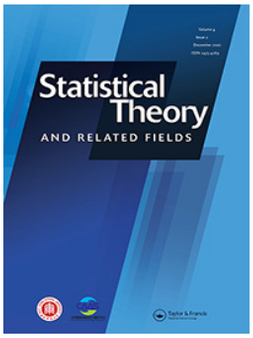

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

A class of admissible estimators of multiple regression coefficient with an unknown variance

Chengyuan Song & Dongchu Sun

To cite this article: Chengyuan Song & Dongchu Sun (2020) A class of admissible estimators of multiple regression coefficient with an unknown variance, Statistical Theory and Related Fields, 4:2, 190-201, DOI: 10.1080/24754269.2019.1653160

To link to this article: https://doi.org/10.1080/24754269.2019.1653160

Published online: 20 Aug 2019.

STATISTICAL THEORY AND RELATED FIELDS
2020, VOL. 4, NO. 2, 190–201
https://doi.org/10.1080/24754269.2019.1653160

A class of admissible estimators of multiple regression coefficient with an unknown variance

Chengyuan Song[a] and Dongchu Sun[a,b]

[a]Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China; [b]Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE, USA

ABSTRACT
Suppose that we observe $\boldsymbol{y} \mid \boldsymbol{\theta}, \tau \sim N_p(\boldsymbol{X\theta}, \tau^{-1}\boldsymbol{I}_p)$, where $\boldsymbol{\theta}$ is an unknown vector with unknown precision τ. Estimating the regression coefficient $\boldsymbol{\theta}$ with known τ has been well studied. However, statistical properties such as admissibility in estimating $\boldsymbol{\theta}$ with unknown τ are not well studied. Han [(2009). *Topics in shrinkage estimation and in causal inference* (PhD thesis). Warton School, University of Pennsylvania] appears to be the first to consider the problem, developing sufficient conditions for the admissibility of estimating means of multivariate normal distributions with unknown variance. We generalise the sufficient conditions for admissibility and apply these results to the normal linear regression model. 2-level and 3-level hierarchical models with unknown precision τ are investigated when a standard class of hierarchical priors leads to admissible estimators of $\boldsymbol{\theta}$ under the normalised squared error loss. One reason to consider this problem is the importance of admissibility in the hierarchical prior selection, and we expect that our study could be helpful in providing some reference for choosing hierarchical priors.

ARTICLE HISTORY
Received 10 June 2019
Revised 18 July 2019
Accepted 21 July 2019

KEYWORDS
Admissible estimators; unknown variance; multivariate normal distributions; hierarchical models

1. Introduction

Consider a multivariate normal model,

$$\begin{aligned} \boldsymbol{y} \mid \boldsymbol{\theta}, \tau &\sim N_p(\boldsymbol{\theta}, \tau^{-1}\boldsymbol{I}_p), \text{ independently,} \\ w \mid \tau &\sim \chi^2_m/\tau, \end{aligned} \tag{1}$$

where $\boldsymbol{y}$ is a $p \times 1$ observation vector, $\boldsymbol{\theta}$ is a p-dimensional vector of unknown parameters, and $\tau > 0$ is the unknown precision. Statistical properties such as admissibility for estimating $\boldsymbol{\theta}$ can be dated back to James-Stein (1961) when the error variance is known, while the admissibility of generalisation of James-Stein estimator of $\boldsymbol{\theta}$ with unknown parameter τ was studied in Judge, Yancey, and Bock (1983), Fraisse, Raoult, Robert, and Roy (1990), Robert (2007) and so on. For estimating $\boldsymbol{\theta}$ with the unknown nuisance parameter τ in the model (1), some authors, such as Strawderman (1973), Maruyama and Strawderman (2005) and Willing and Zhou (2008) studied the minimaxity of Bayesian estimators of $\boldsymbol{\theta}$ under hierarchical priors. The admissibility of a generalised Bayesian estimator of $\boldsymbol{\theta}$ under a class of noninformative priors was recently studied in Han (2009). With additional independent observation $w \mid \tau \sim \tau^{-1}\chi^2_m$, Han (2009) found a set of sufficient conditions for the joint priors of $(\boldsymbol{\theta}, \tau)$, so that the generalised Bayesian estimator of $\boldsymbol{\theta}$ is admissible under the squared error loss. In practice, we often need to consider a normal linear regression model,

$$\boldsymbol{y} \mid \boldsymbol{\theta}, \tau \sim N_n(\boldsymbol{X\theta}, \tau^{-1}\boldsymbol{I}_n), \tag{2}$$

where $\boldsymbol{X}$ is $n \times p$ design matrix with full column rank p, $n > p$. It is of great interest to study the admissibility in estimating the unknown regression coefficients $\boldsymbol{\theta}$ with unknown τ in the normal linear regression model (2).

Several authors have described admissibility as a powerful tool for selecting satisfactory hierarchical generalised Bayesian priors. For example, Berger, Strawderman, and Tang (2005) pointed out that the 'use of objective improper priors in hierarchical modelling is of enormous practical importance, yet little is known about which such priors are good or bad. 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'. For known precision or error variance, Brown (1971) provided the necessary and sufficient condition of the admissible Bayes estimators under quadratic loss, based on a Markovian representation of the estimation problem. Recent papers related to the theoretical studies of the admissibility of estimators of $\boldsymbol{\theta}$ can be found in Berger and Strawderman (1996), Berger et al. (2005), Berger, Sun, and Song (2018), and so on.

However, most of the literature focussed on models of which variances are given, yet in practical problems,

CONTACT Chengyuan Song songchengyuanchina@163.com

the precision or variance is often unknown. For the admissibility in the model (2), to the best of our knowledge, very few results have been obtained because of the technical difficulty. The fundamental tool for proving admissibility for unknown precision is Blyth's method (Blyth, 1951), which proposed a sufficient admissibility condition, relating admissibility of an estimator with the existence of a sequence of prior distributions approximating this estimator. Based on Blyth's results, Han's (2009) found sufficient conditions for the joint priors of $(\boldsymbol{\theta}, \tau)$ for model (1). Sometimes, those sufficient conditions are strict and difficult to satisfy. We will generalise the sufficient conditions for admissibility and apply these results to the normal linear regression model (2). Using the generalised conditions, a 2-level and 3-level hierarchical models with unknown precision τ are investigated when a standard class of hierarchical priors leads to admissible estimators of $\boldsymbol{\theta}$ under the normalised squared error loss. One motivation to consider this problem is the importance of admissibility in the hierarchical prior selection, and we expect that our study could be helpful to provide some reference for choosing hierarchical priors.

The paper is organised as follows. In Section 2, we introduce the sufficient conditions for admissibility of the generalised Bayesian estimators of $\boldsymbol{\theta}$ for the model (1), which is studied by Han (2009). In Section 3, we generalise the sufficient conditions for admissibility and apply these results to the normal linear regression model (2). 2-level and a 3-level hierarchical models with unknown precision τ are investigated in Sections 4 and 5, determining when a standard class of hierarchical priors leads to admissible estimators of $\boldsymbol{\theta}$ under the normalised squared error loss. Finally, some comments are made in Section 6.

2. Han's (2009) results for (1)

Recall the model (1) considered in Han (2009), i.e.

$$(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) \sim N_p(\boldsymbol{\theta}, \tau^{-1}\boldsymbol{I}_p), \quad (w \mid \tau) \sim \tau^{-1}\chi^2_m,$$

where $\boldsymbol{y} = (y_1, \ldots, y_p)'$ and w are independent of each other. Let $\hat{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\theta}}(\boldsymbol{y}, w)$ denote an estimator of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$. Correspondingly, the squared error loss function of $\hat{\boldsymbol{\theta}}$ becomes

$$L(\boldsymbol{\theta}, \tau; \hat{\boldsymbol{\theta}}) = \tau(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{3}$$

Han (2009) studied a class of prior density for $(\boldsymbol{\theta}, \tau)$ with assumption

$$\pi(\boldsymbol{\theta}, \tau) = \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau). \tag{4}$$

Consequently, the generalised Bayes estimator for the normal mean $\boldsymbol{\theta}$ is the posterior mean of $\boldsymbol{\theta}$, given by

$$\hat{\boldsymbol{\theta}}_B(\boldsymbol{y}, w) = \frac{\int_{\mathbb{R}^p}\int_0^\infty \tau\boldsymbol{\theta} f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta}}{\int_{\mathbb{R}^p}\int_0^\infty \tau f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta}}, \tag{5}$$

where

$$f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) \propto \tau^{p/2} \exp\left(-\frac{\tau}{2}\|\boldsymbol{y} - \boldsymbol{\theta}\|^2\right),$$

$$f_2(w \mid \tau) \propto w^{(m-2)/2}\tau^{m/2} \exp\left(-\frac{\tau w}{2}\right).$$

Let $m(\boldsymbol{y}, w, \tau)$ be the marginal likelihood function of $(\boldsymbol{y}, w, \tau)$ with the form

$$m(\boldsymbol{y}, w, \tau) = \int_{\mathbb{R}^p} f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \times \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{\theta}.$$

From Brown (1971), the generalised Bayes estimator in (5) can be expressed as

$$\hat{\boldsymbol{\theta}}_B(\boldsymbol{y}, w) = \boldsymbol{y} + \frac{\int_0^\infty \nabla_{\boldsymbol{y}} m(\boldsymbol{y}, w, \tau)\, \mathrm{d}\tau}{\int_0^\infty \tau m(\boldsymbol{y}, w, \tau)\, \mathrm{d}\tau}, \tag{6}$$

where ∇ denotes the gradient. Let S denote the ball of radius 1 at the origin in IR^p and S^c be the complement of S, defined $a \vee b = \max(a, b)$. For the hierarchical Bayes model (1), Han (2009) studied the admissible generalised Bayes estimators $\hat{\boldsymbol{\theta}}_B(\boldsymbol{y}, w)$ under the following sufficient conditions.

Condition 1. $\int_{S^c}\int_0^\infty (1/\tau)(\pi_0(\boldsymbol{\theta} \mid \tau)/\|\boldsymbol{\theta}\|^2 \log(\|\boldsymbol{\theta}\| \vee 2))\pi_1(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta} < \infty$;
Condition 2. $\int_{S^c}\int_0^\infty \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta} < \infty$;
Condition 3. $\int_{S^c}\int_0^\infty \tau\|\boldsymbol{\theta}\|^2\pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta} < \infty$;
Condition 4. $\int_{S^c}\int_0^\infty (1/\tau)(\|\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta} \mid \tau)\|^2/\pi_0(\boldsymbol{\theta} \mid \tau))\pi_1(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta} < \infty$;
Condition 5. For any positive constant B, $\int_{\|\boldsymbol{\theta}\|^2 < B}\int_{\tau < B} \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta} < \infty$;
Condition 6. Define two sequences of functions

$$h_j(\boldsymbol{\theta}) = \begin{cases} 1, & \text{if} \|\boldsymbol{\theta}\| < 1; \\ 1 - \dfrac{\log(\|\boldsymbol{\theta}\|)}{\log j}, & 1 \le \|\boldsymbol{\theta}\| \le j; \\ 0, & \|\boldsymbol{\theta}\| > j, \end{cases} \quad \text{and}$$

$$l_j(\tau) = \begin{cases} 1, & \text{if } \tau < 1; \\ 1 - \dfrac{\log(\tau)}{\log j}, & 1 \le \tau \le j; \\ 0, & \tau > j. \end{cases} \tag{7}$$

Write $H_j(\boldsymbol{\theta} \mid \tau) = h_j(\boldsymbol{\theta})\pi_0(\boldsymbol{\theta} \mid \tau)$ and $L_j(\tau) = l_j(\tau)\pi_1(\tau)$. There is a constant $C > 0$, such that

$$\frac{\int_S\int_0^\infty \tau f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) H_j(\boldsymbol{\theta} \mid \tau) L_j(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta}}{\int_{S^c}\int_0^\infty \tau f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) H_j(\boldsymbol{\theta} \mid \tau) L_j(\tau)\, \mathrm{d}\tau\, \mathrm{d}\boldsymbol{\theta}} < C, \quad \forall\, \boldsymbol{y}, w.$$

Theorem 2.1 (Han, 2009): *Consider the model* (1) *with the prior densities* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *and* $\pi_1(\tau)$ *satisfying Conditions* 1–6. *If* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *is decreasing with respect to* $\|\theta\|$, *the corresponding generalised Bayes estimator* (6) *for* $\boldsymbol{\theta}$ *is admissible under the squared error loss function* (3).

3. Main results for (2)

We are primarily interested in the normal linear regression model (2). For the model (2), we let $\tilde{\boldsymbol{y}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y}$ denote the least squared estimators of $\boldsymbol{\theta}$, and $w = \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}$ be the usual residual sum squares errors (SSE). Then

$$\tilde{\boldsymbol{y}} \mid \boldsymbol{\theta}, \tau \sim N_p(\boldsymbol{\theta}, \tau^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}), \quad \text{and}$$
$$w \mid \tau \sim \tau^{-1}\chi^2_{n-p}, \tag{8}$$

independently. Here we obtain w automatically with $m = n - p$. For the model (2), consider the normalised squared error loss function of $\hat{\boldsymbol{\theta}}$ given by

$$L(\boldsymbol{\theta}, \tau; \hat{\boldsymbol{\theta}}) = \tau(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{9}$$

The corresponding risk function of $\hat{\boldsymbol{\theta}}$ is

$$R(\boldsymbol{\theta}, \tau; \hat{\boldsymbol{\theta}}) = E_{\boldsymbol{\theta},\tau} L(\boldsymbol{\theta}, \tau; \hat{\boldsymbol{\theta}}), \quad \boldsymbol{\theta} \in \mathbb{R}^p. \tag{10}$$

An estimator $\hat{\boldsymbol{\theta}}_1$ is *inadmissible* if there exists another estimator whose risk function is nowhere bigger and somewhere smaller. If no such better estimator exists, $\hat{\boldsymbol{\theta}}_1$ is *admissible*.

For the model (2), to obtain the admissible estimator of $\boldsymbol{\theta}$ under the normalised squared error loss (9), we define

$$\boldsymbol{\delta}_B(\boldsymbol{y}) = (\boldsymbol{T}')^{-1}\hat{\boldsymbol{\theta}}_B(\boldsymbol{T}'\tilde{\boldsymbol{y}}, w), \tag{11}$$

where $\boldsymbol{T}$ is a $p \times p$ matrix such that $\boldsymbol{T}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{T} = \boldsymbol{I}_p$.

Lemma 3.1: *For the model* (1), *assume the estimator* $\hat{\boldsymbol{\theta}}_B(\boldsymbol{y}, w)$ *in* (6) *is admissible under the loss function* (3). *Then the estimator* $\boldsymbol{\delta}_B(\boldsymbol{y})$ *in* (11) *is admissible under the normalised squared error loss* (9) *for the model* (2).

Proof: Note that the model (2) is equivalent to (8). It yields that

$$(\boldsymbol{T}'\tilde{\boldsymbol{y}} \mid \boldsymbol{\theta}, \tau) \sim N_p(\boldsymbol{T}'\boldsymbol{\theta}, \tau^{-1}\boldsymbol{I}_p). \tag{12}$$

It follows from the admissibility of $\hat{\boldsymbol{\theta}}_B(\boldsymbol{y}, w)$ under the model (1) that the estimator $\hat{\boldsymbol{\theta}}_B(\boldsymbol{T}'\tilde{\boldsymbol{y}}, w)$ for $\boldsymbol{T}'\boldsymbol{\theta}$ is admissible under the loss function

$$\begin{aligned} L(\boldsymbol{T}'\boldsymbol{\theta}, \tau; \hat{\boldsymbol{\theta}}_B(\boldsymbol{T}'\tilde{\boldsymbol{y}}, w)) &= \tau\left[\hat{\boldsymbol{\theta}}_B(\boldsymbol{T}'\tilde{\boldsymbol{y}}, w) - \boldsymbol{T}'\boldsymbol{\theta}\right]' \\ &\quad \times \left[\hat{\boldsymbol{\theta}}_B(\boldsymbol{T}'\tilde{\boldsymbol{y}}, w) - \boldsymbol{T}'\boldsymbol{\theta}\right] \\ &= \tau\left[(\boldsymbol{T}')^{-1}\hat{\boldsymbol{\theta}}_B(\boldsymbol{T}'\tilde{\boldsymbol{y}}, w) - \boldsymbol{\theta}\right]'(\boldsymbol{X}'\boldsymbol{X}) \\ &\quad \times \left[(\boldsymbol{T}')^{-1}\hat{\boldsymbol{\theta}}_B(\boldsymbol{T}'\tilde{\boldsymbol{y}}, w) - \boldsymbol{\theta}\right] \\ &= \tau(\boldsymbol{\delta}_B - \boldsymbol{\theta})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\delta}_B - \boldsymbol{\theta}). \end{aligned}$$

The proof of this lemma is completed. ■

Combining Theorem 1 with Lemma 1, we can reach the following theorem.

Theorem 3.2: *For the model* (2) *with the prior densities* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *and* $\pi_1(\tau)$ *satisfying Conditions* 1–6, *suppose that* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *is decreasing with respect to* $\|\theta\|$, *then* $\boldsymbol{\delta}_B(\boldsymbol{y})$ *defined in* (11) *is the admissible estimator of* $\boldsymbol{\theta}$ *under the normalised squared error loss* (9).

Theorem 2.1 applies to the case where $\pi_0(\boldsymbol{\theta} \mid \tau)$ is spherically symmetric of $\boldsymbol{\theta}$ and decreases in $\|\boldsymbol{\theta}\|$. As discussed in Han (2009), this requirement is not unique and can be replaced by the following condition.

Condition 7. Denote

$$\begin{aligned} \boldsymbol{u}_1 &= \int_0^\infty \int_S f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \nabla_{\boldsymbol{\theta}} \\ &\quad \times \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau, \end{aligned} \tag{13}$$

$$\begin{aligned} z_1 &= \int_0^\infty \int_S \tau f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \\ &\quad \times \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau, \end{aligned} \tag{14}$$

$$\begin{aligned} \boldsymbol{u}_{2j} &= \int_0^\infty \int_S f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \\ &\quad \times \nabla_{\boldsymbol{\theta}} H_j(\boldsymbol{\theta} \mid \tau) L_j(\tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau, \end{aligned} \tag{15}$$

$$\begin{aligned} z_{2j} &= \int_0^\infty \int_S \tau f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \\ &\quad \times H_j(\boldsymbol{\theta} \mid \tau) L_j(\tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau. \end{aligned} \tag{16}$$

We have

$$\frac{\|\boldsymbol{u}_1\|}{z_1} \le \|\boldsymbol{y}\|, \tag{17}$$

$$\frac{\|\boldsymbol{u}_{2j}\|}{z_{2j}} \le \|\boldsymbol{y}\|, \quad \text{for any } j = 1, 2, \ldots. \tag{18}$$

As an immediate corollary, we have the following result.

Theorem 3.3: *For the model* (2), *assume that the prior densities* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *and* $\pi_1(\tau)$ *satisfy Conditions* 1–7. *Then estimator* $\boldsymbol{\delta}_B(\boldsymbol{y})$ *in* (11) *for* $\boldsymbol{\theta}$ *is admissible under the normalised squared error loss* (9).

It might be difficult to show that the $\pi_0(\boldsymbol{\theta} \mid \tau)$ is a decreasing function of $\|\boldsymbol{\theta}\|$. Interestingly, this requirement can be relaxed to the requirement that $\pi_0(\boldsymbol{\theta} \mid \tau)$ is a decreasing function of its component, θ_i^2, for $i = 1, \ldots, p$.

Lemma 3.4: *For the model* (1) *with given* τ, $\pi_0(\boldsymbol{\theta} \mid \tau)$ *is a decreasing function of* θ_i^2, *for* $i = 1, \ldots, p$, *then Condition* 7 *holds.*

Proof: For any given $\boldsymbol{y}$, there is an orthogonal matrix $\boldsymbol{Q}$, such that $\boldsymbol{Qy} = (\|\boldsymbol{y}\|, 0, \ldots, 0)'$. Without loss of generality, we can transform the coordinate system of $\boldsymbol{\theta}$ such that $\boldsymbol{y} = (\|\boldsymbol{y}\|, 0, \ldots, 0)'$. Then,

$$f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) = (2\pi)^{-p/2}\tau^{p/2} \exp\left(-\tau \frac{\|\boldsymbol{y}\|^2 + \|\boldsymbol{\theta}\|^2}{2}\right) \times \exp\left(\tau\theta_1\|\boldsymbol{y}\|\right).$$

It is easy to verify that the *ith* coordinate of $\boldsymbol{u}_1$ is

$$v_{1i} = \int_0^\infty \int_S f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \times \frac{\partial \pi_0(\boldsymbol{\theta} \mid \tau)}{\partial \theta_i} \pi_1(\tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau. \tag{19}$$

Since $\pi_0(\boldsymbol{\theta} \mid \tau)$ is a function of $(\theta_1^2, \ldots, \theta_p^2)$, $f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau)(\partial \pi_0(\boldsymbol{\theta} \mid \tau)/\partial \theta_i)\pi_1(\tau)$ is an odd function for θ_i when $i = 2, \ldots, p$. It yields $v_{12} = \cdots = v_{1p} = 0$. Therefore, $\|\boldsymbol{u}_1\| = |v_{11}|$ and $\|\boldsymbol{u}_1\|/z_1 = |v_{11}|/z_1$.

Let $f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)$ be the joint density of $(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)$, i.e.

$$f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau) = f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \pi_0(\boldsymbol{\theta} \mid \tau) \pi_1(\tau).$$

Using (19), we get

$$\begin{aligned}
&\int_0^\infty \int_S \frac{\partial f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)}{\partial \theta_1} \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau \\
&= \int_0^\infty \int_S \frac{\partial f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau)}{\partial \theta_1} f_2(w \mid \tau) \times \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau \\
&\quad + \int_0^\infty \int_S f_1(\boldsymbol{y} \mid \boldsymbol{\theta}, \tau) f_2(w \mid \tau) \times \frac{\partial \pi_0(\boldsymbol{\theta} \mid \tau)}{\partial \theta_1} \pi_1(\tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau \\
&= \int_0^\infty \int_S (\|\boldsymbol{y}\| - \theta_1)\tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau + v_{11}.
\end{aligned} \tag{20}$$

By the Divergence Theorem (Katz, 2005),

$$\int_0^\infty \int_S \frac{\partial f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)}{\partial \theta_1} \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau = \int_0^\infty \int_{\partial S} f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_2 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau, \tag{21}$$

where ∂S is the boundary of S. Combining (20) and (21), we get

$$v_{11} = \int_0^\infty \int_{\partial S} f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_2 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau - \int_0^\infty \int_S (\|\boldsymbol{y}\| - \theta_1)\tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau.$$

Then, we have

$$\begin{aligned}
\frac{v_{11}}{z_1} &= \frac{\int_0^\infty \int_{\partial S} f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_2 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau}{\int_0^\infty \int_S \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau} - \frac{\int_0^\infty \int_S (\|\boldsymbol{y}\| - \theta_1)\tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau}{\int_0^\infty \int_S \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau} \\
&= \frac{\int_0^\infty \int_{\partial S} f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_2 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau}{\int_0^\infty \int_S \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau} - \|\boldsymbol{y}\| + \frac{\int_0^\infty \int_S \theta_1 \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau}{\int_0^\infty \int_S \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\boldsymbol{\theta}_1\, \mathrm{d}\tau}.
\end{aligned}$$

Clearly,

$$\frac{\int_0^\infty \int_{\partial S} f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_2 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau}{\int_0^\infty \int_S \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_p\, \mathrm{d}\tau} \geq 0.$$

Since $f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)$ can be written by

$$f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau) = (2\pi)^{-p/2}\tau^{p/2} \exp\left(-\frac{\|\boldsymbol{y}\|^2 + \|\boldsymbol{\theta}\|^2}{2}\tau\right) \times \exp(\|\boldsymbol{y}\|\theta_1\tau) f_2(w|\tau)\pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau),$$

and $\pi_0(\boldsymbol{\theta} \mid \tau)$ is symmetric about θ_1, then

$$\frac{\int_0^\infty \int_S \theta_1 \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\tau}{\int_0^\infty \int_S \tau f(\boldsymbol{y}, w, \boldsymbol{\theta}, \tau)\, \mathrm{d}\boldsymbol{\theta}_1\, \mathrm{d}\tau} \geq 0.$$

Therefore,

$$\frac{v_{11}}{z_1} \geq -\|\boldsymbol{y}\|.$$

Since $\pi_0(\boldsymbol{\theta} \mid \tau)$ is an even function for θ_1 and decreasing in θ_1^2, then $v_{11} < 0$. Therefore, we have $\|\boldsymbol{u}_1\|/z_1 = |v_{11}|/z_1 \leq \|\boldsymbol{y}\|$. With the same argument as above, we have $\|\boldsymbol{u}_2\|/z_{2j} \leq \|\boldsymbol{y}\|$, for any $j = 1, 2, \ldots$. ■

Consequently, we obtain the following result.

Theorem 3.5: *For the model* (2), *assume that the prior densities* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *and* $\pi_1(\tau)$ *satisfy Conditions* 1–6. *If for any given* $\tau > 0$, $\pi_0(\boldsymbol{\theta} \mid \tau)$ *is decreasing in* θ_i^2, $i = 1, \ldots, p$. *the estimator* $\boldsymbol{\delta}_B(\boldsymbol{y})$ *in* (11) *for* $\boldsymbol{\theta}$ *is admissible under the normalised squared error loss* (9).

Sometimes, $\pi_0(\boldsymbol{\theta} \mid \tau)$ is not strictly a decreasing function of its component, θ_i^2, for $i = 1, \ldots, p$, but it could be a decreasing function of the components of some given orthogonal transformation. The following lemma shows that such cases also work.

Lemma 3.6: *Consider the model* (1). *Suppose there is an orthogonal matrix* $\boldsymbol{H}$, *such that* $\boldsymbol{u} = \boldsymbol{H\theta} = (u_1, \ldots, u_p)'$, *and* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *is a decreasing function of* u_i^2, *for* $i = 1, \ldots, p$, *then Condition* 7 *holds.*

Proof: It is easy to verify that the Jacobian of the transformation $\boldsymbol{u} = \boldsymbol{H\theta} = (u_1, \ldots, u_p)'$ is $J = |\partial\boldsymbol{\theta}/\partial\boldsymbol{u}| = 1$, and $\|\boldsymbol{\theta}\| = \|\boldsymbol{u}\|$. Note that

$$\|\boldsymbol{u}_1\| = \left\| \int_0^\infty \int_\Omega f_1(\boldsymbol{H}'\boldsymbol{y} \mid \boldsymbol{u}, \tau) f_2(w \mid \tau) \times \nabla_{\boldsymbol{u}} \pi_0(\boldsymbol{H}'\boldsymbol{u} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}\tau \right\|,$$

where $f_1(\boldsymbol{H}'\boldsymbol{y} \mid \boldsymbol{u}, \tau)$ is the normal density function of $\boldsymbol{H}'\boldsymbol{y}$ with mean $\boldsymbol{u}$ and variance $\tau^{-1}\boldsymbol{I}_p$, and $\Omega = \{\boldsymbol{u} : \|\boldsymbol{u}\| \le 1\}$. Similarly, we have

$$z_1 = \int_0^\infty \int_\Omega \tau f_1(\boldsymbol{H}'\boldsymbol{y} \mid \boldsymbol{u}, \tau) f_2(w \mid \tau) \times \pi_0(\boldsymbol{H}'\boldsymbol{u} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}\tau.$$

Since $\pi_0(\boldsymbol{H}'\boldsymbol{u} \mid \tau)$ s a decreasing function of u_i^2, for $i = 1, \ldots, p$, from Lemma 3.4, for any $\boldsymbol{y}$ and w, we have

$$\begin{aligned} &\left\| \int_0^\infty \int_\Omega f_1(\boldsymbol{H}'\boldsymbol{y} \mid \boldsymbol{u}, \tau) f_2(w \mid \tau) \times \nabla_{\boldsymbol{u}} \pi_0(\boldsymbol{H}'\boldsymbol{u} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}\tau \right\| \\ &\le \|\boldsymbol{H}'\boldsymbol{y}\| \int_0^\infty \int_\Omega \tau f_1(\boldsymbol{H}'\boldsymbol{y} \mid \boldsymbol{u}, \tau) f_2(w \mid \tau) \times \pi_0(\boldsymbol{H}'\boldsymbol{u} \mid \tau)\pi_1(\tau)\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}\tau, \end{aligned}$$

i.e. $\|\boldsymbol{u}_1\| \le z_1\|\boldsymbol{y}\|$. With the same argument as above, we have $\|\boldsymbol{u}_2\|/z_{2j} \le \|\boldsymbol{y}\|$, for any $j = 1, 2, \ldots$. ■

Accordingly, we get the following result.

Theorem 3.7: *For the model* (2), *assume that the prior densities* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *and* $\pi_1(\tau)$ *satisfy Conditions* 1-6. *If there is an orthogonal matrix* $\boldsymbol{H}$, *such that* $\boldsymbol{u} = \boldsymbol{H\theta} = (u_1, \ldots, u_p)'$, *and* $\pi_0(\boldsymbol{\theta} \mid \tau)$ *is a decreasing function of* u_i^2, *for* $i = 1, \ldots, p$, *the estimator* $\boldsymbol{\delta}_B(\boldsymbol{y})$ *in* (11) *for* $\boldsymbol{\theta}$ *is admissible under the normalised squared error loss* (9).

In the next two sections, we will apply the above results to a 2-level and a 3-level hierarchical model, with unknown variance and a standard class of hierarchical priors.

4. Admissibility for a 2-level hierarchical model

4.1. g-Prior

For the model (2), we consider the following class of hierarchical prior for $(\boldsymbol{\theta}, \tau)$,

$$(\boldsymbol{\theta} \mid g, \tau) \sim N_p(\boldsymbol{0}, g\tau^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}), \quad \pi_1(\tau) \propto \frac{1}{\tau^k}, \tag{22}$$

where $k \ge 0$. Zellner (1986) proposed this form of the conjugate Normal-Gamma family with $k = 1$. Many authors followed his work, for example, Eaton (1989), Berger, Pericchi, and Varshavsky (1998), Liang, Paulo, Molina, Clyde, and Berger (2008) and Bayarri, Berger, Forte, and Garcła-Donato (2012). From the perspective of model selection, g acts as a dimensionality penalty (Liang et al., 2008). For the choice of g, we study two cases:

Case 1. g is a known positive constant.

Recommendations for g have included the following: Kass & Wasserman's (1995) unit information prior ($g = n$), Foster & George's (1994) risk inflation criterion ($g = p^2$), Fernández, Ley, & Steel's (2001) benchmark prior ($g = \max(n, p^2)$) and so on.

Case 2. g is an unknown parameter, and the prior of g is $\pi_2(g)$.

By integrating out the latent variable g, one can get the conditional prior of $\boldsymbol{\theta}$ given $\tau > 0$,

$$\pi_0(\boldsymbol{\theta} \mid \tau) = \int_0^\infty \left(\frac{\tau}{2g\pi}\right)^{p/2} \times \exp\left\{-\frac{\tau\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta}}{2g}\right\} \pi_2(g)\, \mathrm{d}g, \tag{23}$$

which can be represented as a mixture of g priors.

For Case 2, some priors $\pi_2(g)$ have been previously considered. Here are two examples.

Example 4.1: Inv-Gamma(ν, c), i.e.

$$\pi_2(g) = \frac{c^\nu}{\Gamma(\nu)} g^{-(\nu+1)} \mathrm{e}^{-c/g}. \tag{24}$$

As discussed by Berger and Strawderman (1996), it results in the multivariate t-prior for $\boldsymbol{\theta}$ given $\tau > 0$, namely

$$\pi_0(\boldsymbol{\theta} \mid \tau) \propto \tau^{p/2} \left[1 + \frac{\tau}{2c}\boldsymbol{\theta}'(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{\theta}\right]^{-(p/2+\nu)}. \tag{25}$$

Zellner-Siow (1980) studied the multivariate cauchy prior for $\boldsymbol{\theta}$, which is one special case of (25) with $\nu = 1/2$ and $c = n/2$.

Example 4.2: Robust prior (Bayarri et al., 2012):

$$\begin{aligned} \pi_2(g) &= h_1[h_3(h_2 + p)]^{h_1}(h_2 + g)^{-(h_1+1)} \times 1_{\{g > h_3(h_2+p) - h_2\}} \\ &\propto (h_2 + g)^{-(h_1+1)} 1_{\{g > h_3(h_2+p) - h_2\}}, \end{aligned} \tag{26}$$

where $h_1 > 0$, $h_2 > 0$, and $h_3 \ge h_2/(h_2 + p)$. The prior (26) has its origins in the robust prior introduced by Strawderman (1971), Berger (1980) and Berger (1985). As Bayarri et al. (2012) discussed, the priors proposed by Liang et al. (2008) are particular cases with $h_1 = \frac{1}{2}, h_2 = 1, h_3 = 1/(1 + p)$ (the hyper-g prior) and $h_1 = \frac{1}{2}, h_2 = p, h_3 = \frac{1}{2}$ (the hyper-g/n prior). The prior in Cui and George (2008) has $h_1 = 1, h_2 = 1, h_3 = 1/(1 + p)$.

For the robust prior (26), it is not straightforward to obtain the closed form of the marginal conditional prior for $\boldsymbol{\theta}$ given τ. Alternatively, we attempt to get the boundary of the marginal density of $\boldsymbol{\theta}$ given τ.

Lemma 4.3: *Define*

$$f(u) \equiv \int_0^\infty \frac{1}{v^{r_1}(v+c)^{r_2}} \exp\left(-\frac{u}{2v}\right) \mathrm{d}v, \tag{27}$$

where $r_1 > 1$, $r_2 \geq 0$ and $c > 0$, then there are two positive constants C_1 and C_2, such that

$$\frac{C_1}{u^{r_1-1}(1+u)^{r_2}} \leq f(u) \leq \frac{C_2}{u^{r_1-1}(1+u)^{r_2}},$$

for any $u > 0$.

The proof is given in the Appendix. Applying this lemma to (23), the resulting prior for $\boldsymbol{\theta}$ given τ for robust prior (26) with $h_3 = h_2/(h_2+p)$ has the boundary

$$\begin{aligned}\frac{C_1\tau}{(\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta})^{p/2-1}(1+\tau\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta})^{h_1+1}} &\leq \pi_0(\boldsymbol{\theta} \mid \tau) \\ &\leq \frac{C_2\tau}{(\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta})^{p/2-1}(1+\tau\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta})^{h_1+1}},\end{aligned} \tag{28}$$

where $p > 2$.

4.2. Admissibility

We apply the results in Section 3 to determine when the hierarchical priors (22) result in admissible estimators of $\boldsymbol{\theta}$ under the normalised squared error loss (9).

Theorem 4.4: (*Case* 1) *For the model* (2) *under the hierarchical prior* (22) *with a given g. If $0 \leq k < 1$, the estimator $\boldsymbol{\delta}_B(\boldsymbol{y})$ in* (11) *for $\boldsymbol{\theta}$ is admissible under the normalised squared error loss* (9).

The proof of Theorem 4.4 is similar to the proof of Theorem 4.5 later, thus it is omitted. As discussed by George and Foster (2000), the choice of g effectively controls model selection, with large g typically concentrating the prior on parsimonious models with a few large coefficients, whereas small g tends to concentrate the prior on saturated models with small coefficients. Herein, we consider Case 1 from the perspective of admissibility, not the model selection. From Theorem 4.4, the choice of fix g has no effect on the admissibility of estimators $\boldsymbol{\delta}_B(\boldsymbol{y})$ of $\boldsymbol{\theta}$.

Next, we consider Case 2. The prior density of g satisfies the following conditions:

Condition A1. $\pi_2(g)$ is a continuous function in $(0,\infty)$;
Condition A2. $\exists a \in \mathbb{R}$, $\pi_2(g) = O(g^a)$, as $g \to 0$;
Condition A3. $\exists b \geq 0$, $\pi_2(g) \sim Cg^{-b}$, as $g \to \infty$ for some constant $C > 0$.

Clearly, two examples of $\pi_2(g)$ discussed in Section 4.1 satisfy Condition A1–A3 with appropriate a and b.

Theorem 4.5: (*Case* 2) *For the model* (2) *with the hierarchical prior* (22), *assume $\pi_2(g)$ satisfies Condition* A1–A3. *If $0 \leq k < 1$, $a > k-1$ and $k+b > 3$, the estimator $\boldsymbol{\delta}_B(\boldsymbol{y})$ in* (11) *for $\boldsymbol{\theta}$ is admissible under the normalised squared error loss* (9).

Proof: It is convenient to write $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{H}'\boldsymbol{D}\boldsymbol{H}$, where $\boldsymbol{H}$ is the matrix of eigenvectors corresponding to $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \ldots, d_p)$ with $d_1 \geq \cdots \geq d_p$. Define $\boldsymbol{u} = \boldsymbol{H}\boldsymbol{\theta} = (u_1, \ldots, u_p)'$. From (23), the conditional prior of $\boldsymbol{\theta}$ given $\tau > 0$ is

$$\begin{aligned}\pi_0(\boldsymbol{\theta} \mid \tau) = &\int_0^\infty \left(\frac{\tau}{2g\pi}\right)^{p/2} \\ &\times \exp\left\{-\frac{\tau}{2g}\sum_{i=1}^p \frac{u_i^2}{d_i}\right\} \pi_2(g)\,\mathrm{d}g,\end{aligned}$$

which is a decreasing function of u_i^2, for $i = 1, \ldots, p$. From Theorem 3.7, we just need to verify Condition 1–6. For Condition 1,

$$\begin{aligned}&\int_{S^c}\int_0^\infty \frac{1}{\tau}\frac{\pi_0(\boldsymbol{\theta}|\tau)}{\|\boldsymbol{\theta}\|^2\log(\|\boldsymbol{\theta}\| \vee 2)}\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \\ &= \int_{S^c}\int_0^\infty\int_0^\infty \frac{1}{\tau}\frac{\pi_0(\boldsymbol{\theta} \mid g, \tau)}{\|\boldsymbol{\theta}\|^2\log(\|\boldsymbol{\theta}\| \vee 2)} \\ &\quad \times \pi_1(\tau)\pi_2(g)\,\mathrm{d}\tau\,\mathrm{d}g\,\mathrm{d}\boldsymbol{\theta} \\ &\leq \int_{S^c}\int_0^\infty \frac{g^{-p/2}}{\|\boldsymbol{\theta}\|^2\log(\|\boldsymbol{\theta}\| \vee 2)}\pi_2(g) \\ &\quad \times \left[\int_0^\infty \tau^{p/2-k-1}\exp\left(-\frac{d_p\tau\|\boldsymbol{\theta}\|^2}{2g}\right)\mathrm{d}\tau\right]\mathrm{d}g\,\mathrm{d}\boldsymbol{\theta}.\end{aligned}$$

If $k < p/2$, there is a positive constant C, such that

$$\begin{aligned}&\int_{S^c}\int_0^\infty \frac{1}{\tau}\frac{\pi_0(\boldsymbol{\theta} \mid \tau)}{\|\boldsymbol{\theta}\|^2\log(\|\boldsymbol{\theta}\| \vee 2)}\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \\ &\leq C\int_{S^c}\int_0^\infty \frac{g^{-p/2}}{\|\boldsymbol{\theta}\|^2\log(\|\boldsymbol{\theta}\| \vee 2)} \\ &\quad \times \pi_2(g)\left(\frac{\|\boldsymbol{\theta}\|^2}{g}\right)^{-p/2+k}\mathrm{d}g\,\mathrm{d}\boldsymbol{\theta} \\ &= C\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}\log(\|\boldsymbol{\theta}\| \vee 2)}\mathrm{d}\boldsymbol{\theta} \\ &\quad \times \int_0^\infty g^{-k}\pi_2(g)\,\mathrm{d}g.\end{aligned}$$

By polar coordinate transformation $r = \|\boldsymbol{\theta}\|$, the integration over $\boldsymbol{\theta}$ becomes

$$\begin{aligned}&\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}\log(\|\boldsymbol{\theta}\| \vee 2)}\mathrm{d}\boldsymbol{\theta} \\ &= \int_1^\infty \frac{1}{r^{3-2k}\log(r \vee 2)}\,\mathrm{d}r,\end{aligned}$$

which is finite if $3-2k \geq 1$, i.e. $k \leq 1$. Since $\pi_2(g)$ satisfies Condition A1-A3, there are some postive constants $N_0 < N_1 < N_2$, C_1 and C_2 such that

$$\begin{aligned}\int_0^\infty g^{-k}\pi_2(g)\,\mathrm{d}g &= \int_0^{N_0} g^{-k}\pi_2(g)\,\mathrm{d}g \\ &\quad + \int_{N_0}^{N_1} g^{-k}\pi_2(g)\,\mathrm{d}g \\ &\quad + \int_{N_1}^{\infty} g^{-k}\pi_2(g)\,\mathrm{d}g \\ &\leq C_1 \int_0^{N_0} g^{a-k}\,\mathrm{d}g \\ &\quad + \int_{N_0}^{N_1} g^{-k}\pi_2(g)\,\mathrm{d}g \\ &\quad + C_2 \int_{N_1}^{\infty} g^{-k-b}\,\mathrm{d}g,\end{aligned}$$

which is finite if $a > k-1$, and $k+b > 1$.

For Condition 2,

$$\begin{aligned}\int_{S^c}\int_0^\infty \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} &\leq \int_{S^c}\int_0^\infty g^{-p/2}\pi_2(g) \\ &\quad \times \left[\int_0^\infty \tau^{p/2-k}\exp\left(-\frac{d_p\tau\|\boldsymbol{\theta}\|^2}{2g}\right)\mathrm{d}\tau\right]\mathrm{d}g\,\mathrm{d}\boldsymbol{\theta} \\ &= \frac{\Gamma(1+p/2-k)}{(d_p/2)^{1+p/2-k}}\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}}\mathrm{d}\boldsymbol{\theta} \\ &\quad \times \int_0^\infty g^{1-k}\pi_2(g)\,\mathrm{d}g,\end{aligned}$$

which is finite if $0 \leq k < 1$, $a > k-2$ and $k+b > 2$.

For Condition 3,

$$\begin{aligned}&\int_{S^c}\int_0^\infty \tau\|\boldsymbol{\theta}\|^2\pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \\ &\quad \leq \int_{S^c}\int_0^\infty \|\boldsymbol{\theta}\|^2 g^{-p/2}\pi_2(g) \\ &\qquad \times \left[\int_0^\infty \tau^{1+p/2-k}\exp\left(-\frac{d_p\tau\|\boldsymbol{\theta}\|^2}{2g}\right)\mathrm{d}\tau\right]\mathrm{d}g\,\mathrm{d}\boldsymbol{\theta} \\ &\quad = \frac{\Gamma(2+p/2-k)}{(d_p/2)^{2+p/2-k}}\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}}\mathrm{d}\boldsymbol{\theta} \\ &\qquad \times \int_0^\infty g^{2-k}\pi_2(g)\,\mathrm{d}g,\end{aligned}$$

which is finite if $0 \leq k < 1$, $a > k-3$ and $k+b > 3$.

For Condition 4, note that

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta} \mid \tau) &= -\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta}\tau^{p/2+1}\int_0^\infty g^{-(p/2+1)} \\ &\quad \times \exp\left(-\frac{\tau\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta}}{2g}\right)\pi_2(g)\,\mathrm{d}g.\end{aligned}$$

By Cauchy-Schwarz inequality, it yields

$$\begin{aligned}\|\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta} \mid \tau)\|^2 &\leq d_1^2\|\boldsymbol{\theta}\|^2\tau^{p+2}\int_0^\infty g^{-p/2} \\ &\quad \times \exp\left(-\frac{\tau\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta}}{2g}\right)\pi_2(g)\,\mathrm{d}g \\ &\quad \times \int_0^\infty g^{-(p/2+2)}\exp\left(-\frac{\tau\boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\theta}}{2g}\right) \\ &\quad \times \pi_2(g)\,\mathrm{d}g \\ &\leq d_1^2\|\boldsymbol{\theta}\|^2\tau^{p/2+2}\pi_0(\boldsymbol{\theta} \mid \tau) \\ &\quad \times \int_0^\infty g^{-(p/2+2)}\exp\left(-\frac{d_p\tau\|\boldsymbol{\theta}\|^2}{2g}\right) \\ &\quad \times \pi_2(g)\,\mathrm{d}g.\end{aligned}$$

Therefore,

$$\begin{aligned}&\int_{S^c}\int_0^\infty \frac{1}{\tau}\frac{\|\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta} \mid \tau)\|^2}{\pi_0(\boldsymbol{\theta} \mid \tau)}\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \\ &\quad \leq d_1^2\int_{S^c}\int_0^\infty\int_0^\infty \|\boldsymbol{\theta}\|^2\tau^{p/2+1}g^{-(p/2+2)} \\ &\qquad \times \exp\left(-\frac{d_p\tau\|\boldsymbol{\theta}\|^2}{2g}\right)\pi_2(g)\,\mathrm{d}g\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \\ &\quad = d_1^2\int_{S^c}\int_0^\infty \|\boldsymbol{\theta}\|^2 g^{-(p/2+2)}\pi_2(g) \\ &\qquad \times \left[\int_0^\infty \tau^{p/2+1-k}\exp\left(-\frac{\tau\|\boldsymbol{\theta}\|^2}{2g}\right)\mathrm{d}\tau\right]\mathrm{d}g\,\mathrm{d}\boldsymbol{\theta} \\ &\quad = \frac{\Gamma(2+p/2-k)}{(d_p/2)^{2+p/2-k}}\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}}\mathrm{d}\boldsymbol{\theta} \\ &\qquad \times \int_0^\infty g^{-k}\pi_2(g)\,\mathrm{d}g,\end{aligned}$$

which is finite if $0 \leq k < 1$, $a > k-1$, and $k+b > 1$.

For Condition 5,

$$\begin{aligned}&\int_{\|\boldsymbol{\theta}\|^2<B}\int_{\tau<B}\pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \\ &\quad \leq \int_0^\infty\int_0^B g^{-p/2}\tau^{p/2-k} \\ &\qquad \times \left[\int_{\|\boldsymbol{\theta}\|^2<B}\exp\left(-\frac{d_p\tau\|\boldsymbol{\theta}\|^2}{2g}\right)\mathrm{d}\boldsymbol{\theta}\right]\pi_2(g)\,\mathrm{d}\tau\,\mathrm{d}g.\end{aligned}$$

By polar coordinate transformation $r = \|\boldsymbol{\theta}\|^2$, the integration over $\boldsymbol{\theta}$ becomes

$$\begin{aligned}&\int_{\|\boldsymbol{\theta}\|^2<B}\exp\left(-\frac{d_p\tau\|\boldsymbol{\theta}\|^2}{2g}\right)\mathrm{d}\boldsymbol{\theta} \\ &\quad = \int_0^B r^{p/2-1}\exp\left(-\frac{d_p\tau r}{2g}\right)\mathrm{d}r \\ &\quad \leq C\tau^{-p/2}g^{p/2}.\end{aligned}$$

Therefore,

$$\int_{\|\boldsymbol{\theta}\|^2<B}\int_{\tau<B}\pi_0(\boldsymbol{\theta}\mid\tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \le C\int_0^B \tau^{-k}\,\mathrm{d}\tau\int_0^\infty \pi_2(g)\,\mathrm{d}g,$$

which is finite if $0\le k<1$, $a>-1$, and $b>1$.

Combining these restrictions, we can find that when $0\le k<1$, $a>k-1$ and $k+b>3$, Conditions 1–5 hold. As Han (2009) discussed, Condition 6 is very mild. Proceeding in an analogous way on page 47 of Han (2009), Condition 6 holds. By Theorem 3.7, the estimator $\boldsymbol{\delta}_B(\boldsymbol{y})$ in (11) for $\boldsymbol{\theta}$ is admissible. ■

We are also interested in admissible estimators under Inv-Gamma and robust prior for g. Using Theorem 4.5, we have the following results.

Theorem 4.6: *For the model* (2) *with the hierarchical prior* (22), *assume* $\pi_2(g)$ *is Inv-Gamma*(v,c). *If* $0\le k<1$, *and* $v>2-k$, *the estimator* $\boldsymbol{\delta}_B(\boldsymbol{y})$ *in* (11) *for* $\boldsymbol{\theta}$ *is admissible.*

Proof: By Theorem 4.5 with any constant $a>k-1$ and $b=v+1$, the result holds. ■

Theorem 4.7: *For the model* (2) *with the hierarchical prior* (22), *assume* $\pi_2(g)$ *is robust prior* (26). *If* $0\le k<1$, *and* $h_1>2-k$, *the estimator* $\boldsymbol{\delta}_B(\boldsymbol{y})$ *in* (11) *for* $\boldsymbol{\theta}$ *is admissible.*

Proof: From Theorem 4.5 with $a=0$ and $b=h_1+1$, the proof is completed. ■

5. Admissibility for a 3-level hierarchical model

We also study a 3-level hierarchical model and determine which elements of the hierarchical prior class lead to admissible estimators of the $\boldsymbol{\theta}$ under normalised squared error loss.

5.1. The model and priors

Consider the following 3-level hierarchical model

$$\begin{aligned}
&\textit{Level } 1: (\boldsymbol{y}\mid\boldsymbol{\theta},\tau)\sim N_n(\boldsymbol{X\theta},\tau^{-1}\boldsymbol{I}_n);\\
&\textit{Level } 2: (\boldsymbol{\theta}\mid\boldsymbol{\beta},\tau,g)\sim N_p(\boldsymbol{Z\beta},g\tau^{-1}\boldsymbol{A});\\
&\textit{Level } 3: (\boldsymbol{\beta}\mid\lambda,\tau)\sim N_s(\boldsymbol{0},\lambda\tau^{-1}\boldsymbol{B}),
\end{aligned}\tag{29}$$

where $\boldsymbol{Z}$ is a given $p\times s$ matrix with full rank s, $\boldsymbol{\beta}$ is the $s\times 1$ unknown vector, $\boldsymbol{A}$ and $\boldsymbol{B}$ are a $p\times p$ and $s\times s$ known covariate matrix, respectively, and λ is an unknown hyperparameter. To simplify the computation, without loss of generality, we set $\boldsymbol{A}=\boldsymbol{I}_p$ and $\boldsymbol{B}=\boldsymbol{I}_s$.

Assume $\pi_1(\tau)\propto\tau^{-k}$, and the prior of g satisfies the Condition A1–A3. The prior $\pi_3(\lambda)$ satisfies the following conditions.

Condition B1. $\pi_3(\lambda)$ is a continuous function in $(0,\infty)$;
Condition B2. $\exists c_1$, $\pi_3(\lambda)=O(\lambda^{c_1})$, as $\lambda\to 0$;
Condition B3. $\exists c_2\ge 0$, $\pi_3(\lambda)\sim C\lambda^{-c_2}$ as $\lambda\to\infty$ for some constant $C>0$.

5.2. Admissibility

The following lemma is needed.

Lemma 5.1: *For the 3-level hierarchical model* (29), *assume* $\pi_1(\tau)\propto\tau^{-k}$, $\pi_2(g)$ *satisfies Condition* A1–A3, *and* $\pi_3(\lambda)$ *satisfies Condition* B1–B3. *Then*

$$\pi_0(\boldsymbol{\theta}\mid\tau)\propto\tau^{p/2}\iint\exp\left[-\frac{1}{2}\tau\boldsymbol{\theta}'(g\boldsymbol{I}_p+\lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta}\right]\times(g+\lambda)^{-s/2}g^{-(p-s)/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda.\tag{30}$$

Proof: Note that

$$\begin{aligned}
\pi_0(\boldsymbol{\theta}\mid\tau)&\propto\int_0^\infty\int_0^\infty\int_{\mathbb{R}^s}\pi_0(\boldsymbol{\theta}\mid\boldsymbol{\beta},\tau,g)\,\pi(\boldsymbol{\beta}\mid\lambda)\pi_2(g)\pi_3(\lambda)\,\mathrm{d}\boldsymbol{\beta}\,\mathrm{d}g\,\mathrm{d}\lambda\\
&\propto\tau^{(p+s)/2}\int_0^\infty\int_0^\infty\int_{\mathbb{R}^s}\times\exp\left(-\frac{\tau\|\boldsymbol{\theta}-\boldsymbol{Z\beta}\|^2}{2g}-\frac{\tau\|\boldsymbol{\beta}\|^2}{2\lambda}\right)\times g^{-p/2}\lambda^{-s/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}\boldsymbol{\beta}\,\mathrm{d}g\,\mathrm{d}\lambda.
\end{aligned}$$

Define $\boldsymbol{\beta}_0=g^{-1}(\lambda^{-1}\boldsymbol{I}_s+g^{-1}\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\theta}$, we have

$$\begin{aligned}
\frac{\tau\|\boldsymbol{\theta}-\boldsymbol{Z\beta}\|^2}{g}+\frac{\tau\|\boldsymbol{\beta}\|^2}{\lambda}&=\tau(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\times(\lambda^{-1}\boldsymbol{I}_s+g^{-1}\boldsymbol{Z}'\boldsymbol{Z})\times(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\tau\boldsymbol{\theta}'\times(g\boldsymbol{I}_p+\lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta}.
\end{aligned}$$

Then the marginal distribution of θ given τ,

$$\pi_0(\boldsymbol{\theta}\mid\tau)\propto\tau^{p/2}\int_0^\infty\int_0^\infty|\lambda^{-1}\boldsymbol{I}_s+g^{-1}\boldsymbol{Z}'\boldsymbol{Z}|^{-1/2}\times\exp\left[-\frac{1}{2}\tau\boldsymbol{\theta}'(g\boldsymbol{I}_p+\lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta}\right]\times g^{-p/2}\lambda^{-s/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda,$$

which is proportional to (30). The proof is completed. ■

Theorem 5.2: *For the 3-level hierarchical model* (29), *assume* $\pi_1(\tau)\propto\tau^{-k}$, $\pi_2(g)$ *satisfies Condition* A1–A3,

and $\pi_3(\lambda)$ satisfies Condition B1–B3. *Then, the estimator $\boldsymbol{\delta}_B(\boldsymbol{y})$ in* (11) *for $\boldsymbol{\theta}$ is admissible if $0 \le k < 1$, $b > 3-k$, $c_2 > 3+(p-s)/2-k$, and one of the following conditions holds,*

(i) $p > s, a > (p-s)/2+1, c_1 > -1$;
(ii) $p = s, a > -1, c_1 > -1+k$;
(iii) $p = s, a > -1+k, c_1 > -1$.

Proof: It is convenient to write $\boldsymbol{ZZ}' = \boldsymbol{\Gamma}'\boldsymbol{D}\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is the matrix of eigenvectors corresponding to $\boldsymbol{D} = \mathrm{diag}(z_1, z_2, \ldots, z_p)$ with $z_1 \ge \cdots \ge z_p$. Herein, we denote $\boldsymbol{u} = \boldsymbol{\Gamma}\boldsymbol{\theta} = (u_1, \ldots, u_p)'$. Therefore, from Lemma 5.1,

$$\pi_0(\boldsymbol{\theta} \mid \tau) \propto \tau^{p/2} \iint \exp\left[-\frac{\tau}{2}\sum_{i=1}^{p}(g+z_i\lambda)^{-1}u_i^2\right] \times (g+\lambda)^{-s/2} g^{-(p-s)2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda,$$

which is a decreasing function of u_i^2, for $i = 1, \ldots, p$. In addition, there are two positive constant C_1 and C_2, such that

$$\pi_0(\boldsymbol{\theta} \mid \tau) \le C_1\tau^{p/2} \iint \exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right) \times (g+\lambda)^{-s/2} g^{-(p-s)/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda.$$

For the technical reasons, we first consider Condition 2. Note that

$$\begin{aligned}\int_{S^c}\int_0^\infty \pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} &\le C_1 \int_{S^c}\int_0^\infty\int_0^\infty \times \left\{\int_0^\infty \tau^{p/2-k}\exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right)\mathrm{d}\tau\right\} \\ &\quad \times (g+\lambda)^{-s/2}g^{-(p-s)/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}g\,\mathrm{d}\boldsymbol{\theta} \\ &= C_1\frac{\Gamma(1+p/2-k)}{C_2^{1+p/2-k}}\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}}\mathrm{d}\boldsymbol{\theta} \\ &\quad \times \iint (\lambda+g)^{1+(p-s)/2-k}g^{-(p-s)/2} \\ &\quad \times \pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda.\end{aligned}$$

The integration over $\boldsymbol{\theta}$ is finite if $k < 1$. For simplicity, denote $l = 1+(p-s)/2-k$ and $h = (p-s)/2$. If $0 \le k < 1$, we have $l > 0$.

Note that

$$\begin{aligned}&\int_0^\infty\int_0^\infty (\lambda+g)^{1+(p-s)/2-k}g^{-(p-s)/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda \\ &= \left\{\int_0^1\int_0^1 + \int_0^1\int_0^\infty + \int_1^\infty\int_0^1 + \int_1^\infty\int_1^\infty\right\} \times (\lambda+g)^l g^{-h}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda \\ &\equiv I_1 + I_2 + I_3 + I_4. \qquad (31)\end{aligned}$$

Clearly,

$$I_1 \le 2^l\int_0^1 g^{-h}\pi_2(g)\,\mathrm{d}g \times \int_0^1 \pi_3(\lambda)\,\mathrm{d}\lambda,$$

which is finite if $a > h-1$ and $c_1 > -1$. Clearly,

$$I_4 \le 2^l\int_1^\infty g^{l-h}\pi_2(g)\,\mathrm{d}g \times \int_1^\infty \lambda^l\pi_3(\lambda)\,\mathrm{d}\lambda,$$

which is finite if $b > 1-h+l$ and $c_2 > 1+l$. Similarly, it is easy to verify that $I_2 + I_3$ is finite if $a > h-1$, $b > 1-h+l$, $c_1 > -1$ and $c_2 > 1+l$. Therefore, (31) is finite if $a > (p-s)/2-1, b > 2-k, c_1 > -1$ and $c_2 > 2+(p-s)/2-k$.

Similarly, for Condition 3,

$$\begin{aligned}&\int_{S^c}\int_0^\infty \tau\|\boldsymbol{\theta}\|^2\pi_0(\boldsymbol{\theta} \mid \tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta} \\ &\le C_1'\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}}\mathrm{d}\boldsymbol{\theta} \\ &\quad \times \iint (\lambda+g)^{2+(p-s)/2-k}g^{-(p-s)/2} \\ &\quad \times \pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda, \qquad (32)\end{aligned}$$

where C_1' is a postive constant.

Clearly, $2+(p-s)/2-k > 0$ if $0 \le k < 1$. As in the proof of Condition 2, (32) is finite if $0 \le k < 1, a > (p-s)/2-1, b > 3-k, c_1 > -1$ and $c_2 > 3+(p-s)/2-k$.

For Condition 4, from Lemma 5.1, note that

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta} \mid \tau) = -\tau^{p/2+1}\iint &(g\boldsymbol{I}_p + \lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta} \\ &\times \exp\left[-\frac{1}{2}\tau\boldsymbol{\theta}'(g\boldsymbol{I}_p+\lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta}\right] \\ &\times (g+\lambda)^{-s/2}g^{-(p-s)/2}\pi_3(\lambda) \\ &\times \pi_2(g)\,\mathrm{d}g\,\mathrm{d}\lambda.\end{aligned}$$

We will consider two cases, i.e. $p > s$ and $p = s$, respectively. If $p > s$, $\|(g\boldsymbol{I}_p + \lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta}\| \le g^{-1}\|\boldsymbol{\theta}\|$. It yields

$$\begin{aligned}\|\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta} \mid \tau)\|^2 &\le \|\boldsymbol{\theta}\|^2\tau^{p+2} \\ &\quad \times \left(\iint \exp\left[-\frac{1}{2}\tau\boldsymbol{\theta}'(g\boldsymbol{I}_p+\lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta}\right]\right. \\ &\quad \left.\times (g+\lambda)^{-s/2}g^{-(p-s)/2-1}\pi_3(\lambda)\pi_2(g)\,\mathrm{d}g\,\mathrm{d}\lambda\right)^2 \\ &\le \|\boldsymbol{\theta}\|^2\tau^{p/2+2}\pi_0(\boldsymbol{\theta} \mid \tau) \\ &\quad \times \iint \exp\left[-\frac{1}{2}\tau\boldsymbol{\theta}'(g\boldsymbol{I}_p+\lambda\boldsymbol{ZZ}')^{-1}\boldsymbol{\theta}\right] \\ &\quad \times (g+\lambda)^{-s/2}g^{-(p-s)/2-2} \\ &\quad \times \pi_3(\lambda)\pi_2(g)\,\mathrm{d}g\,\mathrm{d}\lambda \\ &\le C_1|\boldsymbol{\theta}\|^2\tau^{p/2+2}\pi_0(\boldsymbol{\theta} \mid \tau)\iint \exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right) \\ &\quad \times (g+\lambda)^{-s/2}g^{-(p-s)/2-2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda.\end{aligned}$$

In the second step, we apply the Cauchy-Schwartz inequality. Therefore,

$$\begin{aligned}
&\int_{S^c}\int_0^\infty \frac{1}{\tau}\frac{\|\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta}\mid\tau)\|^2}{\pi_0(\boldsymbol{\theta}\mid\tau)}\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta}\\
&\quad\le C_1\int_{S^c}\iint \|\boldsymbol{\theta}\|^2\left\{\int_0^\infty \tau^{1+p/2-k}\right.\\
&\qquad\left.\times\exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right)\mathrm{d}\tau\right\}(g+\lambda)^{-s/2}g^{-(p-s)/2-2}\\
&\qquad\times\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda\,\mathrm{d}\boldsymbol{\theta}\\
&\quad= C_1'\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}}\mathrm{d}\boldsymbol{\theta}\times\iint(\lambda+g)^{(p-s)/2+2-k}\\
&\qquad\times g^{-(p-s)/2-2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda. \qquad (33)
\end{aligned}$$

As in the proof of Condition 3, (33) is finite if $0\le k<1$, $a>(p-s)/2+1$, $b>1-k$, $c_1>-1$ and $c_2>3+(p-s)/2-k$.

If $p=s$, there is a positive constant C_3, such that

$$\|(g\boldsymbol{I}_p+\lambda\boldsymbol{Z}\boldsymbol{Z}')^{-1}\boldsymbol{\theta}\|\le C_3(g+\lambda)^{-1}\|\boldsymbol{\theta}\|.$$

Therefore, using the Cauchy-Schwartz inequality,

$$\begin{aligned}
\|\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta}\mid\tau)\|^2 &\le C_4\|\boldsymbol{\theta}\|^2\tau^{p/2+2}\pi_0(\boldsymbol{\theta}\mid\tau)\\
&\quad\times\iint\exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right)\\
&\quad\times(g+\lambda)^{-p/2-2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda,
\end{aligned}$$

where $C_4=C_1C_3^2$. Thus,

$$\begin{aligned}
&\int_{S^c}\int_0^\infty\frac{1}{\tau}\frac{\|\nabla_{\boldsymbol{\theta}}\pi_0(\boldsymbol{\theta}\mid\tau)\|^2}{\pi_0(\boldsymbol{\theta}\mid\tau)}\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta}\\
&\quad\le C_4\int_{S^c}\iint\|\boldsymbol{\theta}\|^2\\
&\qquad\times\left\{\int_0^\infty\tau^{1+p/2-k}\exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right)\mathrm{d}\tau\right\}\\
&\qquad\times(g+\lambda)^{-p/2-2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda\,\mathrm{d}\boldsymbol{\theta}\\
&\quad= C_4'\int_{S^c}\frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}}\mathrm{d}\boldsymbol{\theta}\\
&\qquad\times\iint(\lambda+g)^{-k}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda, \qquad (34)
\end{aligned}$$

where C_4' is a positive constant. Note that the integration over $\boldsymbol{\theta}$ is finite if $0\le k<1$.

If $k\ge 0$,

$$\begin{aligned}
&\int_0^\infty\int_0^\infty(\lambda+g)^{-k}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda\\
&\quad\le\int_0^\infty\pi_2(g)\,\mathrm{d}g\int_0^\infty\lambda^{-k}\pi_3(\lambda)\,\mathrm{d}\lambda,
\end{aligned}$$

which is finite if $a>-1, b>1, c_1>-1+k$ and $c_2>1-k$. Meanwhile,

$$\begin{aligned}
&\int_0^\infty\int_0^\infty(\lambda+g)^{-k}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda\\
&\quad\le\int_0^\infty g^{-k}\pi_2(g)\,\mathrm{d}g\int_0^\infty\pi_3(\lambda)\,\mathrm{d}\lambda,
\end{aligned}$$

which is finite if $a>-1+k, b>1-k, c_1>-1$ and $c_2>1$.

For Condition 5, note that

$$\begin{aligned}
&\int_{\|\boldsymbol{\theta}\|^2<B}\int_{\tau<B}\pi_0(\boldsymbol{\theta}\mid\tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta}\\
&\quad\le C_1\int_0^B\iint\tau^{p/2-k}\\
&\qquad\times\left\{\int_{\|\boldsymbol{\theta}\|^2<B}\exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right)\mathrm{d}\boldsymbol{\theta}\right\}\\
&\qquad\times(g+\lambda)^{-s/2}g^{-(p-s)/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}g\,\mathrm{d}\tau.
\end{aligned}$$

By polar coordinate transformation $r=\|\boldsymbol{\theta}\|^2$, the integration over $\boldsymbol{\theta}$ becomes

$$\begin{aligned}
&\int_{\|\boldsymbol{\theta}\|^2<B}\exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right)\mathrm{d}\boldsymbol{\theta}\\
&\quad=\int_0^B r^{p/2-1}\exp\left(-C_2\frac{\tau r}{g+\lambda}\right)\mathrm{d}\\
&\quad\le C_5\tau^{-p/2}(g+\lambda)^{p/2},
\end{aligned}$$

where C_5 is a positive constant. Therefore,

$$\begin{aligned}
&\int_{\|\boldsymbol{\theta}\|^2<B}\int_{\tau<B}\pi_0(\boldsymbol{\theta}\mid\tau)\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta}\\
&\quad\le C_5\int_0^B\tau^{-k}\,\mathrm{d}\tau\times\int_0^\infty\int_0^\infty(\lambda+g)^{(p-s)/2}\\
&\qquad\times g^{-(p-s)/2}\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda, \qquad (35)
\end{aligned}$$

which is finite if $0\le k<1$, $a>(p-s)/2-1, b>1, c_1>-1$ and $c_2>1+(p-s)/2$.

Combining the above results, Conditions 2–5 hold if (k,a,b,c_1,c_2) satisfy the conditions as this theorem states. For Condition 1,

$$\begin{aligned}
&\int_{S^c}\int_0^\infty\frac{1}{\tau}\frac{\pi_0(\boldsymbol{\theta}\mid\tau)}{\|\boldsymbol{\theta}\|^2\log(\|\boldsymbol{\theta}\|\vee 2)}\pi_1(\tau)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{\theta}\\
&\quad\le C_1\int_{S^c}\iint\frac{1}{\|\boldsymbol{\theta}\|^2\log(\|\boldsymbol{\theta}\|\vee 2)}\\
&\qquad\times\left\{\int_0^\infty\tau^{p/2-k-1}\exp\left(-C_2\frac{\tau\|\boldsymbol{\theta}\|^2}{g+\lambda}\right)\mathrm{d}\tau\right\}\\
&\qquad\times(g+\lambda)^{-s/2}g^{-(p-s)2}\\
&\qquad\times\pi_2(g)\pi_3(\lambda)\,\mathrm{d}g\,\mathrm{d}\lambda\,\mathrm{d}\boldsymbol{\theta}
\end{aligned}$$

$$\leq C_6 \int_{S^c} \frac{1}{\|\boldsymbol{\theta}\|^{2+p-2k}} \mathrm{d}\boldsymbol{\theta} \times \iint (\lambda + g)^{(p-s)/2-k} g^{-(p-s)/2} \times \pi_2(g)\pi_3(\lambda)\, \mathrm{d}g\, \mathrm{d}\lambda, \tag{36}$$

where C_6 is a positive constant. If $p = s$, (36) can be proceeded as (34). If $p > s$, it is easy to verify that (36) is finite if $0 \leq k < 1$, $a > (p-s)/2 + 1$, $b > 3 - k$, $c_2 > 3 + (p-s)/2 - k$, $c_1 > -1$. Proceeding in an analogous way on page 47 of Han (2009), Condition 6 also holds. By Theorem 3.7, estimator (11) for $\boldsymbol{\theta}$ are admissible. ■

6. Comments

In Section 2, we listed the sufficient conditions for admissibility of the estimators of $\boldsymbol{\theta}$ with unknown τ, which was developed by Han (2009). In Section 3, we generalise the sufficient conditions for admissibility and apply these results to the normal linear regression model (2). We have to admit that those sufficient conditions are still not optimal enough. Sometimes, we can't obtain satisfactory results utilising the conditions directly. In our paper, we consider $\pi_1(\tau) \propto \tau^{-k}$ for the prior of τ. The condition of k for admissibility is $0 \leq k < 1$. Unfortunately, we can't prove the admissibility for the boundary point $k = 1$, which is of great interest since it is the natural extension of Stein's harmonic prior (Stein, 1981) to the unknown variance problem. In follow-up work, we will try to explore the more powerful sufficient conditions for admissibility of the estimators of $\boldsymbol{\theta}$ with unknown τ. One promising method for this problem may be by Blyth's method (Blyth, 1951), discovering an appropriate sequence of finite measures.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The project was supported by the 111 Project of China (No. B14019), the National Natural Science Foundation of China [Grant No. 11671146].

Notes on contributors

Chengyuan Song is a PhD candidate in the College of Statistics, East China Normal University, Shanghai, China. His research interests include Bayesian statistics.

Dr. Dongchu Sun received his PhD. in 1991 from Department of Statistics, Purdue University, under the guidence of Professor James O. Berger.

References

Bayarri, M. J., Berger, J. O., Forte, A., & Garcła-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, *40*, 1550–1577.

Berger, J. O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *The Annals of Statistics*, *8*, 716–761.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer-Verlag Inc.

Berger, J. O., Pericchi, L. R., & Varshavsky, J. A. (1998). Bayes factors and marginal distributions in invariant situations. *Sankhya, Series A, Indian Journal of Statistics*, *60*, 307–321.

Berger, J. O., & Strawderman, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *The Annals of Statistics*, *24*, 931–951.

Berger, J. O., Strawderman, W., & Tang, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *The Annals of Statistics*, *33*, 606–646.

Berger, J. O., Sun, D., & Song, C. (2018). An objective prior for hyperparameters in normal hierarchical models. Submitted.

Blyth, C. (1951). On minimax statistical decision procedures and their admissibility. *The Annals of Mathematical Statistics*, *22*, 22–42.

Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *The Annals of Mathematical Statistics*, *42*, 855–903.

Cui, W., & George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, *138*, 888–900.

Eaton, M. L. (1989). Group invariance applications in statistics. *Institute of Mathematical Statistics*, *1*, 1–133.

Fernández, C., Ley, E., & Steel, M. F. J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*, 381–427.

Foster, D. P., & George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, *22*, 1947–1975.

Fraisse, A., Raoult, J., Robert, C., & Roy, M. (1990). Une condition nécessaire d'admissibilité et ses conséquences sur les estimateurs à rétrécisseur de la moyenne d'un vecteur normal. *Canadian Journal of Statistics*, *18*, 213–220.

George, E. I., & Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, *87*, 731–747.

Han, X. (2009). *Topics in shrinkage estimation and in causal inference* (PhD thesis). Warton School, University of Pennsylvania.

James, W., & Stein, C. (1961). *Estimation with quadratic loss.Proceedings Fourth Berkeley Symposium in Mathematical Statistics and Probability* (Vol. 1, pp. 361–3679).

Judge, G., Yancey, T., & Bock, M. (1983). Pre-test estimation under squared error loss. *Economics Letters*, *11*, 347–352.

Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, *90*, 928–934.

Katz, V. J. (2005). The history of Stokes' theorem. *Mathematics Magazine*, *52*, 146–156.

Liang, F., Paulo, R., Molina, G., Clyde, M., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, *103*, 410–423.

Maruyama, Y., & Strawderman, W. (2005). A new class of generalized Bayes minimax ridge regression estimators. *The Annals of Statistics*, *33*, 1753–1770.

Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. New York: Springer.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, *9*, 1135–1151.

Strawderman, W. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *The Annals of Mathematical Statistics*, *42*, 385–388.

Strawderman, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *The Annals of Statistics*, *1*, 1189–1194.

Willing, R., & Zhou, G. (2008). Generalized Bayes minimax estimators of the mean of multivariate normal distribution with unknown variance. *Journal of Multivariate Analysis*, *99*, 2208–2220.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Goel, P. and Zellner, A., Eds., *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti* (pp. 233–243). New York: Elsevier Science Publishers, Inc.

Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 585–603). Valencia: University Press.

Appendix

A.1 Proof of Lemma 4.3

To simplify the computation, without loss of generality, we set $c = 1$. Let $x = u/v$, then

$$f(u) = \int_0^\infty \frac{1}{(u/x)^a (u/x+1)^b} \frac{u}{x^2} \exp(-x/2)\,\mathrm{d}x$$
$$= u^{1-a} \int_0^\infty \frac{x^{a+b-2}}{(u+x)^b} \exp(-x/2)\,\mathrm{d}x.$$

One just needs to consider

$$\frac{C_1}{(1+u)^b} \le \int_0^\infty \frac{x^{a+b-2}}{(u+x)^b} \exp(-x/2)\,\mathrm{d}x \le \frac{C_2}{(1+u)^b}. \tag{A1}$$

The integral can be written by

$$\left\{ \int_0^1 + \int_1^\infty \right\} \frac{x^{a+b-2}}{(u+x)^b} \exp(-x/2)\,\mathrm{d}x \equiv I_1 + I_2.$$

For the lower bound of I_1 and I_2, we have

$$I_1 \ge \int_0^1 \frac{x^{a+b-2}}{(u+1)^b} \exp(-x/2)\,\mathrm{d}x = C_1^* \frac{1}{(1+u)^b},$$
$$I_2 \ge \int_1^\infty \frac{x^{a-2}}{(u+1)^b} \exp(-x/2)\,\mathrm{d}x = C_2^* \frac{1}{(1+u)^b},$$

where $C_1^* = \int_0^1 x^{a+b-2} \exp(-x/2)\,\mathrm{d}x$ and $C_2^* = \int_1^\infty x^{a-2} \exp(-x/2)\,\mathrm{d}x$. For the upper bound of I_1 and I_2, we have

$$I_1 \le \int_0^1 \frac{x^{a+b-2}}{(u+1)^b x^b} \exp(-x/2)\,\mathrm{d}x = C_3^* \frac{1}{(1+u)^b},$$
$$I_2 \le \int_1^\infty \frac{x^{a+b-2}}{(u+1)^b} \exp(-x/2)\,\mathrm{d}x = C_4^* \frac{1}{(1+u)^b},$$

where $C_3^* = \int_0^1 x^{a-2} \exp(-x/2)\,\mathrm{d}x$ and $C_4^* = \int_1^\infty x^{a+b-2} \exp(-x/2)\,\mathrm{d}x$. Therefore, let $C_1 = C_1^* + C_2^*$ and $C_2 = C_3^* + C_4^*$. We get (A1). The lemma is proved.