$(10) \quad f(\mathbf{x}, w, \boldsymbol{\theta}, \eta^2) = f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2)$$

$$(11) \quad g^*(\mathbf{x}, w, \eta^2) = \int_{\mathbb{R}^p} f(\mathbf{x}, w, \boldsymbol{\theta}, \eta^2) d\boldsymbol{\theta},$$

where $f(\mathbf{x}, w, \boldsymbol{\theta}, \eta^2)$ is the joint density function of $\mathbf{x}, w, \boldsymbol{\theta}, \eta^2$, and $g^*(\mathbf{x}, w, \eta^2)$ is the marginal joint density function of \mathbf{x}, w, η^2 .

The generalized Bayes estimator in (9) can be expressed as

LEMMA 1.

$$(12) \quad \boldsymbol{\delta}_G(\mathbf{x}, w) = \mathbf{x} + \frac{\int_0^\infty \nabla_{\mathbf{x}} g^*(\mathbf{x}, w, \eta^2) d\eta^2}{\int_0^\infty \eta^2 g^*(\mathbf{x}, w, \eta^2) d\eta^2}.$$

To find sufficient conditions for admissible generalized Bayes estimators, the following Condition 1 & 2 are fundamental. These generalize conditions in Brown & Hwang (1982).

CONDITION 1.

$$(13) \quad \int_{S^c} \int_0^\infty \frac{1}{\eta^2} \frac{g(\boldsymbol{\theta}; \eta^2)}{\|\boldsymbol{\theta}\|^2 \log^2(\|\boldsymbol{\theta}\| \vee 2)} \pi(\eta^2) d\eta^2 d\boldsymbol{\theta} < \infty.$$

where S denote the ball of radius 1 at the origin in \mathbb{R}^p , S^c is the complement of S and $a \vee b$ is defined as $\max(a, b)$. This is called the Growth Condition on the prior density functions.

The Growth Condition suggests that $g(\boldsymbol{\theta}; \eta^2)$ should not have heavy tails. In particular, this condition will be satisfied if

$$(14) \quad \int_0^\infty g(\boldsymbol{\theta}; \eta^2) \frac{\pi(\eta^2)}{\eta^2} d\eta^2 = O\left(\frac{1}{\|\boldsymbol{\theta}\|^{p-2}}\right).$$

Conversely the condition will not be satisfied if for some $r < p - 2$,

$$(15) \quad \int_0^\infty g(\boldsymbol{\theta}; \eta^2) \frac{\pi(\eta^2)}{\eta^2} d\eta^2 \asymp \frac{1}{\|\boldsymbol{\theta}\|^r}$$

as $\|\boldsymbol{\theta}\| \rightarrow \infty$.

Some commonly used prior density functions do not satisfy Condition 1. In Gelman, Carlin, Stern & Rubin (2003), on page 74, the authors suggested a noninformative prior distribution. Assume the prior on $(\boldsymbol{\theta}, \log \eta)$ is uniform or, equivalently,

$$p(\boldsymbol{\theta}, \eta^2) \propto \frac{1}{\eta^2}.$$

It is easy to check that this prior does not satisfy the Growth Condition.

To simplify our computation, we define $J^*(h_1, h_2)$ for generic functions h_1 and h_2 with domains $\mathbb{R}^p \times \mathbb{R}^+$ and \mathbb{R}^+ :

$$(16) \quad J^*(h_1, h_2) = \int_{\mathbb{R}^p} f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) h_1(\boldsymbol{\theta}; \eta^2) h_2(\eta^2) d\boldsymbol{\theta}.$$

If $h_1 = (h_1^1, \dots, h_1^p)^T$ is a vector, (16) is interpreted as a vector whose i th coordinate is $\int_{\mathbb{R}^p} f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) h_1^i(\boldsymbol{\theta}; \eta^2) h_2(\eta^2) d\boldsymbol{\theta}$. Note that for the prior density functions $g(\boldsymbol{\theta}; \eta^2)$ and $\pi(\eta^2)$, $J^*(g, \pi)$ is equivalent to $g^*(\mathbf{x}, w, \eta^2)$ as defined in (11), but in the notation for $J^*(g, \pi)$ we emphasize the effects of the prior density functions $g(\boldsymbol{\theta}; \eta^2)$ and $\pi(\eta^2)$.

CONDITION 2.

$$(17) \quad \int_{\mathbb{R}^p} \int_0^\infty \int_0^\infty \eta^2 J^*(g \parallel \frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{1}{\eta^2} \frac{\nabla_{\boldsymbol{\theta}} g}{g} \parallel^2, \pi) d\eta^2 dw d\mathbf{x} < \infty.$$

This is called the Asymptotic Flatness Condition on the prior density functions.

In application, Condition 2 may not be easy to verify. A more transparent but slightly less general condition is

CONDITION 3.

$$(18) \quad \int_{\mathbb{R}^p} \int_0^\infty \frac{1}{\eta^2} \frac{\|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}; \eta^2)\|^2}{g(\boldsymbol{\theta}; \eta^2)} \pi(\eta^2) d\eta^2 d\boldsymbol{\theta} < \infty.$$

Condition 3 is satisfied by most examples as shown in section 3.2.

LEMMA 2. *When Condition 3 holds, then Condition 2 also holds.*

CONDITION 4.

$$(19) \quad \int_{\|\boldsymbol{\theta}\| < B} \int_{\eta^2 < B} G(d\boldsymbol{\theta}, d\eta^2) < \infty, \quad \forall B < \infty.$$

Condition 4 is used to guarantee finiteness of the sequence of measures G_j described in (33) and (34) of section 5. It is separate from the main part of the technical proof in Theorem 1. Condition 4 is very mild. It will be shown later that Condition 4 is satisfied by most M-S (2005) priors and others.

THEOREM 1. *Consider the hierarchical Bayes model (8). If the prior density functions $g(\boldsymbol{\theta}; \eta^2)$, and $\pi(\eta^2)$ satisfy Condition 1, 2 and 4, then the corresponding generalized Bayes estimator (9) for the normal mean $\boldsymbol{\theta}$ is admissible.*

Remark 3: For the known variance case, Brown (1971) provides a complete class theorem: admissible estimators must be generalized Bayes. However, for the unknown variance case, we are not sure if such statement is still correct. Our current sufficient conditions are just for generalized Bayes estimators. We cannot rule out the possibility that there exist other admissible estimators which are not generalized Bayes.

3.2. *Application to a Hierarchical Prior Setting.* Maruyama and Strawderman (2005) considered the following hierarchical Bayes model to study minimaxity:

$$(20) \quad \begin{aligned} \mathbf{X} &\sim N_p(\boldsymbol{\theta}, \eta^{-2} \mathbf{I}_p) & W &\sim \eta^{-2} \chi_m^2 \\ \boldsymbol{\theta} | \nu, \eta^2 &\sim N_p(0, \nu \eta^{-2} \mathbf{I}_p) \\ h(\nu) d\nu &\propto \nu^b (1 + \nu)^{-a-b-2} d\nu & \pi(\eta^2) d\eta^2 &\propto (\eta^2)^{-k-1/2} d\eta^2. \end{aligned}$$

Wells & Zhou (2008) also considered a generalization in which $h(\nu)$ has a more general form and investigated minimaxity in this general setting. Like Maruyama & Strawderman (2005) they investigated (only) minimaxity, not admissibility. When we apply combination of Theorem 1 and Lemma 2 to

the hierarchical Bayes model above with the Maruyama and Strawderman (2005) prior, we obtain a class of admissible generalized Bayes estimators for the normal mean θ .

THEOREM 2. *In the hierarchical Bayes model (20), when $-a - 3/2 < k \leq 1/2$, $a > -2$ and $b > 0$, the corresponding generalized Bayes estimators (9) for the normal mean θ are admissible.*

Remark 4: Theorem 3 below makes clear that this condition is almost sharp. We conjecture that even when

$$(21) \quad -a - 3/2 < k \leq 1/2, \quad a > -2, \quad b = 0 \quad \text{or}$$

$$(22) \quad -a - 3/2 = k \leq 1/2, \quad a \geq -2, \quad b \geq 0,$$

the corresponding generalized Bayes estimator is admissible. Numerical results such as those shown in Figure 2 also reveal such estimators can have desirable risk functions.

Han (2009) contains a proof of this conjectured admissibility when $-a + 1/2 < k < 1/2$ and $b \geq 0$, but we do not yet have a complete proof for all of the boundary points in (21) and (22). In particular we do not have a proof for $k = 1/2$, $a = -2$, $b = 0$. A solution for this case would be of particular interest since this is the natural extension of Stein's harmonic prior (Stein (1973, 1981)) to our unknown variance problem. See also our Figure 1 in Section 4.1. The proof for the values $-a + 1/2 < k < 1/2$ and $b \geq 0$ discussed above involves a considerable extension of the argument in the current paper, and will appear elsewhere.

It should be noted that for any procedure satisfying (21) there is an admissible procedure with $b = \epsilon > 0$ satisfying the conditions in Theorem 2, and such that the risk functions of the two procedures are uniformly close to each other. In this sense we may at least say that the procedures satisfying (21) are nearly admissible even if we are so far unable to prove that some of the limiting estimators are actually admissible. Among these, the estimator for $a = -2$, $b = 0$ is of particular interest as noted above and in Section 4.

Remark 5: Maruyama & Strawderman (2005) and Wells & Zhou (2008) give the sufficient conditions for δ_G in model (20) to exist:

$$a > -p/2 - 1, \quad k < p/2 + m/2 + 3/2, \quad b > -1.$$

Theorems 1 and 2 implicitly include the conclusion that the posterior distributions exist under the stated conditions, along with the integrals in the numerator and denominator of (9).

M-S (2005) showed that when $b \geq 0$, $m/2 - k - 1/2 > a > -p/2 - 1$, and $0 \leq \frac{p/2+a+1}{m/2-k-1/2-a} \leq \frac{2p-4}{m+2}$, their prior leads to minimax generalized Bayes estimators. Combining with this result, we have Corollary 1.

COROLLARY 1. *If the following set of restrictions are satisfied:*
(23)

$$b > 0 \quad a > -2 \quad -a-3/2 < k \leq \min(1/2, \frac{2-2p-m}{2p-4}a + \frac{pm/2-2p-3m}{2p-4}),$$

then the corresponding generalized Bayes estimators (9) are admissible and minimax.

These conditions are discussed further in section 4.1 and illustrated by Figure 1.

Remark 6: The set of restrictions (23) also implies that $p \geq 3$ and $-2 < a < p/2 - 3$ should be satisfied. Therefore, Corollary 1 shows that there exist admissible and minimax generalized Bayes estimators for $p \geq 3$ in model (20) as long as a, b, k for the priors satisfy (23). We should point out the restriction for k does not involve b beyond the requirement that $b > 0$.

3.3. *Necessary Condition for Admissibility in a Hierarchical Prior Setting.* We provide a necessary condition for admissibility of generalized Bayes estimators in the hierarchical Bayes model (20), which will be compared with the results in Theorem 2 and illustrated in section 3.2. The purpose here is to investigate the sharpness of our sufficient conditions for admissibility in Theorem 2.

THEOREM 3. *In the hierarchical Bayes model (20), if δ_G is an admissible generalized Bayes estimator, then k, a, p and m should satisfy*

$$(24) \quad (a+2)m + (k+a+3/2)p + 1 - 2k \geq 0.$$

Remark 7: In Theorem 3, (24) $\Leftrightarrow a \geq -\frac{p-2}{m+p}k - \frac{2m+\frac{3}{2}p+1}{m+p}$. For finite p , when $m \rightarrow \infty$, $\frac{p}{m} \rightarrow 0$, then (24) is $a \geq -2$. This matches the sufficient condition $a > -2$ in Theorem 2 for a . When $m \rightarrow \infty$, it means that the variance of the normal distribution is known. This result is also consistent with the well known result for admissibility and inadmissibility in the known variance case. Another interesting case is that when $k = 1/2$, then no matter what values for p and m , (24) is $a \geq -2$. Also (24) $\Leftrightarrow k \geq -\frac{m+p}{p-2}a - \frac{2m+\frac{3}{2}p+1}{p-2}$. For finite m , when $p \rightarrow \infty$, $\frac{m}{p} \rightarrow 0$, then (24) is $k \geq -a - 3/2$. This matches the sufficient condition $k > -a - 3/2$ in Theorem 2 for k .

4. Numerical Analysis.

4.1. *Hierarchical Prior Selection.* Figure 1 shows the region of admissible and minimax generalized Bayes estimators for the M-S (2005) estimators. In this figure $p = 10, b > 0$, and the figure shows the relationship among k , a and m for admissibility and minimaxity.

The lowest sloped line [red] is for $k = -a - 3/2$. This is the lower boundary for admissibility but not inclusive. The remaining sloped lines [in various colors] are for $\frac{2-2p-m}{2p-4}a + \frac{pm/2-2p-3m}{2p-4}$ with $m = 16, 20, 25, 30$ respectively. These are upper boundaries for minimaxity. The line at $k = 1/2$ is an upper boundary for admissibility according to our Theorem 2. For example, when $m = 16$, the admissible and minimax region is the shadowed triangle surrounded by the red lines and the blue line. We use the dotted line to emphasize that the boundary is not currently included in the proven regions for admissibility and minimaxity.

There are two interesting and special points on the boundary of the admissible and minimax region in Figure 1. The upper boundaries of the minimaxity region and the lower boundary of the admissibility region all pass through $a = 2, k = -7/2$.

The other special point is for $a = -2, k = 1/2$. The case when $b = 0, a = -2$ and $k = 1/2$ corresponds to the special situation in which $g(\boldsymbol{\theta}) \propto 1/||\boldsymbol{\theta}||^{p-2}$, and $\pi(\eta^2) \propto 1/(\eta^2)^{k-1/2}$. The density for $\boldsymbol{\theta}$ is the well known harmonic prior. See Stein (1981). In this case $\boldsymbol{\theta}$ is independent of η^2 . This prior is discussed further at the end of Section 4.2.

The necessary condition for admissibility in Theorem 3 can best be understood by Figure 2. The shadowed region is for admissible estimators in Theorem 2. The dashed line [red] is for $k = -a - 3/2$. The horizontal line [red] is for $k = 1/2$. The remaining sloped lines [in various colors] are for $k = -\frac{m+p}{p-2}a - \frac{2m+\frac{3}{2}p+1}{p-2}$ with $(p, m) = (10, 250), (50, 150), (150, 40), (250, 10)$ respectively. These are upper boundaries for inadmissibility due to Theorem 3. The vertical line [black] is for $a = -2$. This is the right boundary for inadmissibility when p is finite and m is ∞ . The dashed line [red] is also the upper boundary for inadmissibility when m is finite and p is ∞ . When the ratio $\frac{m}{p}$ increases from 0 to ∞ , the boundary for inadmissibility moves clockwise from the dashed line $k = -a - 3/2$ to the vertical line $a = -2$.

4.2. *Comparison with James-Stein Positive Part Estimator.* We will compare the generalized Bayes estimators in (20) with the James-Stein positive part estimators. Applying the prior density functions in (20) to expression

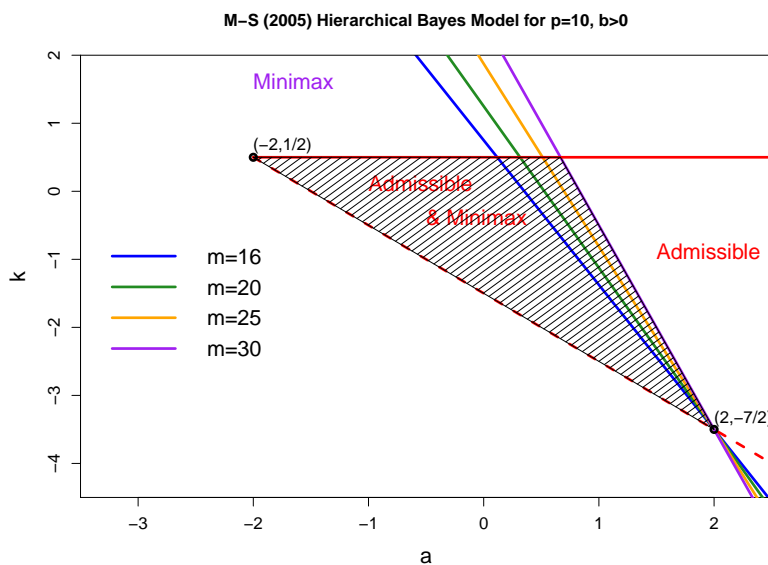


FIG 1. *Admissibility and Minimarity: The region of minimax estimators includes all those falling to the left and below the boundaries marked with values of m .*

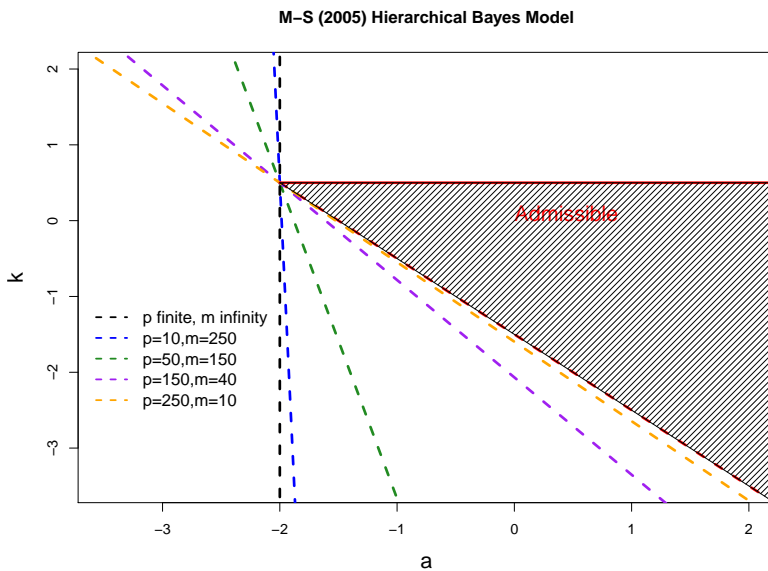


FIG 2. *Admissibility and Inadmissibility: The region of estimators proven to be inadmissible includes all those falling to the left and below the boundaries marked with values of p and m .*

(9) yields

$$\begin{aligned}
\delta_G(\mathbf{x}, w) &= \frac{\int_0^1 t^{a+p/2}(1-t)^{b+1}(w/2 + \|\mathbf{x}\|^2 t/2)^{-m/2-p/2+k-3/2} dt}{\int_0^1 t^{a+p/2}(1-t)^b(w/2 + \|\mathbf{x}\|^2 t/2)^{-m/2-p/2+k-3/2} dt} \mathbf{x} \\
&= \left(1 - \frac{\int_0^1 t^{a+p/2+1}(1-t)^b(1+t\|\mathbf{x}\|^2/w)^{-m/2-p/2+k-3/2} dt}{\int_0^1 t^{p/2+a}(1-t)^b(1+t\|\mathbf{x}\|^2/w)^{-m/2-p/2+k-3/2} dt}\right) \mathbf{x} \\
(25) \quad &= \left(1 - \frac{Et/(1+t\|\mathbf{x}\|^2/w)^{m/2+p/2-k+3/2}}{E1/(1+t\|\mathbf{x}\|^2/w)^{m/2+p/2-k+3/2}}\right) \mathbf{x}.
\end{aligned}$$

In the last step, $t \sim \text{Beta}(a + p/2 + 1, b + 1)$.

Traditionally, for estimation of the normal mean with an unknown variance, James-Stein positive part estimators are widely used. Let

$$(26) \quad \delta_{J-S}^+(\mathbf{x}, w) = \left(1 - \frac{cw}{m\|\mathbf{x}\|^2}\right)_+ \mathbf{x}.$$

Our goal is to find the relationship between c and p, m, a, b, k , and use admissible and minimax generalized Bayes estimator to mimic the James-Stein positive part estimator. Since the James-Stein positive part estimator is asymptotically optimal when $\|\boldsymbol{\theta}\|/\sigma$ is large, we search for corresponding admissible and minimax generalized Bayes estimator which has the same asymptotic property.

Let $z = \|\mathbf{x}\|^2/w$ and

$$(27) \quad \phi(z) = z \frac{\int_0^1 t^{p/2+a+1}(1-t)^b(1+zt)^{-p/2-m/2+k-3/2} dt}{\int_0^1 t^{p/2+a}(1-t)^b(1+zt)^{-p/2-m/2+k-3/2} dt},$$

then the generalized Bayes estimator can be written as

$$(28) \quad \delta_G(\mathbf{x}, w) = \left(1 - \frac{\phi(z)}{z}\right) \mathbf{x}.$$

It is easy to show that $\phi(z)$ is monotonically increasing in z , and

$$(29) \quad \lim_{z \rightarrow \infty} \phi(z) = (p/2 + a + 1)/(m/2 - k - 1/2 - a).$$

Also see Maruyama & Strawderman (2005).

Comparing with James-Stein positive part estimator, we have

$$(30) \quad \frac{c}{m} = \frac{p/2 + a + 1}{m/2 - k - 1/2 - a}.$$

Note that b is not involved in this relation.

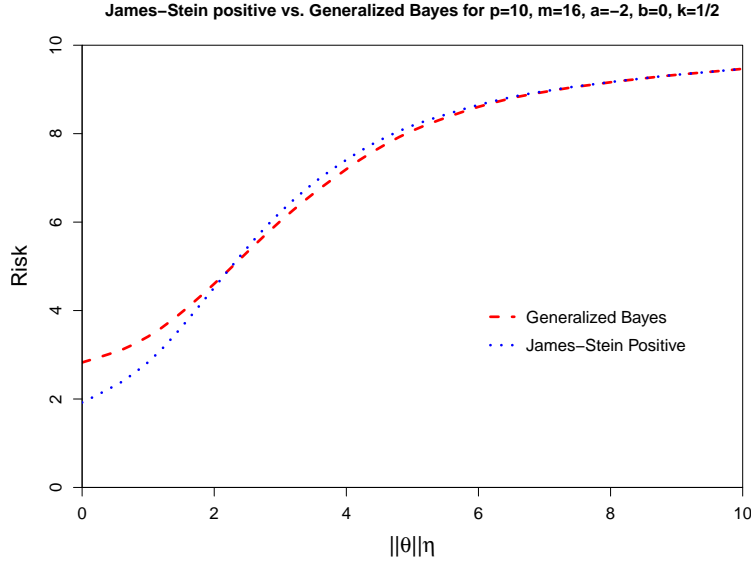


FIG 3. *James-Stein positive part estimator is very close to a minimax and nearly admissible generalized Bayes estimator. The risk plot is based on 30000 round Monte Carlo simulation.*

For the James-Stein positive part estimator, the most popular choice is $c = \frac{m}{m+2}(p - 2)$. The comparison of the generalized Bayes estimators with the James-Stein positive part estimator leads to $k = 1/2$ and $a = -2$. As already noted, this set of values is on the boundary of our sufficient conditions for admissibility. Also see our Figure 1. We consider the case when $p = 10, m = 16, b = 0, a = -2, k = 1/2$. The comparison of risk functions for this generalized Bayes estimator with the corresponding James-Stein positive part estimator is shown in Figure 3. As noted earlier, our Corollary 1 shows that there are admissible estimators whose risk function is arbitrarily close to that of this generalized Bayes estimator.

Remark 8: Formula (27) and (28) give an expression for the formal Bayes estimators under (20). It is evident from this that the estimator is scale invariant even though the prior is scale invariant only for $k = 1/2$. The prior is also rotation invariant. These facts imply that the risk function of the M-S estimators depends only on $\|\theta\|\eta = \|\theta\|/\sigma$. This fact is used in producing the plots in Figure 3, 4, 5 and 6.

The expression (27) also clarifies somewhat the role of k in model (20). Note that the functional form of the estimator depends only on $-m/2 + k =$

$-(m - 2k)/2$. Thus, a problem having m, k and another problem having $m', k' = k + \frac{m'-m}{2}$ will have functionally identical estimators.

4.3. Performance of Generalized Bayes Estimators. We first provide an example to emphasize that only minimaxity is not enough for evaluation of an estimator's performance. We will propose a sequence of minimax but inadmissible generalized Bayes estimators based on the M-S (2005) model in (20). To illustrate their results and demonstrate the relevance of admissibility we choose several of their priors for $p = 10, m = 16, a = -5, b = 0$. We pick $k = -10, -5, 0, 5, 10$ and make plots of the risk for the corresponding generalized Bayes procedures. According to results of M-S (2005) all of these estimators are minimax, since $k < 91/8$. We also include the risk of the James-Stein positive part estimator in the plot. The illustration is shown in Figure 4.

The estimator with smaller k value is dominated by the estimator with larger k value when $k = -10, -5, 0, 5, 10$. The best of these estimators is the one for $k = 10$. All the five estimators are minimax, but the estimators with $k = -10, -5, 0, 5$ are certainly not admissible and do not have as desirable risk functions compared with the two best choices. This suggests that these four estimators should not be used. The estimator with $k = 10$ is dominated by the James-Stein positive part estimator when $\|\boldsymbol{\theta}\|\eta$ is small, and slightly dominates this James-Stein estimator in a range of values. The plot demonstrates that minimaxity is not enough for evaluation of an estimator's performance, even when the estimator is generalized Bayes.

The following example more clearly demonstrates that the sufficient conditions in Theorem 2 are almost sharp. We choose $p = 10, m = 16, b = 1, k = 1/2$, but pick $a = 0, -2, -3$. $a = 0$ is in the range of admissibility; $a = -2$ is on the boundary for the range of admissibility in Theorem 2 but not inclusive, and on the boundary for the range of inadmissibility in Theorem 3 but not inclusive; $a = -3$ is outside of the range of admissibility and in the range of inadmissibility in Theorem 3. The comparison is shown in Figure 5. From Figure 5, it is clear that when $a = -3$ the generalized Bayes estimator is dominated by the estimator for $a = -2$ when $\|\boldsymbol{\theta}\|\eta \leq 10$. More detailed risk comparison for $a = -3$ vs. $a = -2$ is shown in Figure 6. On the range $\|\boldsymbol{\theta}\|\eta$ from 0 to 45, the risk function for $a = -2$ statistically significantly dominates the risk function for $a = -3$. For sufficiently large $\|\boldsymbol{\theta}\|\eta$, by the following Proposition 1, this risk dominance is still correct. This is consistent with the fact that $a = -3$ is outside of the admissibility range in Theorem 2 and falls with the scope of Theorem 3.

PROPOSITION 1. *When $\|\boldsymbol{\theta}\|\eta$ is sufficiently large and $k = 1/2$, the risk*

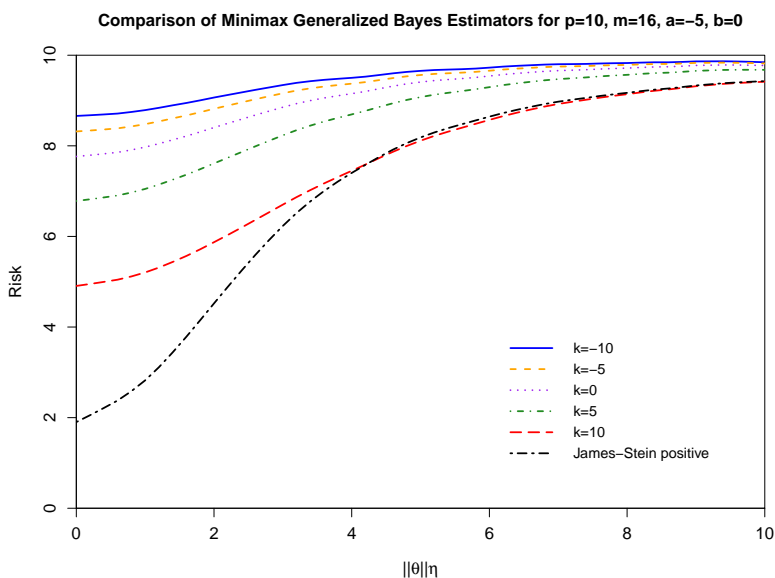


FIG 4. *Inadmissible Minimax Generalized Bayes Estimators: The estimator with smaller k value is dominated by the estimator with larger k value when $k = -10, -5, 0, 5, 10$. The risk plot is based on 30000 round Monte Carlo simulation.*

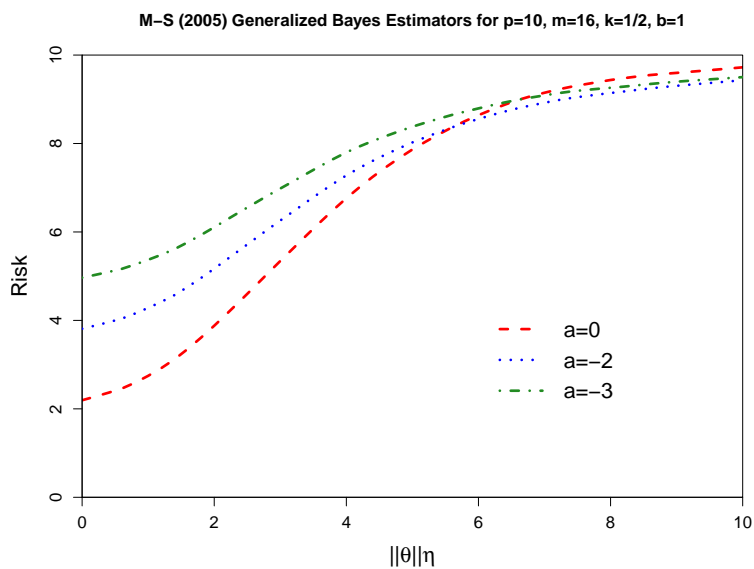


FIG 5. *$a = 0$ is in the range of admissibility in Theorem 2; $a = -2$ is on the boundary of admissibility in Theorem 2 but not inclusive and on the boundary of inadmissibility in Theorem 3 but not inclusive; $a = -3$ is in the range of inadmissibility in Theorem 3. The risk plot is based on 30000 round Monte Carlo simulation.*

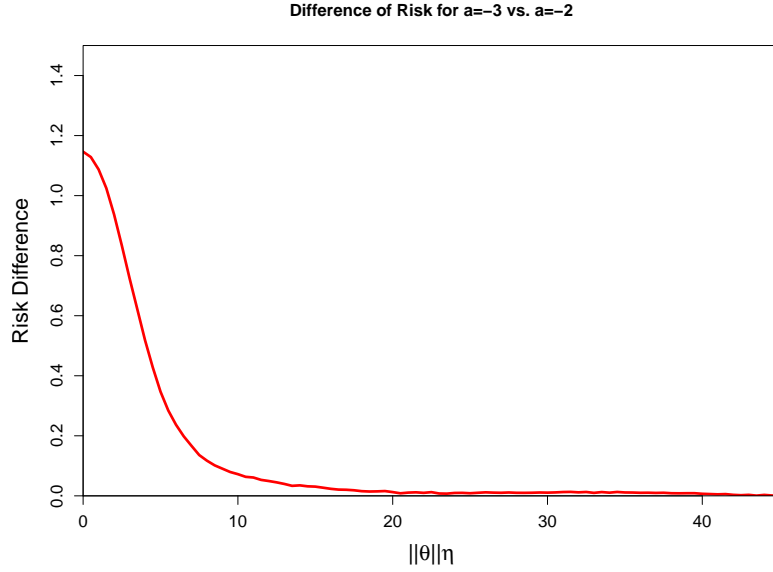


FIG 6. *Difference in risk for $a = -3$ vs. $a = -2$ when $p = 10$, $m = 16$, $k = 1/2$ and $b = 1$ on the range $\|\boldsymbol{\theta}\|\eta$ from 0 to 45 based on 30000 round Monte Carlo simulation.*

function for $a = -2$ dominates the case for $a = -3$.

The proof of Proposition 1 is similar to the proof of Theorem 3, thus it is omitted.

5. Proofs. In the section, we provide proofs for the results in Section 3.

Proof of Lemma 1 can be found in Brown (1971).

The proof in Theorem 1 is based on the method in Blyth (1951).

LEMMA 3. (*Blyth's method*) Let $\boldsymbol{\delta}$ be any estimator. Let K denote a non-empty compact subset of the parameter space. Suppose there is a sequence of finite measures G_j such that

$$(31) \quad \inf\{G_j(K) : j = 1, \dots\} > 0 \quad \int_0^\infty \int_{\mathbb{R}^p} R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}) G_j(d\boldsymbol{\theta}, d\eta^2) - B(G_j) \rightarrow 0,$$

then $\boldsymbol{\delta}$ is admissible.

Proof of Lemma 3 can be found in Stein (1955), Farrell (1964) and Brown (1971).

In Blyth's method, we need to calculate the difference between two integrals for the risk function with respect to $G_j(d\boldsymbol{\theta}, d\eta^2)$. Thus Lemma 4 will be useful for our proof.

LEMMA 4.

$$(32) \quad \int_0^\infty \int_{\mathbb{R}^p} R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}) G(d\boldsymbol{\theta}, d\eta^2) - B(G) \\ = \int_0^\infty \int_{\mathbb{R}^p} \|\boldsymbol{\delta}_G(x, w) - \boldsymbol{\delta}(\mathbf{x}, w)\|^2 \left(\int_0^\infty \eta^2 g^*(\mathbf{x}, w, \eta^2) d\eta^2 \right) d\mathbf{x} dw.$$

Proof of Lemma 4 can be found in James & Stein (1961).

Proof of Theorem 1: Start with the prior density functions. Let $g_j(\boldsymbol{\theta}; \eta^2) = k_j^2(\boldsymbol{\theta})g(\boldsymbol{\theta}; \eta^2)$, and $\pi_j(\eta^2) = l_j^2(\eta^2)\pi(\eta^2)$, where

$$(33) \quad k_j(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } \|\boldsymbol{\theta}\| < 1 \\ 1 - \frac{\log(\|\boldsymbol{\theta}\|)}{\log j} & \text{if } 1 \leq \|\boldsymbol{\theta}\| \leq j \\ 0 & \text{if } \|\boldsymbol{\theta}\| > j \end{cases}$$

$$(34) \quad l_j(\eta^2) = \begin{cases} 1 & \text{if } \eta^2 < 1 \\ 1 - \frac{\log(\eta^2)}{\log j} & \text{if } 1 \leq \eta^2 \leq j \\ 0 & \text{if } \eta^2 > j \end{cases}$$

$k_j(\boldsymbol{\theta})$ is as in Brown (1971) and in Brown & Hwang (1982) (and similar to a choice in James & Stein (1961)). But the product structure in π_j and form of l_j are additional and crucial choices. Deriving the best possible admissibility results for the vary conditions in Remark 4 may require a different form of prior sequence.

The main step is to apply Lemma 1, 4, and Cauchy-Schwartz inequality in an appropriate way. We show that

$$(35) \quad \Delta_j = \int \int [R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_G) - R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_{G_j})] g_j(\boldsymbol{\theta}; \eta^2) \pi_j(\eta^2) d\boldsymbol{\theta} d\eta^2$$

is uniformly upper bounded when $g(\boldsymbol{\theta}; \eta^2)$ and $\pi(\eta^2)$ satisfy Conditions 1 and 2. It is also true that the integrand in (35) converges pointwise to 0. Hence by the dominated convergence theorem, $\Delta_j \rightarrow 0$. Then the proof is complete.

By Lemma 1 and 4, we have

$$\begin{aligned}
\Delta_j &= \int_0^\infty \int_{\mathbb{R}^p} [R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_G) - R(\boldsymbol{\theta}, \eta^2, \boldsymbol{\delta}_{G_j})] g_j(\boldsymbol{\theta}; \eta^2) \pi_j(\eta^2) d\boldsymbol{\theta} d\eta^2 \\
&= \int_0^\infty \int_{\mathbb{R}^p} \|\boldsymbol{\delta}_{G_j}(\mathbf{x}, w) - \boldsymbol{\delta}_G(\mathbf{x}, w)\|^2 \left(\int_0^\infty \eta^2 g_j^*(\mathbf{x}, w, \eta^2) d\eta^2 \right) d\mathbf{x} dw \\
&= \int_0^\infty \int_{\mathbb{R}^p} \left\| \frac{\int \nabla_{\mathbf{x}} g_j^*(\mathbf{x}, w, \eta^2) d\eta^2}{\int_0^\infty \eta^2 g_j^*(\mathbf{x}, w, \eta^2) d\eta^2} - \frac{\int \nabla_{\mathbf{x}} g^*(\mathbf{x}, w, \eta^2) d\eta^2}{\int \eta^2 g^*(\mathbf{x}, w, \eta^2) d\eta^2} \right\|^2 \\
&\quad \times \left(\int \eta^2 g_j^*(\mathbf{x}, w, \eta^2) d\eta^2 \right) d\mathbf{x} dw.
\end{aligned}$$

Make the transformation $\mathbf{t} = \boldsymbol{\theta} - \mathbf{x}$ to obtain

$$\begin{aligned}
\nabla_{\mathbf{x}} g^*(\mathbf{x}, w, \eta^2) &= \nabla_{\mathbf{x}} \int_{\mathbb{R}^p} f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2) d\boldsymbol{\theta} \\
&= \nabla_{\mathbf{x}} \int_{\mathbb{R}^p} f(\mathbf{x}|\mathbf{t}, \eta^2) f(w|\eta^2) g(\mathbf{t} + \mathbf{x}; \eta^2) \pi(\eta^2) d\mathbf{t} \\
&= \int_{\mathbb{R}^p} f(\mathbf{t}|\mathbf{0}, \eta^2) f(w|\eta^2) \nabla_{\mathbf{x}} g(\mathbf{t} + \mathbf{x}; \eta^2) \pi(\eta^2) d\mathbf{t} \\
&= \int_{\mathbb{R}^p} f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) \nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2) d\boldsymbol{\theta}
\end{aligned}$$

So if we let $J^*(g, \pi) = \int_{\mathbb{R}^p} f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2) d\boldsymbol{\theta}$, then

$$\nabla_{\mathbf{x}} J^*(g, \pi) = J^*(\nabla_{\boldsymbol{\theta}} g, \pi).$$

Hence

$$\begin{aligned}
\Delta_j &= \int_0^\infty \int_{\mathbb{R}^p} \left\| \frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g_j, \pi_j) d\eta^2}{\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2} \right\|^2 \\
&\quad \times \left(\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2 \right) d\mathbf{x} dw \\
&= \int_0^\infty \int_{\mathbb{R}^p} \left\| \frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{\int_0^\infty J^*(k_j^2 \nabla_{\boldsymbol{\theta}} g, \pi_j) d\eta^2}{\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2} \right. \\
&\quad \left. - \frac{\int_0^\infty J^*(g \nabla_{\boldsymbol{\theta}} k_j^2, \pi_j) d\eta^2}{\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2} \right\|^2 \left(\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2 \right) d\mathbf{x} dw
\end{aligned}$$

We need to find an upper bound for Δ_j uniformly in j , such that $\Delta_j \rightarrow 0$ by the dominated convergence theorem.

To obtain a uniform upper bound for Δ_j , consider the following A_j and B_j :

$$\begin{aligned}
 A_j &\equiv \int_0^\infty \int_{\mathbb{R}^p} \left\| \frac{\int_0^\infty J^*(g \nabla_{\boldsymbol{\theta}} k_j^2, \pi_j) d\eta^2}{\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2} \right\|^2 \left(\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2 \right) d\mathbf{x} dw \\
 &= 4 \int_0^\infty \int_{\mathbb{R}^p} \left\| \frac{\int_0^\infty J^*(g k_j \nabla_{\boldsymbol{\theta}} k_j, \pi_j) d\eta^2}{\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2} \right\|^2 \left(\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2 \right) d\mathbf{x} dw \\
 &\leq 4 \int_0^\infty \int_{\mathbb{R}^p} \left(\int_0^\infty \frac{1}{\eta^2} J^*(g \|\nabla_{\boldsymbol{\theta}} k_j\|^2, \pi_j) d\eta^2 \right) d\mathbf{x} dw \\
 &= 4 \int_0^\infty \int_{\mathbb{R}^p} \frac{1}{\eta^2} g(\boldsymbol{\theta}; \eta^2) \|\nabla_{\boldsymbol{\theta}} k_j\|^2 \pi_j(\eta^2) \\
 &\quad \times \left(\int_0^\infty \int_{\mathbb{R}^p} f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(w|\eta^2) d\mathbf{x} dw \right) d\boldsymbol{\theta} d\eta^2 \\
 &= 4 \int_0^\infty \int_{\mathbb{R}^p} \frac{1}{\eta^2} g(\boldsymbol{\theta}; \eta^2) \|\nabla_{\boldsymbol{\theta}} k_j\|^2 \pi_j(\eta^2) d\boldsymbol{\theta} d\eta^2 \\
 &\leq 4 \int_0^\infty \int_{\mathbb{R}^p} \frac{1}{\eta^2} g(\boldsymbol{\theta}; \eta^2) \|\nabla_{\boldsymbol{\theta}} k_j\|^2 \pi(\eta^2) d\boldsymbol{\theta} d\eta^2
 \end{aligned}$$

In the third step, we use Cauchy Schwartz inequality.

By the definition of $k_j(\boldsymbol{\theta})$,

$$\|\nabla_{\boldsymbol{\theta}} k_j(\boldsymbol{\theta})\|^2 = \frac{1}{\|\boldsymbol{\theta}\|^2 \log^2 j} I_{(1,j)}(\|\boldsymbol{\theta}\|) \leq \frac{1}{\|\boldsymbol{\theta}\|^2 \log^2(\|\boldsymbol{\theta}\| \vee 2)} I_{(1,j)}(\|\boldsymbol{\theta}\|).$$

Note that $\|\nabla_{\boldsymbol{\theta}} k_j(\boldsymbol{\theta})\|^2 \rightarrow 0$ for each $\boldsymbol{\theta} \in \mathbb{R}^n$. The Growth Condition then yields that

$$\int_{\mathbb{R}^p} \int_0^\infty \frac{1}{\eta^2} \sup_j [\|\nabla_{\boldsymbol{\theta}} k_j(\boldsymbol{\theta})\|^2] g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2) d\eta^2 d\boldsymbol{\theta} < \infty$$

By the dominated convergence theorem, $A_j \rightarrow 0$ as $j \rightarrow \infty$.

$$\begin{aligned}
 B_j &\equiv \int_0^\infty \int_{\mathbb{R}^p} \left\| \frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{\int_0^\infty J^*(k_j^2 \nabla_{\boldsymbol{\theta}} g, \pi_j) d\eta^2}{\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2} \right\|^2 \\
 &\quad \times \left(\int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2 \right) d\mathbf{x} dw \\
 &= \int_0^\infty \int_{\mathbb{R}^p} \frac{\left\| \int_0^\infty \eta^2 J^*(g_j, \pi_j) d\eta^2 \frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \int_0^\infty J^*(k_j^2 \nabla_{\boldsymbol{\theta}} g, \pi_j) d\eta^2 \right\|^2}{\int_0^\infty \eta^2 J^*(g k_j^2, \pi_j) d\eta^2} d\mathbf{x} dw
 \end{aligned}$$

$$\begin{aligned}
&= \int_0^\infty \int_{\mathbb{R}^p} \frac{\|\int_0^\infty \eta^2 J^*(g_j(\frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{1}{\eta^2} \frac{\nabla_{\boldsymbol{\theta}} g}{g}), \pi_j) d\eta^2\|^2}{\int_0^\infty \eta^2 J^*(g k_j^2, \pi_j) d\eta^2} d\mathbf{x} d\mathbf{w} \\
&\leq \int_0^\infty \int_{\mathbb{R}^p} \int_0^\infty \eta^2 J^*(g_j \|\frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{1}{\eta^2} \frac{\nabla_{\boldsymbol{\theta}} g}{g}\|^2, \pi_j) d\eta^2 d\mathbf{x} d\mathbf{w} \\
&\leq \int_0^\infty \int_{\mathbb{R}^p} \int_0^\infty \eta^2 J^*(g \|\frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{1}{\eta^2} \frac{\nabla_{\boldsymbol{\theta}} g}{g}\|^2, \pi) d\eta^2 d\mathbf{x} d\mathbf{w}
\end{aligned}$$

In the fourth step, we apply Cauchy Schwartz inequality.

The integrand of B_j goes to zero as $j \rightarrow \infty$. By the Asymptotic Flatness Condition, B_j is upper bounded uniformly in j . Then by the dominated convergence theorem, $B_j \rightarrow 0$. Applying triangular inequality, it is easy to show that $\Delta_j \leq 2A_j + 2B_j$. Combination of the results for A_j and B_j leads to the result that $\Delta_j \rightarrow 0$ as $j \rightarrow \infty$. Condition 4 guarantees that the sequence of measures described in (33) and (34) are finite. By Lemma 3 (Blyth's method), the proof is complete.

Proof of Lemma 2: By

$$\begin{aligned}
&\int_0^\infty \eta^2 J^*(g \|\frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{1}{\eta^2} \frac{\nabla_{\boldsymbol{\theta}} g}{g}\|^2, \pi) d\eta^2 \\
&= \int_0^\infty \frac{1}{\eta^2} J^*(\frac{\|\nabla_{\boldsymbol{\theta}} g\|^2}{g}, \pi) d\eta^2 - \frac{\|\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2\|^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} \\
&\leq \int_0^\infty \frac{1}{\eta^2} J^*(\frac{\|\nabla_{\boldsymbol{\theta}} g\|^2}{g}, \pi) d\eta^2,
\end{aligned}$$

we have

$$\begin{aligned}
&\int_{\mathbb{R}^p} \int_0^\infty \int_0^\infty \eta^2 J^*(g \|\frac{\int_0^\infty J^*(\nabla_{\boldsymbol{\theta}} g, \pi) d\eta^2}{\int_0^\infty \eta^2 J^*(g, \pi) d\eta^2} - \frac{1}{\eta^2} \frac{\nabla_{\boldsymbol{\theta}} g}{g}\|^2, \pi) d\eta^2 d\mathbf{w} d\mathbf{x} \\
&\leq \int_{\mathbb{R}^p} \int_0^\infty \int_0^\infty \frac{1}{\eta^2} J^*(\frac{\|\nabla_{\boldsymbol{\theta}} g\|^2}{g}, \pi) d\eta^2 d\mathbf{w} d\mathbf{x} \\
&= \int_{\mathbb{R}^p} \int_0^\infty \int_0^\infty \int_{\mathbb{R}^p} \frac{1}{\eta^2} f(\mathbf{x}|\boldsymbol{\theta}, \eta^2) f(s|\eta^2) \frac{\|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}; \eta^2)\|^2}{g(\boldsymbol{\theta}; \eta^2)} \pi(\eta^2) d\boldsymbol{\theta} d\eta^2 d\mathbf{w} d\mathbf{x} \\
&= \int_0^\infty \int_{\mathbb{R}^p} \frac{1}{\eta^2} \frac{\|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}; \eta^2)\|^2}{g(\boldsymbol{\theta}; \eta^2)} \pi(\eta^2) d\boldsymbol{\theta} d\eta^2.
\end{aligned}$$

Thus in Theorem 1, when prior density functions $g(\boldsymbol{\theta}; \eta^2)$ and $\pi(\eta^2)$ satisfy Condition 1, 3 and 4, the corresponding generalized Bayes estimator for normal mean $\boldsymbol{\theta}$ is admissible.

Proof of Theorem 2: We apply Theorem 1 and Lemma 2 to the hierarchical Bayes model, and check Condition 1, 3 and 4 for the prior density functions.

For Condition 1:

$$\begin{aligned}
 & \int_{S^c} \int_0^\infty \frac{1}{\eta^2} \frac{g(\boldsymbol{\theta}; \eta^2)}{\|\boldsymbol{\theta}\|^2 \log^2(\|\boldsymbol{\theta}\| \vee 2)} \pi(\eta^2) d\eta^2 d\boldsymbol{\theta} \\
 &= \int_{S^c} \int_0^\infty \int_0^\infty \frac{1}{\eta^2} \frac{g(\boldsymbol{\theta}; \nu, \eta^2)}{\|\boldsymbol{\theta}\|^2 \log^2(\|\boldsymbol{\theta}\| \vee 2)} h(\nu) \pi(\eta^2) d\eta^2 d\nu d\boldsymbol{\theta} \\
 &= \int_{S^c} \int_0^\infty \frac{\nu^{b-k-1/2} (1+\nu)^{-a-b-2}}{\|\boldsymbol{\theta}\|^{p-2k+1} \log^2(\|\boldsymbol{\theta}\| \vee 2)} d\nu d\boldsymbol{\theta}
 \end{aligned}$$

In the second step, we integrate with respect to $d\eta^2$.

By polar coordinate transformation,

$$\int_{S^c} \frac{1}{\|\boldsymbol{\theta}\|^{p-2k+1} \log^2(\|\boldsymbol{\theta}\| \vee 2)} d\boldsymbol{\theta} \propto \int_1^\infty \frac{z^{p-1}}{z^{p-2k+1} \log^2(z \vee 2)} dz.$$

Thus when $p - 2k + 1 - (p - 1) \geq 1$, that is $k \leq 1/2$,

$$\int_{S^c} \frac{1}{\|\boldsymbol{\theta}\|^{p-2k+1} \log^2(\|\boldsymbol{\theta}\| \vee 2)} d\boldsymbol{\theta} < \infty.$$

By variable transformation, let $t = 1/(1 + \nu)$, then

$$\int_0^\infty \nu^{b-k-1/2} (1+\nu)^{-a-b-2} d\nu = \int_0^1 t^{a+k+1/2} (1-t)^{b-k-1/2} dt.$$

When $b - k - 1/2 > -1$, and $a + k + 1/2 > -1$, the last line is finite.

Hence when $k \leq 1/2$, $k < b + 1/2$, $k > -a - 3/2$, the Maruyama and Strawderman (2005) prior satisfies Condition 1.

Before we check Condition 3, we first calculate the order of magnitude for $g(\boldsymbol{\theta}; \eta^2)$.

$$\begin{aligned}
 (36) \quad g(\boldsymbol{\theta}; \eta^2) &= \int_0^\infty g(\boldsymbol{\theta}; \nu, \eta^2) h(\nu) d\nu \\
 &\propto \int_0^\infty \frac{\eta^p \nu^b}{\nu^{p/2} (1+\nu)^{a+b+2}} \exp\left(-\frac{\eta^2 \|\boldsymbol{\theta}\|^2}{2\nu}\right) d\nu \\
 &\propto \int_0^\infty \frac{\eta^p \zeta^{p/2+a+b}}{\zeta^b (\zeta+1)^{a+b+2}} \exp(-\eta^2 \|\boldsymbol{\theta}\|^2 \zeta/2) d\zeta \\
 &\propto \int_0^\infty \frac{\eta^p \xi^{p/2+a} (\eta^2 \|\boldsymbol{\theta}\|^2)^b}{(\eta^2 \|\boldsymbol{\theta}\|^2)^{p/2-2} (\eta^2 \|\boldsymbol{\theta}\|^2 + \xi)^{a+b+2} (\eta^2 \|\boldsymbol{\theta}\|^2)} \exp(-\xi/2) d\xi \\
 &\propto \frac{\eta^{2+2b}}{\|\boldsymbol{\theta}\|^{p-2-2b}} \int_0^\infty \frac{\xi^{p/2+a}}{(\eta^2 \|\boldsymbol{\theta}\|^2 + \xi)^{a+b+2}} \exp(-\xi/2) d\xi.
 \end{aligned}$$

In the third step, we let $\zeta = 1/\nu$, and in the fourth step, we let $\xi = \eta^2 \|\boldsymbol{\theta}\|^2 \zeta$.

This implies that

$$(37) \quad g(\boldsymbol{\theta}; \eta^2) \approx \frac{\eta^{2+2b}}{\|\boldsymbol{\theta}\|^{p-2-2b} (\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}}.$$

Here we use the symbol “ \approx ” to mean “is of the exact order of”; ie, $a(x) \approx b(x)$ means there is an $\epsilon > 0$ such that $\epsilon a(x) < b(x) < \epsilon^{-1} a(x)$ for all x .

We can also show that

$$(38) \quad \nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}; \eta^2) \approx \frac{-\eta^{2+2b} \boldsymbol{\theta}}{\|\boldsymbol{\theta}\|^{p-2-2b} (\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} \left(\frac{1}{\|\boldsymbol{\theta}\|^2} + \frac{(a+b+2)\eta^2}{\eta^2 \|\boldsymbol{\theta}\|^2 + 1} \right).$$

To proof (37), the integration in (36) can be written as

$$\left\{ \int_0^1 + \int_1^\infty \right\} \frac{\xi^{p/2+a}}{(\eta^2 \|\boldsymbol{\theta}\|^2 + \xi)^{a+b+2}} \exp(-\xi/2) d\xi \equiv I_1 + I_2$$

Note that $a + b + 2 > 0$. For the upper bound of $I_1 + I_2$:

$$\begin{aligned} I_1 &= \int_0^1 \frac{\xi^{p/2+a}}{(\eta^2 \|\boldsymbol{\theta}\|^2 / \xi + 1)^{a+b+2} \xi^{a+b+2}} \exp(-\xi/2) d\xi \\ &< \int_0^1 \frac{\xi^{p/2+a}}{(\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2} \xi^{a+b+2}} \exp(-\xi/2) d\xi \\ &\equiv \frac{1}{(\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} A. \end{aligned}$$

The second step is correct because $\xi < 1$.

$$\begin{aligned} I_2 &\leq \int_1^\infty \frac{\xi^{p/2+a}}{(\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} \exp(-\xi/2) d\xi \\ &\equiv \frac{1}{(\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} B. \end{aligned}$$

The first step is correct because $\xi > 1$.

For the lower bound of $I_1 + I_2$:

$$\begin{aligned} I_1 &> \int_0^1 \frac{\xi^{p/2+a}}{(\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} \exp(-\xi/2) d\xi \\ &\equiv \frac{1}{(\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} C. \\ I_2 &= \int_1^\infty \frac{\xi^{p/2+a}}{(\eta^2 \|\boldsymbol{\theta}\|^2 / \xi + 1)^{a+b+2} \xi^{a+b+2}} \exp(-\xi/2) d\xi \end{aligned}$$

$$\begin{aligned}
 &> \int_1^\infty \frac{\xi^{p/2+a}}{(\eta^2\|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}\xi^{a+b+2}} \exp(-\xi/2) d\xi \\
 &\equiv \frac{1}{(\eta^2\|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} D.
 \end{aligned}$$

Therefore, we have shown that

$$I_1 + I_2 \approx \frac{1}{(\eta^2\|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}},$$

which implies (37).

To prove (38), We can differentiate with respect to $\boldsymbol{\theta}$ on both sides of (36), and then use the same argument as above. Therefore, (38) is correct.

For Condition 3:

$$\begin{aligned}
 &\int_{\mathbb{R}^p} \int_0^\infty \frac{1}{\eta^2} \frac{\|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}; \eta^2)\|^2}{g(\boldsymbol{\theta}; \eta^2)} \pi(\eta^2) d\eta^2 d\boldsymbol{\theta} \\
 &\approx \int_{\mathbb{R}^p} \int_0^\infty \frac{\|\boldsymbol{\theta}\|^2}{\|\boldsymbol{\theta}\|^{p-2b-2} (\eta^2\|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} \\
 &\quad \times \left(\frac{1}{\|\boldsymbol{\theta}\|^2} + \frac{(a+b+2)\eta^2}{\eta^2\|\boldsymbol{\theta}\|^2 + 1} \right)^2 \frac{1}{\eta^{2k-2b}} d\eta d\boldsymbol{\theta} \\
 &\approx \int_{\mathbb{R}^p} \int_0^\infty \left(\frac{1}{\|\boldsymbol{\theta}\|^2} + \frac{(a+b+2)\|\boldsymbol{\theta}\|^2\eta^4}{(\eta^2\|\boldsymbol{\theta}\|^2 + 1)^2} \right) \\
 &\quad \times \frac{1}{\|\boldsymbol{\theta}\|^{p-2b-2} (\eta^2\|\boldsymbol{\theta}\|^2 + 1)^{a+b+2} \eta^{2k-2b}} d\eta d\boldsymbol{\theta} \\
 &\propto \int_0^\infty \int_0^\infty \left(1 + \frac{(a+b+2)^2 r^4 \eta^4}{(\eta^2 r^2 + 1)^2} \right) \frac{1}{r^{1-2b} (\eta^2 r^2 + 1)^{a+b+2} \eta^{2k-2b}} d\eta dr \\
 (39) \quad &\approx \int_0^\infty \int_0^\infty \frac{1}{r^{1-2b} (\eta^2 r^2 + 1)^{a+b+2} \eta^{2k-2b}} d\eta dr.
 \end{aligned}$$

In the third step, we change variables to $r = \|\boldsymbol{\theta}\|$. To make (39) be finite, when $\eta \rightarrow \infty$ we need $2(a+b+2) + 2k - 2b > 1$, when $\eta \rightarrow 0$ we need $2k - 2b < 1$, when $r \rightarrow 0$ we need $1 - 2b < 1$ and when $r \rightarrow \infty$ we need $1 - 2b + 2(a+b+2) > 1$. Combining the restrictions for (39) to be finite, we have

$$-a - 3/2 < k < b + 1/2, \quad b > 0, \quad a > -2.$$

For Condition 4, when $k < b + 1/2$ and $b > 0$, we have

$$\int_{\|\boldsymbol{\theta}\| < B} \int_{\eta^2 < B} g(\boldsymbol{\theta}; \eta^2) \pi(\eta^2) d\eta^2 d\boldsymbol{\theta}$$

$$\begin{aligned}
&\approx \int_{\|\boldsymbol{\theta}\| < B} \int_{\eta^2 < B} \frac{\eta^{2+2b}}{\|\boldsymbol{\theta}\|^{p-2-2b} (\eta^2 \|\boldsymbol{\theta}\|^2 + 1)^{a+b+2}} \frac{1}{\eta^{2k}} d\eta d\boldsymbol{\theta} \\
&\approx \int_{\|\boldsymbol{\theta}\| < B} \frac{1}{\|\boldsymbol{\theta}\|^{p-2}} d\boldsymbol{\theta} \\
&\propto \int_{r < B} r^{1+2b} dr < \infty.
\end{aligned}$$

Combining all the results above, when $-a - 3/2 < k \leq 1/2$, $a > -2$ and $b > 0$, the corresponding Maruyama and Strawderman (2005) priors lead to admissible generalized Bayes estimators for the normal mean $\boldsymbol{\theta}$.

To prove Theorem 3, the following lemma is very useful:

LEMMA 5. *Let $\boldsymbol{\delta}_0$ be any admissible procedure and Θ be the parameter space of $\boldsymbol{\theta}$ and η . If there is a continuous function $H : \Theta \rightarrow (c, \infty)$ for $c > 0$ and a procedure $\boldsymbol{\delta}'$ with $R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}') < \infty$ for all $\boldsymbol{\theta}$ and η , and*

$$\liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} H(\boldsymbol{\theta}, \eta)(R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_0) - R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}')) = \lambda > 0$$

then $\boldsymbol{\delta}_0$ is generalized Bayes for some (generalized) prior G satisfying

$$\int \int H^{-1}(\boldsymbol{\theta}, \eta) G(d\boldsymbol{\theta}, d\eta) < \infty.$$

This result has been given by Brown (1979, 1980). It shows that procedures which can be improved on in a neighborhood of infinity are either inadmissible or are generalized Bayes for a (possibly improper) prior whose rate of growth at infinity is of an appropriate order.

Proof of Theorem 3: We apply Lemma 5 to the hierarchical Bayes model (20) and obtain the necessary condition for $\boldsymbol{\delta}_G$ to be admissible. Let $\boldsymbol{\delta}_G(x, w)$ be an admissible generalized Bayes estimator. Let $H(\boldsymbol{\theta}, \eta) = \max(\|\boldsymbol{\theta}\|^2 \eta^2, 1)$. First we will show that

$$\int_0^\infty \int_0^\infty \int_{\mathbb{R}^p} H^{-1}(\boldsymbol{\theta}, \eta) G(d\boldsymbol{\theta}, d\eta, d\nu) = \infty$$

Proof:

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \int_{\mathbb{R}^p} H^{-1}(\boldsymbol{\theta}, \eta) G(d\boldsymbol{\theta}, d\eta, d\nu) \\
&\propto \int_0^\infty \int_0^\infty \int_{\mathbb{R}^p} \frac{1}{\max(\|\boldsymbol{\theta}\|^2 \eta^2, 1)} \frac{\eta^p}{\nu^{p/2}} \exp\left(-\frac{\eta^2 \|\boldsymbol{\theta}\|^2}{2\nu}\right) \\
&\quad \times \nu^b (1 + \nu)^{-a-b-2} \frac{1}{\eta^{2k+1}} d\boldsymbol{\theta} d\eta^2 d\nu
\end{aligned}$$

$$\begin{aligned}
 &> \int_0^\infty \int_0^\infty \int_{\|\boldsymbol{\theta}\|^2 \eta^2 \leq 1} \frac{\eta^p}{\nu^{p/2}} \exp\left(-\frac{\eta^2 \|\boldsymbol{\theta}\|^2}{2\nu}\right) \\
 &\quad \times \nu^b (1+\nu)^{-a-b-2} \frac{1}{\eta^{2k+1}} d\boldsymbol{\theta} d\eta^2 d\nu \\
 &\approx \int_0^\infty \int_{\|\boldsymbol{\theta}\|^2 \eta^2 \leq 1} \eta^{2b-2k+3} \|\boldsymbol{\theta}\|^{-p+2b+2} (\|\boldsymbol{\theta}\|^2 \eta^2 + 1)^{-a-b-2} d\boldsymbol{\theta} d\eta^2 \\
 (40) \quad &\propto \int_0^\infty \int_{R^2 \eta^2 \leq 1} \frac{1}{(\eta^2)^{k-b-3/2}} \frac{1}{R^{-1-2b}} \frac{1}{(R^2 \eta^2 + 1)^{a+b+2}} dR d\eta^2.
 \end{aligned}$$

In the third step, we use the result in the proof of Theorem 2 that

$$\int \nu^{b-p/2} \exp\left(-\frac{\eta^2 \|\boldsymbol{\theta}\|^2}{2\nu}\right) (1+\nu)^{-a-b-2} d\nu \approx \|\boldsymbol{\theta}\eta\|^{-p+2b+2} (\|\boldsymbol{\theta}\eta\|^2 + 1)^{-a-b-2}.$$

In the fourth step, we use the polar coordinate transformation for $\boldsymbol{\theta}$.

(40) is infinity because

$$\begin{aligned}
 &\int_0^\infty \int_{R^2 \eta^2 \leq 1} \frac{1}{(\eta^2)^{k-b-3/2}} \frac{1}{R^{-1-2b}} \frac{1}{(R^2 \eta^2 + 1)^{a+b+2}} dR d\eta^2 \\
 &> \int_0^\infty \int_0^{1/R^2} \frac{1}{(\eta^2)^{k-b-3/2}} \frac{1}{R^{-1-2b}} \max\left(\left(\frac{1}{2}\right)^{a+b+2}, 1\right) d\eta^2 dR \\
 &\propto \int_0^\infty \frac{1}{R^{4-2k}} dR = \infty.
 \end{aligned}$$

The proof is complete.

If we compare the risk function of $\boldsymbol{\delta}_G$ with that of the James-Stein estimator $\boldsymbol{\delta}_{J-S}$ in the neighborhood of infinity, then

$$\liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} H(\boldsymbol{\theta}, \eta) (R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_G) - R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_{J-S})) = \lambda > 0$$

is not true. Otherwise, it is a contradiction with Lemma 5.

LEMMA 6. When $\frac{p/2+a+1}{m/2-k-1/2-a} < \frac{p-2}{m+2}$ and $H(\boldsymbol{\theta}, \eta) = \max(\|\boldsymbol{\theta}\|^2 \eta^2, 1)$, then

$$\liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} H(\boldsymbol{\theta}, \eta) (R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_G) - R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_{J-S})) = \lambda > 0.$$

Proof of Lemma 6: Let $z = \|\mathbf{x}\|^2/w$. We compare the risk function for $\boldsymbol{\delta}_G$ in (27) and (28) with the risk function for James-Stein estimator $\boldsymbol{\delta}_{J-S}$ when $\|\boldsymbol{\theta}\|\eta \rightarrow \infty$, where

$$\begin{aligned}
 \boldsymbol{\delta}_G &= \left(1 - \frac{\phi(z)}{z}\right) \mathbf{x} \equiv \mathbf{x} + \gamma_G(\mathbf{x}, w) \\
 \boldsymbol{\delta}_{J-S} &= \left(1 - \frac{c}{z}\right) \mathbf{x} \equiv \mathbf{x} + \gamma_{J-S}(\mathbf{x}, w)
 \end{aligned}$$

with $c = \frac{p-2}{m+2}$.

The difference between two risk functions can be calculated by Stein's unbiased estimate of risk method:

$$\begin{aligned}
& R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_G) - R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_{J-S}) \\
&= 2E_{\boldsymbol{\theta}, \eta}(\mathbf{x} - \boldsymbol{\theta})^T \gamma_G(\mathbf{x}, w) \eta^2 - 2E_{\boldsymbol{\theta}, \eta}(\mathbf{x} - \boldsymbol{\theta})^T \gamma_{J-S}(\mathbf{x}, w) \eta^2 \\
&\quad + E_{\boldsymbol{\theta}, \eta} \|\gamma_G(\mathbf{x}, w)\|^2 \eta^2 - E_{\boldsymbol{\theta}, \eta} \|\gamma_{J-S}(\mathbf{x}, w)\|^2 \eta^2 \\
&= 2E_{\boldsymbol{\theta}, \eta} \nabla_{\mathbf{x}} \cdot [\gamma_G(\mathbf{x}, w) - \gamma_{J-S}(\mathbf{x}, w)] \\
&\quad + \eta^2 E_{\boldsymbol{\theta}, \eta} [\|\gamma_G(\mathbf{x}, w)\|^2 - \|\gamma_{J-S}(\mathbf{x}, w)\|^2],
\end{aligned}$$

where

$$\begin{aligned}
E_{\boldsymbol{\theta}, \eta} \nabla_{\mathbf{x}} \cdot \gamma_G(\mathbf{x}, w) &= -E_{\boldsymbol{\theta}, \eta} \left[(p-2) \frac{\phi(z)}{z} + 2\phi'(z) \right] \\
E_{\boldsymbol{\theta}, \eta} \nabla_{\mathbf{x}} \cdot \gamma_{J-S}(\mathbf{x}, w) &= -c E_{\boldsymbol{\theta}, \eta} \frac{(p-2)}{z}.
\end{aligned}$$

When $\|\boldsymbol{\theta}\| \rightarrow \infty$ and $\eta \rightarrow \infty$, $z = \|\mathbf{x}\|^2/w \rightarrow \infty$. $\phi(z)$ is monotonically increasing in z and

$$\begin{aligned}
\lim_{z \rightarrow \infty} \phi(z) &= \frac{p/2 + a + 1}{m/2 - k - 1/2 - a} \equiv d \\
\lim_{z \rightarrow \infty} \phi'(z) &= 0.
\end{aligned}$$

So if $d \leq c - \epsilon$ for any $\epsilon > 0$, we have

$$\begin{aligned}
& \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} H(\boldsymbol{\theta}, \eta) (R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_G) - R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_{J-S})) \\
&= \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \eta^2 [-2(p-2) E_{\boldsymbol{\theta}, \eta} \left[\frac{1}{z} (\phi(z) - c) \right] - 4E_{\boldsymbol{\theta}, \eta} \phi'(z)] \\
&\quad + \eta^2 E_{\boldsymbol{\theta}, \eta} \left[\frac{w}{z} (\phi(z) - c) (\phi(z) + c) \right] \\
&\geq \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \eta^2 [-2(p-2) E_{\boldsymbol{\theta}, \eta} \left[\frac{\phi(z) - c}{z} \right] + \eta^2 E_{\boldsymbol{\theta}, \eta} \left[\frac{w}{z} (\phi(z) - c) (2c - \epsilon) \right]] \\
&= \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \eta^2 [E_{\boldsymbol{\theta}, \eta} [(\phi(z) - c) \left(\frac{-2(p-2)}{z} + \eta^2 \frac{w}{z} 2c \right)] + \epsilon \eta^2 E_{\boldsymbol{\theta}, \eta} \left[\frac{w}{z} (c - \phi(z)) \right]] \\
&\geq E_{\boldsymbol{\theta}, \eta} \left[(d - c) \left(\liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \eta^2 \left(\frac{-2(p-2)\chi_m^2}{\eta^2 \|\mathbf{x}\|^2} + \frac{2c(\chi_m^2)^2}{\eta^2 \|\mathbf{x}\|^2} \right) \right) \right] \\
&\quad + \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \eta^2 \epsilon \eta^2 E_{\boldsymbol{\theta}, \eta} \left[\frac{w}{z} (c - d) \right]
\end{aligned}$$

$$\begin{aligned}
 &= (d - c)E\left[\liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \frac{\|\boldsymbol{\theta}\|^2}{\|\mathbf{x}\|^2}\right]E(-2(p-2)\chi_m^2 + 2c(\chi_m^2)^2) \\
 &\quad + \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \eta^2 \epsilon \eta^2 E_{\boldsymbol{\theta}, \eta} \left[\frac{w}{z}(c-d)\right] \\
 &= \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \eta^2 \epsilon \eta^2 E_{\boldsymbol{\theta}, \eta} \left[\frac{w}{z}(c-d)\right] \\
 &\geq \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \epsilon^2 E_{\boldsymbol{\theta}, \eta} \left[\frac{1}{\|\mathbf{x}\|^2}\right] E_{\eta}[(\chi_m^2)^2] \\
 &= \liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} \|\boldsymbol{\theta}\|^2 \epsilon^2 (2m + m^2) E_{\boldsymbol{\theta}, \eta} \left[\frac{1}{\|\mathbf{x}\|^2}\right].
 \end{aligned}$$

In the fifth step, we use the independence of \mathbf{X} and W . In the sixth step, we use the fact that $c = \frac{p-2}{m+2}$, $E\chi_m^2 = m$ and $E(\chi_m^2)^2 = m(m+2)$.

It is well known that $\|\mathbf{x}\|^2$, $\frac{1}{\eta^2}$ times a noncentral χ_p^2 with noncentrality parameter $\|\boldsymbol{\theta}\|^2$, is distributed the same as $\frac{1}{\eta^2}$ times a random variable V where $V|L \sim \chi_{p+2L}^2$ and $L \sim \text{Poisson}(\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2)$. Hence

$$E_{\boldsymbol{\theta}, \eta} \frac{1}{\|\mathbf{x}\|^2} = E_{\boldsymbol{\theta}, \eta} \left[\frac{\eta^2}{\chi_{p+2L}^2}\right] = E_{\boldsymbol{\theta}, \eta} [E[\frac{\eta^2}{\chi_{p+2L}^2} | L]] = E_{\boldsymbol{\theta}, \eta} \frac{\eta^2}{p-2+2L},$$

where

$$\begin{aligned}
 E_{\boldsymbol{\theta}, \eta} \frac{1}{p-2+2L} &= \sum_{l=0}^{\infty} \frac{1}{p-2+2l} \frac{e^{-\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2}}{l!} \left(\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2\right)^l \\
 &> \sum_{l=1}^{\infty} \frac{1}{(p-2)2l} \frac{e^{-\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2}}{l!} \left(\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2\right)^l \\
 &> \frac{1}{2(p-2)} \sum_{l=1}^{\infty} \frac{e^{-\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2}}{(l+1)!} \left(\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2\right)^l \\
 &= \frac{1}{2(p-2)} \frac{1}{\|\boldsymbol{\theta}\|^2\eta^2} (1 - e^{-\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2} - e^{-\frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2} \frac{1}{2}\|\boldsymbol{\theta}\|^2\eta^2).
 \end{aligned}$$

When $H(\boldsymbol{\theta}, \eta) = \max(\|\boldsymbol{\theta}\|^2\eta^2, 1)$, we have

$$\liminf_{\|\boldsymbol{\theta}\| \rightarrow \infty, \eta \rightarrow \infty} H(\boldsymbol{\theta}, \eta) (R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_G) - R(\boldsymbol{\theta}, \eta, \boldsymbol{\delta}_{J-S})) > 0.$$

The proof of Lemma 6 is complete.

By Lemma 6, if $\boldsymbol{\delta}_G$ is an admissible generalized Bayes estimator, then $\frac{p/2+a+1}{m/2-k-1/2-a} \geq \frac{p-2}{m+2}$. Otherwise, it is a contradiction with Lemma 5. After some algebra, it is equivalent that $(a+2)m + (k+a+3/2)p + 1 - 2k \geq 0$. The proof of Theorem 3 is complete.

Acknowledgements. We thank the Editor Professor Peter Bühlmann, the Associate Editor and two referees for constructive comments which help us improve the presentation of this paper.

REFERENCES

- [1] Baranchik, A. J. (1970) A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Statistics* 41,642-645.
- [2] Blyth, C. R. (1951) On minimax statistical procedures and their admissibility. *Annals of Statistics* 22,22-24.
- [3] Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer, New York.
- [4] Berger, J. O. and Bernardo, J. M. (1992) On the development of the reference prior method. In *Bayesian Statistics* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.)4 35-60. Oxford University Press.
- [5] Berger, J. O. and Strawderman, W. E. (1996) Choice of hierarchical priors: Admissibility in estimation of normal means. *Annals of Statistics* 24, 931-951.
- [6] Berger, J. O., Strawderman, W. E. and Tang, D. (2005) Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Annals of Statistics* 33, 606-646.
- [7] Blattberg, R. and George, E. (1991) Shrinkage estimation of price and promotional elasticities: seemingly unrelated equations. *Journal of the American Statistical Association* Vol 86, No. 414, 304-315.
- [8] Brown, L. D. (1971) Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Annals of Statistics* 42, 855-903.
- [9] Brown, L. D. (1979) A heuristic method for determining admissibility of estimators-with applications. *Annals of Statistics* 7, 960-994.
- [10] Brown, L. D. (1980) A necessary condition for admissibility. *Annals of Statistics* 8, 540-545.
- [11] Brown, L. D. and Hwang, J. (1982) A unified admissibility proof. *Statistical Decision Theory and Related Topics III* (S.S. Gupta and J.O. Berger, eds) 1, 205-230. Academic Press, New York. MR0705290
- [12] DuMouchel, W. and Harris, J. (1983) Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association* Vol 78, No. 382, 293-308.
- [13] Efron, B. and Morris, C. (1971) Limiting the risk of Bayes and empirical Bayes estimators—part I: the Bayes case. *Journal of the American Statistical Association*, Vol 66, No. 336, 807-815.
- [14] Efron, B. and Morris, C. (1972) Limiting the risk of Bayes and empirical Bayes estimators—part I: the empirical Bayes case. *Journal of the American Statistical Association*, Vol 67, No. 337, 130-139.
- [15] Efron, B. and Morris, C. (1975) Data analysis using Stein's estimator and its generalizations *Journal of the American Statistical Association*, Vol 70, No. 350, 311-319.
- [16] Farrell, R.H. (1964) Estimators of a location parameter in the absolutely continuous case. *Annals of Statistics* 3, 949-998.
- [17] Fay, R. and Herriot, R. (1979) Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* Vol 74, No. 366, 269-277.

- [18] Gelfand, A., Hills, S., Racine-Poon, A. and Smith, A. (1990) Illustration of Bayesian inference in Normal data models using Gibbs sampling. *Journal of the American Statistical Association* Vol 85, No. 412, 972-985.
- [19] Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2003) Bayesian Data Analysis (second edition). *Chapman and Hall/CRC Texts in Statistical Science*
- [20] Han, X. (2009) Topics in Shrinkage Estimation and in Causal Inference. *Ph.D. Dissertation Paper*.
- [21] Hui, S. and Berger, J. (1983) Empirical Bayes estimation of rates in longitudinal studies. *Journal of the American Statistical Association* Vol 78, No. 384, 753-760.
- [22] James, W. and Stein, C. (1961) Estimation with quadratic loss. *Proceedings Fourth Berkeley Symposium in Mathematical Statistics and Probability* 1, 361-3679.
- [23] Jorion, P. (1986) Bayes-Stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis* Vol 21, No.3, 279-292.
- [24] Lehmann, E. L. and Casella, G. (1998) Theory of Point Estimation, 2nd Edition. *Springer*.
- [25] Maruyama, Y. and Strawderman, W. E. (2005) A new class of generalized Bayes minimax ridge regression estimators. *Annals of Statistics* 4, 1753-1770.
- [26] Stein, C. (1955) A necessary and sufficient condition for admissibility. *Annals of Statistics* 3, 518-522.
- [27] Stein, C. (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings Third Berkeley Symposium in Mathematical Statistics and Probability* 1, 197-206.
- [28] Stein, C. (1973) Estimation of the mean of a multivariate normal distribution. In *Proc. Prague Symposium on Asymptotic Statistics* (J. Hájek, ed.) 2 345-381. Univ. Karlova, Prague.
- [29] Stein, C. (1981) Estimation of the mean of a multivariate normal distribution. *Annals of Statistics* 9, 1135-1151.
- [30] Strawderman, W. E. (1971) Proper Bayes minimax estimators of the multivariate normal mean. *Annals of Statistics* 42, 385-388.
- [31] Strawderman, W. E. (1973) Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *Annals of Statistics* 1, 1189-1194.
- [32] Wells, M. T. and Zhou, G. (2008) Generalized Bayes minimax estimators of the mean of multivariate normal distribution with unknown variance. *Journal of Multivariate Analysis* 99, 2208-2220.
- [33] Zellner, A. and Hong, C. (1989) Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics* 40, 183-202.

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
400 JON M. HUNTSMAN HALL
3730 WALNUT ST.
PHILADELPHIA, PA 19104-6340
E-MAIL: lbrown@wharton.upenn.edu

DEPARTMENT OF OPERATIONS RESEARCH
& FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
ROOM 214, SHERRERD HALL
PRINCETON, NJ 08544
E-MAIL: xhan@princeton.edu