

ON EMPIRICAL BAYES ESTIMATION OF MULTIVARIATE REGRESSION COEFFICIENT

R. J. KARUNAMUNI AND L. WEI

Received 11 November 2005; Revised 19 April 2006; Accepted 4 May 2006

We investigate the empirical Bayes estimation problem of multivariate regression coefficients under squared error loss function. In particular, we consider the regression model $Y = X\beta + \varepsilon$, where Y is an m -vector of observations, X is a known $m \times k$ matrix, β is an unknown k -vector, and ε is an m -vector of unobservable random variables. The problem is squared error loss estimation of β based on some “previous” data Y_1, \dots, Y_n as well as the “current” data vector Y when β is distributed according to some unknown distribution G , where Y_i satisfies $Y_i = X\beta_i + \varepsilon_i$, $i = 1, \dots, n$. We construct a new empirical Bayes estimator of β when $\varepsilon_i \sim N(0, \sigma^2 I_m)$, $i = 1, \dots, n$. The performance of the proposed empirical Bayes estimator is measured using the mean squared error. The rates of convergence of the mean squared error are obtained.

Copyright © 2006 Hindawi Publishing Corporation. All rights reserved.

1. Introduction

In an EB problem (Robbins [15, 16]), there is a sequence of independent random vectors $(Y_1, \beta_1), (Y_2, \beta_2), \dots, (Y_{n+1}, \beta_{n+1})$, where the Y_j 's are observable whereas the β_j 's are not observable. Furthermore, the β_j 's are independent identically distributed (i.i.d.) according to some unknown distribution G and, given $\beta_j = \beta$, Y_j is distributed according to some conditional distribution with density $f(y | \beta)$. In this paper, we will assume that Y_j and β_j are related by a multiple linear regression model given by

$$Y_j = X\beta_j + \varepsilon_j, \quad (1.1)$$

$j = 1, 2, \dots, n + 1$, where Y_j is an m -vector of observations, X is a known $m \times k$ matrix, β_j is a k -vector, and ε_j is an m -vector of unobservable random variables. We will further assume that the conditional distribution of ε_j , given $\beta_j = \beta$, is a multivariate normal distribution $N(0, \sigma^2 I_m)$, where σ^2 is an unknown constant and I_m denotes the identity matrix of order $m \times m$. The objective is to estimate β_{n+1} based on the observations (Y_1, \dots, Y_{n+1})

2 Empirical Bayes estimation

under a squared error loss function given by

$$L(d, \beta) = \|d - \beta\|^2, \quad (1.2)$$

where $\|a\|^2 = a'a = \sum_{i=1}^m a_i^2$ for any m -vector $a = (a_1, \dots, a_m)'$. Denote $\beta_{n+1} = \beta$ and $Y_{n+1} = y$. Let $\hat{\beta}_n(y)$ denote an empirical Bayes (EB) estimator of β under the loss (1.2) based on (Y_1, \dots, Y_n) as well as $Y_{n+1} = y$. Then the sequence $\{\hat{\beta}_n\}$ is said to be asymptotically optimal if the “regret” $EL(\hat{\beta}_n, \beta) - EL(\beta_G, \beta) \rightarrow 0$ as $n \rightarrow \infty$, where expectation E is with respect to all the random variables involved, and β_G denotes the Bayes estimator of β with respect to G under the loss function (1.2).

Martz and Krutchkoff [8] have considered the EB estimation problem in the general linear model of (1.1) above. They have, however, only studied their EB estimator of β through some Monte Carlo simulations and have shown that for certain priors their EB estimator performed better than the usual least squares estimator. Wind [21] has considered EB estimation of β when the error vector ε is assumed to have 0 mean and covariance $\sigma^2 I_m$, but is not assumed to take a specific parametric form, for example, normal. For priors specifying their means and variances, he has exhibited restricted asymptotically optimal EB estimators of β . Singh [17] extended their work and exhibited two classes of estimators $\hat{\phi}$ and $\tilde{\phi}$ for β , one for the case when nothing is known about the support of the prior, and the other for the case when it is known that the prior distribution has a compact support. He showed that $\hat{\phi}$ is asymptotically optimal (a.o.) with rates $O(n^{-1+\eta})$ of the corresponding regret uniformly over a class of all priors satisfying certain moment conditions dependent on η , whereas $\tilde{\phi}$ is shown to be a.o. with rates $O(n^{-1+\eta})$ uniformly over the class of all priors with compact support. Singh’s [17] results are for the case of known error variance σ^2 of ε . Recently, Zhang and Wei [22], Wei and Zhang [20], and Wei [19] extended Singh’s work to the case of unknown error variance of ε . They discussed asymptotic optimality and convergence rates of their EB estimators of the regression coefficient β .

All of the works mentioned above have employed the regret as a measure of goodness in order to study the performance of their proposed EB estimators. In this paper, we study the performance of our EB estimator using the “mean squared error” (MSE) criterion. Our rationale to study the MSE of our EB estimator is motivated by the work of Pensky [11], where she has argued that MSE is a more appropriate criterion as far as applications are concerned. We will construct an EB estimator of β (for the unknown σ^2 case) based on some improved estimators of multivariate normal mixture density and its first partial derivatives. In the work of Zhang and Wei [22], Wei and Zhang [20], and Wei [19], they have used kernel-type estimators of the preceding functions. We show that the rate of convergence of the MSE of the proposed EB estimator is of the order $O(n^{-1}(\log n)^{m+1})$, showing a considerable improvement over the MSEs of the estimators presented in the above papers.

This paper is organized as follows. Section 2 contains the Bayes estimator of β under the loss function (1.2). Section 3 presents a new method of density estimation of a multivariate normal mixture and its first partial derivatives. Section 4 contains the proposed

EB estimator and the main result. Two applications are discussed in Section 5. Section 6 gives some concluding remarks. The proofs are deferred to the appendix.

2. The Bayes estimator of regression coefficient

In this section, we derive the Bayes estimator of β under the loss function (1.2) for the model (1.1). First, we will assume that $k < m$ and that $X'X$ is nonsingular throughout, where X is as defined in (1.1). Further, assume that the error variance σ^2 is bounded away from zero and finite, that is, $0 < \gamma_0 \leq \sigma^2 \leq \gamma_1 < \infty$ for some known constants γ_0 and γ_1 . We assume further that the prior distribution G of β is a member of the family given by

$$\mathcal{F}_{\omega_0} = \left\{ G : \int_{\Omega} \|\beta\|^{\omega_0} dG(\beta) < \infty \right\}, \quad (2.1)$$

where $\omega_0 \geq 2$. Then the conditional density of Y given β is

$$f(y | \beta) = (2\pi\sigma^2)^{-m/2} \exp \left\{ \frac{-\|y - X\beta\|^2}{2\sigma^2} \right\}, \quad (2.2)$$

and the marginal density of Y with respect to G is given by

$$f(y) = \int_{\Omega} f(y | \beta) dG(\beta). \quad (2.3)$$

By (2.2) and [19, Lemma A.1], we obtain that

$$\begin{aligned} \sigma^2 \nabla f(y) &= - \int_{\mathcal{R}} (y - X\beta) f(y | \beta) dG(\beta) \\ &= -yf(y) + X \int_{\Omega} \beta f(y | \beta) dG(\beta), \end{aligned} \quad (2.4)$$

where $\nabla f(y)$ is the gradient of f at y and is defined by

$$\nabla f(y) = \left(\frac{\partial f(y)}{\partial y_1}, \dots, \frac{\partial f(y)}{\partial y_m} \right)'. \quad (2.5)$$

Since $(X'X)$ is invertible by our assumption, from (2.5) we obtain

$$\int_{\Omega} \beta f(y | \beta) dG(\beta) = (X'X)^{-1} X' [yf(y) + \sigma^2 \nabla f(y)]. \quad (2.6)$$

Under the loss function (1.2), the Bayes estimator of β is the posterior expected loss, that is,

$$\begin{aligned} \delta_G(y) &= E(\beta | Y = y) \\ &= \frac{\int \beta f(y | \beta) dG(\beta)}{f(y)} \\ &= (X'X)^{-1} X' [y + \sigma^2 \psi(y)], \end{aligned} \quad (2.7)$$

4 Empirical Bayes estimation

where

$$\psi(y) = \frac{\nabla f(y)}{f(y)} = (\psi_1(y), \dots, \psi_m(y))' \quad (2.8)$$

with $\psi_i(y) = (\partial f(y)/\partial y_i)/f(y)$, $i = 1, \dots, m$.

In view of (2.7) and (2.8), an estimator of $\delta_G(y)$ can be constructed by developing estimators of $f(y)$ and $\psi(y)$ given by (2.3) and (2.8), respectively, based on a sequence of observations $\{Y_i\}_{i=1}^n$ from $f(y)$. This density estimation problem is discussed in the next section.

3. Estimation of a multivariate normal mixture density and its first partial derivatives

Let $\phi(t)$ denote the characteristic function of $f(y)$ given by (2.3) above. Then, it is easy to show that

$$\begin{aligned} \phi(t) &= \int e^{it'y} f(y) dy \\ &= \frac{1}{(2\pi\sigma^2)^{m/2}} \int_{\mathbb{R}^m} \int_{\mathbb{R}} \exp\left\{-\frac{\|y - X\beta\|^2}{2\sigma^2}\right\} e^{it'y} dG(\beta) dy \\ &= e^{-\sigma^2\|t\|^2/2} \phi_G(X't), \end{aligned} \quad (3.1)$$

where ϕ_G denotes the characteristic function of the prior distribution G . Since $\int |\phi(t)| dt < \infty$, by Fourier inversion theorem, we have

$$f(y) = \frac{1}{(2\pi)^m} \int e^{-it'y} \phi(t) dt, \quad (3.2)$$

here and in what follows, all integrals without limits are taken to be over \mathbb{R}^m , the m -dimensional Euclidean space. It is easy to show from (3.1) and (3.2) that for any $0 < M < \infty$,

$$\left| (2\pi)^m f(y) - \int_{-M}^M \cdots \int_{-M}^M e^{-it'y} \phi(t) dt \right| \leq \frac{c}{(\sqrt{\gamma_0})^{m+1}} \frac{1}{M^m} e^{-m(M^2\gamma_0)/2}, \quad (3.3)$$

where $c > 0$ is a constant, see the appendix. Since, for large M , the right-hand side of the inequality (3.3) is small, one can consider estimating $\int_{-M}^M \cdots \int_{-M}^M e^{-it'y} \phi(t) dt$ with large M in order to estimate $f(y)$ given by (3.2). We define

$$\tilde{f}_M(y) = \frac{1}{(2\pi)^m} \int_{[-M, M]^m} e^{-it'y} \hat{\phi}_n(t) dt, \quad (3.4)$$

where $[-M, M]^m$ denotes the m -dimensional interval $[-M, M] \times \cdots \times [-M, M]$ and $\hat{\phi}_n(t) = n^{-1} \sum_{j=1}^n e^{it'Y_j}$, the empirical characteristic function of an i.i.d. sample Y_1, \dots, Y_n from $f(y)$ given by the formula (3.2). Then the $\tilde{f}_M(y)$ is an unbiased estimator of $(1/(2\pi)^m) \int_{[-M, M]^m} e^{-it'y} \phi(t) dt$. With an appropriate choice of M , as a function of n , $\tilde{f}_M(y)$ can be used to estimate density $f(y)$ given by (3.2). A similar estimator has been studied

by Singh and Pensky [18]. This estimator provides optimal convergence rates, however, its common drawback is that it has heavy tails (it is not integrable since its Fourier transform is discontinuous). It would be therefore advantageous to use a smoother estimator that will have the same convergence rates and will decrease faster at infinity. Thus, we define the following modified estimator as our proposed estimator of $f(y)$:

$$\hat{f}_M(y) = \frac{1}{(2\pi)^m} \int g_M(\mathbf{t}) e^{-it'y} \hat{\phi}_n(t) dt, \quad (3.5)$$

where $g_M(\mathbf{t}) = \prod_{i=1}^m g_M(t_i)$ with $g_M(t_i)$ is given by

$$g_M(t_i) = \begin{cases} 1 & \text{if } -M + a \leq t_i \leq M - a, \\ -\frac{1}{a}(t_i - M) & \text{if } M - a \leq t_i \leq M, \\ \frac{1}{a}(t_i + M) & \text{if } -M \leq t_i \leq -M + a, \\ 0 & \text{otherwise,} \end{cases} \quad (3.6)$$

where $a = M - M^{-1}e^{-M^2\gamma_0/2}$, and M will be chosen so that a satisfies $0 < a < M$. Note that $g_M(t_i)$ is bounded and continuous on $(-\infty, \infty)$, and vanishes outside $[-M, M]$. Further, the Fourier transform of $\hat{f}_M(y)$ is continuous on $(-\infty, \infty)$. It can be shown that the mean squared error of $\hat{f}_M(y)$ satisfies the following inequality:

$$E(\hat{f}_M(y) - f(y))^2 \leq c_1(M^{-1}e^{-M^2\gamma_0/2})^{2m} + c_2n^{-1}M^{2m}, \quad (3.7)$$

where c_1 and c_2 are the positive constants independent of n , y , and M (see the appendix). A similar bound holds for the mean squared error of $\tilde{f}_M(y)$. By choosing $M = ((\gamma_0 m)^{-1} \log n)^{1/2}$, we now obtain from (3.7) that

$$E(\hat{f}_M(y) - f(y))^2 \leq c_3n^{-1}(\log n)^m, \quad (3.8)$$

for some constant $c_3 > 0$ independent of n and y .

For estimating $\nabla f(y) = (\partial f/\partial y_1, \dots, \partial f/\partial y_m)'$, the gradient of f at y , observe that since $\int |t\phi(t)| dt < \infty$, from (3.2),

$$\frac{\partial f(y)}{\partial y_j} = \frac{1}{(2\pi)^m} \int (-it_j) e^{-it'y} \phi(t) dt. \quad (3.9)$$

Then an unbiased estimator of $\int_{-N}^N \dots \int_{-N}^N (-it_j) e^{-it'y} \phi(t) dt$ is

$$\tilde{f}_{N,j}^{(1)}(y) = \frac{1}{(2\pi)^m} \int_{[-N,N]^m} (-it_j) e^{-it'y} \hat{\phi}_n(t) dt, \quad (3.10)$$

with $[-N, N]^m = [-N, N] \times \dots \times [-N, N]$ for some finite number $0 < N < \infty$ and $\hat{\phi}_n(t) = n^{-1} \sum_{j=1}^n e^{it'Y_j}$. Again a smooth version of $\tilde{f}_{N,j}^{(1)}(y)$ is defined as

$$\hat{f}_{N,j}^{(1)}(y) = \frac{1}{(2\pi)^m} \int g_N(\mathbf{t}) (-it_j) e^{-it'y} \hat{\phi}_n(t) dt, \quad (3.11)$$

6 Empirical Bayes estimation

where $g_N(\mathbf{t}) = \prod_{i=1}^m g_N(t_i)$ with $g_N(t_i)$ is given by (3.6) with M replaced by N . By choosing $N = ((\gamma_0 m)^{-1} \log n)^{1/2}$, it can be shown that (see the appendix)

$$E \left(\hat{f}_{N,j}^{(1)}(y) - \frac{\partial f(y)}{\partial y_j} \right)^2 \leq c_4 n^{-1} (\log n)^{m+1}, \quad (3.12)$$

for some constant $c_4 > 0$ independent of n and y , where $\partial f(y)/\partial y_j$ is given by (3.9).

Remark 3.1. The density estimators (3.5) and (3.11) present a multivariate extension of similar estimators that have been investigated by O'Bryan and Susarla [9] for the univariate case. Their estimators are, however, not smooth. The rates of convergence mentioned at (3.8) and (3.12) are much faster than those which have been obtained by using the kernel-type density estimators; see, for example, Wei [19] and Wei and Zhang [20]. The reason for better rates attained here is that we have exploited the fact that the density $f(y)$ to be estimated is a mixture of multivariate normal densities.

4. Empirical Bayes estimator and main results

In view of (2.7), an EB estimator of $\delta_G(y)$ can be obtained by replacing $\psi(y)$ in (2.7) with an estimator of $\psi(y)$ given by (2.8) based on Y_1, \dots, Y_n , given $Y_{n+1} = y$. For this purpose, following Pensky [11] and Penskaya [10], we introduce the function

$$\Delta(x, y, \alpha) = xy^{-1} (1 + \alpha x^{2\tau} y^{-2\tau})^{-\omega}, \quad \tau = 1, 2, \dots; \omega > 0; 2\tau\omega \geq 1, \alpha > 0; \quad (4.1)$$

and define our proposed EB estimator of $\psi(y)$ as

$$\hat{\psi}(y) = (\hat{\psi}_1(y), \dots, \hat{\psi}_m(y))', \quad (4.2)$$

where

$$\hat{\psi}_j(y) = \Delta(\hat{f}_{N,j}^{(1)}(y), \hat{f}_M(y), \alpha_n(y)), \quad j = 1, \dots, m, \quad (4.3)$$

with $\hat{f}_M(y)$ and $\hat{f}_{N,j}^{(1)}(y)$ are given by (3.5) and (3.11), respectively, $j = 1, \dots, m$, and $\alpha_n(y)$ is a positive number that is to be specified later. Note that $\hat{\psi}(y)$ defined by (4.2) depends on n but this is suppressed for notational convenience here and in what follows.

The (nuisance) parameter σ^2 in (2.7) is estimated by

$$\hat{\sigma}^2 = \begin{cases} \tilde{\sigma}^2 & \text{if } \gamma_0 \leq \tilde{\sigma}^2 \leq \gamma_1, \\ \gamma_1 & \text{if } \tilde{\sigma}^2 > \gamma_1, \\ \gamma_0 & \text{if } \tilde{\sigma}^2 < \gamma_0, \end{cases} \quad (4.4)$$

where γ_0 and γ_1 are as given in Section 2; see, before formula (2.1), and

$$\tilde{\sigma}^2 = n^{-1} \sum_{i=1}^n \tilde{\sigma}_{(i)}^2, \quad \tilde{\sigma}_{(i)}^2 = \frac{Y_i' H Y_i}{s}, \quad (4.5)$$

with $H = I_m - X(X'X)^{-1}X'$ and $s = m - k$. Note that H is an idempotent matrix such that $\text{rank}(H) = s$ and $s\tilde{\sigma}_{(i)}^2/\sigma^2 \sim \chi_{(s)}^2$, the chi-square distribution with s degrees of freedom. The estimator $\tilde{\sigma}^2$ has been previously studied by Rao [14] under the same model as in (1.1).

Finally, our proposed EB estimator of $\delta_G(y)$ given by (2.7) is defined by

$$\hat{\phi}(y) = (X'X)^{-1}X'(y + \hat{\sigma}^2\hat{\psi}(y)), \tag{4.6}$$

where $\hat{\psi}$ and $\hat{\sigma}^2$ are given by (4.2) and (4.4), respectively.

We will measure the performance of $\hat{\phi}(y)$ by the quantity

$$R_n(y) = E\|\hat{\phi}(y) - \delta_G(y)\|^2, \tag{4.7}$$

where the expectation E is with respect to density $\prod_{i=1}^n f(y_i)$ with $f(y)$ given by (2.3). In typical EB problems, the quality of an EB estimator is measured by the regret. In the present problem, the regret of $\hat{\phi}$ is equal to $\int R_n(y)f(y)dy$ with $R_n(y)$ given by (4.7). However, Pensky [11] pointed out at least two advantages of using (4.7) compared to the regret. First, $R_n(y)$ enables one to calculate the mean squared error for the given observation $Y_{n+1} = y$ which is the interesting quantity. Second, by using the risk function (4.7), we eliminate the influence on the risk function of the observations having very low probabilities. So, the use of $R_n(y)$ provides a way of getting EB estimators with better convergence rates. The next theorem is the main result of this paper, which establishes the rates of convergence of $R_n(y)$.

THEOREM 4.1. *Let $\delta_G(y)$ be given by (2.7) and let $\hat{\phi}(y)$ be defined by (4.6) with $M = N = ((y_0m)^{-1} \log n)^{1/2}$ in (3.5) and (3.11), respectively. Let $\alpha_n(y)$ in (4.3) be given by $\alpha_n(y) = n^{-2\tau/(2\tau+1)}(\log n)^{\tau(4m+1+\varepsilon)/(2\tau+1)}$ for some number $0 < \varepsilon < 1$ and τ is as defined in (4.1). Then*

$$\lim_{n \rightarrow \infty} n(\log n)^{-(m+1)} E\|\hat{\phi}(y) - \delta_G(y)\|^2 < \infty. \tag{4.8}$$

Remark 4.2. The inequality established in Theorem 4.1 means that there exists some constant $c_0 > 0$ such that for every large n ,

$$E_n\|\hat{\phi}(y) - \delta_G(y)\|^2 < c_0n^{-1}(\log n)^{m+1}. \tag{4.9}$$

The preceding inequality shows a very fast convergence rate for the pointwise mean squared error of the proposed EB estimator $\hat{\phi}(y)$. It is easy to show that for the EB estimators of Zhang and Wei [22], Wei and Zhang [20], and Wei [19], which are based on kernel-type density estimates, the rate of the mean squared error is only of the order $n^{-\rho}$ with ρ ($0 < \rho < 1$) determined by the behavior of the prior distribution G . Thus, the proposed estimator has a faster convergence rate for the mean squared error compared to other competitors.

Remark 4.3. The purpose of introducing the function $\Delta(x, y, \alpha)$ given at (4.1) is to construct a mean square consistent estimator of the ratio $\psi(y) = \Delta f(y)/f(y)$ given by (2.8). Other possible forms of estimators of a ratio are given in Penskaya [10]. The “ τ ” in (4.1) is chosen as a positive integer because it makes 2τ an even positive number. Therefore,

8 Empirical Bayes estimation

the ratio $x^{2\tau}/y^{2\tau}$ in (4.1) is a well-defined quantity. In applications, one can simply take $\tau = 1$. Furthermore, for w in (4.1), any positive number can be chosen so that $2\tau w \geq 1$ for the chosen value of τ . Thus, it is very easy to implement the proposed EB estimator $\hat{\phi}(y)$ defined by (4.6) in practical applications.

Remark 4.4. In the model (1.1), we have assumed that $E(\varepsilon_j | \beta_j) = 0$ and $\text{Var}(\varepsilon_j | \beta_j) = \sigma^2 I_m$, where σ^2 is an unknown constant and I_m denotes the identity matrix of order $m \times m$. The preceding variance assumption can be replaced by a more general assumption that $\text{Var}(\varepsilon_j | \beta_j) = \sigma^2 V$, where V is a known nonsingular matrix. In this case, the proposed EB estimator defined at (4.6) takes the form

$$\hat{\phi}(y) = (X' V^{-1} X)^{-1} X' V^{-1} (y + \hat{\sigma}^2 \hat{\psi}(y)), \quad (4.10)$$

where $\hat{\psi}(y)$ is given by (4.2) and $\hat{\sigma}^2$ is given by (4.4) with $\tilde{\sigma}^2$ is replaced by $\tilde{\sigma}_1^2$, where

$$\tilde{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (Y_i' V^{-1} Y_i - Y_i' V^{-1} X \hat{\beta}_i) \quad (4.11)$$

with $\hat{\beta}_i = (X' V^{-1} X)^{-1} X' V^{-1} Y_i$.

5. Applications

In this section, we discuss two applications. In each case, the construction of the Bayes estimator as well as an empirical Bayes estimator is exhibited.

5.1. One-way classification. Suppose we have data in I groups, with J_i ($i = 1, \dots, I$) observations in each group as follows:

$$\begin{aligned} \text{Group 1 : } & Y_{11}, Y_{12}, \dots, Y_{1J_1} \text{ mean} = \bar{Y}_1, \\ \text{Group 2 : } & Y_{21}, Y_{22}, \dots, Y_{2J_2} \text{ mean} = \bar{Y}_2, \\ & \dots \\ \text{Group } I : & Y_{I1}, Y_{I2}, \dots, Y_{IJ_I} \text{ mean} = \bar{Y}_I. \end{aligned} \quad (5.1)$$

The usual fixed-effects analysis of variance model for such a situation is

$$Y_{ij} = \mu_i + e_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, J_i, \quad (5.2)$$

where e_{ij} are i.i.d. (independently and identically distributed) as $N(0, \sigma^2)$. Then, in regression terms, we can write the model (5.2) as

$$Y = X\beta + \varepsilon, \quad (5.3)$$

where $Y = (Y_{11}, Y_{12}, \dots, Y_{1I_1}; Y_{21}, Y_{22}, \dots, Y_{2I_2}; \dots; Y_{I_1}, Y_{I_2}, \dots, Y_{I I_I})'$, $\varepsilon = (e_{11}, e_{12}, \dots, e_{1I_1}; e_{21}, e_{22}, \dots, e_{2I_2}; \dots; e_{I_1}, e_{I_2}, \dots, e_{I I_I})'$, $\beta = (\mu_1, \mu_2, \dots, \mu_I)'$, and

$$X = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 0 & \cdots & 0 \\ \hline \cdots & \cdots & \cdots & \cdots & \cdots \\ \hline 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{5.4}$$

Note that $\varepsilon \sim N(0, \sigma^2 I_m)$ with $m = \sum_{i=1}^I J_i$. Then

$$X'X = \begin{bmatrix} J_1 & & & & \\ & J_2 & & & \\ & & \ddots & & \\ & & & J_{I-1} & \\ & & & & J_I \end{bmatrix}, \quad X'Y = \begin{bmatrix} J_1 \bar{Y}_1 \\ J_2 \bar{Y}_2 \\ \vdots \\ J_I \bar{Y}_I \end{bmatrix}. \tag{5.5}$$

Since $X'X$ is a diagonal matrix with J_i in the i th diagonal position and is zero elsewhere, its inverse is a diagonal matrix with $1/J_i$ in the i th diagonal position. Then the Bayes estimator of β given by (2.7) takes the form

$$\delta_G(Y) = \bar{Y} + \sigma^2 (X'X)^{-1} X' \psi(Y), \tag{5.6}$$

where $\bar{Y} = (\bar{Y}_1, \dots, \bar{Y}_I)'$, X and $(X'X)^{-1}$ are given by (5.4) and (5.5), respectively, and $\psi(y)$ is given by (2.8). An empirical Bayes estimator of (5.6) can be constructed by replacing σ^2 and $\psi(y)$ by estimators $\hat{\sigma}^2$ and $\hat{\psi}(y)$ defined by (4.2) and (4.4), respectively, compare with (4.6). For instance, $\tilde{\sigma}^2$ defined by (4.5) now becomes

$$\tilde{\sigma}^2 = \frac{1}{n(m-I)} \sum_{i=1}^n Y_i' (I_m - X(X'X)^{-1}X') Y_i, \tag{5.7}$$

where X given by (5.4), $m = \sum_{i=1}^I J_i$, and $\{Y_i\}_{i=1}^n$ are some independent auxiliary data satisfying the same form as (5.3), that is, $Y_i = X\beta_i + \varepsilon_i$, $i = 1, \dots, n$. For $\hat{\psi}(y)$, one could use the estimator $\tilde{f}_M(y)$, given by (3.4), and its first partial derivatives and then employ

the function (4.1) to form an estimator as in (4.2) and (4.3). Note that $\tilde{f}_M(y)$ is equal to

$$\tilde{f}_M(y) = \frac{M^m}{n} \sum_{j=1}^n \left\{ \prod_{i=1}^m K(M(y_i - Y_{ji})) \right\}, \quad (5.8)$$

where $y = (y_1, \dots, y_m)'$, $Y_j = (Y_{j1}, \dots, Y_{jm})'$, $j = 1, \dots, n$, and $K(x) = \sin x / \pi x$. Thus, $\tilde{f}_M(y)$ has the usual kernel estimator form in a multivariate setting.

5.2. Two-way classification. Consider an experiment in which two factors A and B are allowed to vary. Suppose there are I levels of A and J levels of B , and let Y_{ijk} be the k th experimental observation ($k = 1, 2, \dots, K$, $K > 1$) on the combination of the i th level of A with the j th level of B . We assume that the Y_{ijk} are independently distributed as $N(\mu_{ij}, \sigma^2)$ so that

$$Y_{ijk} = \mu_{ij} + e_{ijk}, \quad i = 1, 2, \dots, I; \quad j = 1, 2, \dots, J; \quad k = 1, 2, \dots, K, \quad (5.9)$$

where the e_{ijk} are i.i.d. $N(0, \sigma^2)$. By writing

$$\begin{aligned} Y' &= (Y_{111}, Y_{112}, \dots, Y_{11K}, Y_{121}, Y_{122}, \dots, Y_{12K}, \dots, Y_{I11}, Y_{I12}, \dots, Y_{IJK}), \\ \beta' &= (\mu_{11}, \mu_{12}, \dots, \mu_{1J}, \mu_{21}, \mu_{22}, \dots, \mu_{2J}, \dots, \mu_{I1}, \mu_{I2}, \dots, \mu_{IJ}), \end{aligned} \quad (5.10)$$

and so forth, and using the same approach which led to (5.3), we find that (5.9) can be expressed in the regression form

$$Y = X\beta + \varepsilon, \quad (5.11)$$

where ε is $N(0, \sigma^2 I_m)$, $m = IJK$, and X is $m \times IJ$ of rank IJ . Then the Bayes estimator of β given by (2.7) is obtained by using the X and Y of (5.9). Again, an empirical Bayes estimator can be constructed by replacing σ^2 and $\psi(y)$ by their estimators based on some auxiliary data $\{Y_i\}_{i=1}^n$ as in Section 5.1.

6. Concluding remarks

In this paper, we investigated empirical Bayes estimation in a multivariate regression model of the form $Y = X\beta + \varepsilon$. An empirical Bayes estimator of β was constructed with very fast rates of convergence of the corresponding mean squared error. The proposed empirical Bayes estimator is based on a newly developed multivariate density estimator of a normal mixture density and its first derivatives. The preceding results also extend similar results of the univariate case of a normal mixture density available in the literature. The results of the present paper can be easily extended to more complicated linear models such as $Y = Z\theta + X\beta + \varepsilon$ and $Y = U\mu + Z\alpha + X\beta + \varepsilon$.

The model and the estimator we developed in this paper are applicable in a wide variety of contexts. For instance, they can be used in the construction of a selection index for choosing individuals with a high intrinsic genetic value. For example, β_i may represent unknown genetic parameters and Y_i are observable characteristics on the i th individual, while $p'\beta_i$ for given p is the genetic value to be estimated in terms of observed Y_i . Early examples of such estimation by computing the regression of $p'\beta_i$ on Y_i (suggested by R. A. Fisher) are due to Fairfield Smith [6]. A detailed study of the estimation problem from a decision theoretic view point with an estimated parametric prior distribution of β_i is given by Rao [13]. Some applications are given by Rao [12, 13], and further developments of parametric estimation theory are developed by Rao [14], Efron and Morris [3–5], and Bunke and Gladitz [2], among others.

Appendix

Proof of (3.3). From (3.2), for any $0 < M < \infty$, we have

$$\begin{aligned}
 & \left| (2\pi)^m f(y) - \int_{-M}^M \cdots \int_{-M}^M e^{-it'y} \phi(t) dt \right| \\
 & \leq c \int_M^\infty \cdots \int_M^\infty |\phi(t)| dt \\
 & \leq c \int_M^\infty \cdots \int_M^\infty e^{-\sigma^2 t'/2} dt \\
 & = c \left(\int_M^\infty e^{-\sigma^2 t_1^2/2} dt_1 \right)^m \\
 & = c \left(\sigma^{-1} \int_{M\sigma}^\infty e^{-s^2/2} ds \right)^m \tag{A.1} \\
 & \leq c (\gamma_0)^{-1/2} \left(\int_{M\sqrt{\gamma_0}}^\infty e^{-s^2/2} ds \right)^m \\
 & \leq c (\gamma_0)^{-m/2} \left(\frac{1}{M\sqrt{\gamma_0}} e^{-M^2\gamma_0/2} \right)^m \\
 & = \frac{c}{(\gamma_0)^{(m+1)/2}} \frac{1}{M^m} e^{-mM^2\gamma_0/2}, \quad (s = t_1\sigma) \quad (\sigma^2 \geq \gamma_0),
 \end{aligned}$$

where $c > 0$ is a constant independent of n , y , and M . This completes the proof. \square

Proof of (3.7). From (3.5), we write

$$\begin{aligned}
 E\hat{f}_M(y) & = \frac{1}{(2\pi)^m} \int_{[-M, M]^m} \mathbf{g}_M(\mathbf{t}) e^{-it'y} E(\phi_n(t)) dt \\
 & = \frac{1}{(2\pi)^m} \int_{[-M, M]^m} \mathbf{g}_M(\mathbf{t}) e^{-it'y} \phi(t) dt.
 \end{aligned} \tag{A.2}$$

12 Empirical Bayes estimation

From (3.2) and (A.1), we now have

$$E\hat{f}_M(y) - f(y) = \frac{1}{(2\pi)^m} \int_{[-M, M]^m} (g_M(\mathbf{t}) - 1) e^{-it'y} \phi(t) dt - \frac{1}{(2\pi)^m} \int_{([-M, M]^m)^c} e^{-it'y} \phi(t) dt, \quad (\text{A.3})$$

where $([-M, M]^m)^c$ is the complement of the set $[-M, M]^m$. As in (A.1), it is easy to show that the second term on the right-hand side of (A.3) is bounded as follows:

$$\left| \frac{1}{(2\pi)^m} \int_{([-M, M]^m)^c} e^{-it'y} \phi(t) dt \right| \leq c \int_M^\infty \cdots \int_M^\infty |\phi(t)| dt \leq c \frac{1}{(\gamma_0)^{(m+1)/2}} \frac{1}{M^m} e^{-mM^2\gamma_0/2}, \quad (\text{A.4})$$

where $c > 0$ is a positive constant. Using the definition of $g(\mathbf{t})$, the first term on the right-hand side of (A.3) can be bounded as follows:

$$\begin{aligned} & \left| \int_{[-M, M]^m} (g_M(\mathbf{t}) - 1) e^{-it'y} \phi(t) dt \right| \\ & \leq \int_{[-M, M]^m} |g_M(\mathbf{t}) - 1| dt \\ & = \left\{ \int_{-M}^{-M+a} (1 - g_M(t)) dt + \int_{M-a}^M (1 - g_M(t)) dt \right\}^m \\ & = \left(\frac{2}{M} e^{-M^2\gamma_0/2} \right)^m. \end{aligned} \quad (\text{A.5})$$

Using the i.i.d. property of Y_1, \dots, Y_n , we obtain the variance of $\hat{f}_M(y)$ as

$$\begin{aligned} \text{Var}(\hat{f}_M(y)) &= \frac{1}{n} \text{Var} \left\{ \frac{1}{(2\pi)^m} \int_{[-M, M]^m} g_M(t) e^{-it'y} e^{itY_1} dt \right\} \\ &\leq \frac{1}{n(2\pi)^m} E \left\{ \int_{[-M, M]^m} g_M(t) e^{-it'y} e^{itY_1} dt \right\}^2 \\ &\leq n^{-1} (\pi^{-1} M^2)^m. \end{aligned} \quad (\text{A.6})$$

By combining (A.3) to (A.6), we now complete the proof of (3.7). \square

Proof of (3.12). From (3.9) and (3.11), we have

$$E\hat{f}_{N,j}^{(1)}(y) - \frac{\partial f(y)}{\partial y_j} = \frac{1}{(2\pi)^m} \int_{[-N, N]^m} (g_N(\mathbf{t}) - 1) (-it_j) e^{-it'y} \phi(t) dt - \frac{1}{(2\pi)^m} \int_{([-N, N]^m)^c} (-it_j) e^{-it'y} \phi(t) dt. \quad (\text{A.7})$$

The second term on the right-hand side of (A.7) can be bounded as follows:

$$\begin{aligned}
& \left| \frac{1}{(2\pi)^m} \int_{([-N, N]^m)^c} (-it_j) e^{it'y} \phi(t) dt \right| \\
& \leq \frac{c_1}{(2\pi)^m} \int_N^\infty \cdots \int_N^\infty |t_j \phi(t)| dt \\
& = c_2 \int_N^\infty \cdots \int_N^\infty |t_j| e^{-\sigma^2 \sum_{j=1}^m t_j^2/2} dt_1 \cdots dt_m \\
& \leq c_2 \left(\int_N^\infty |t_j| e^{-\sigma^2 t_j^2/2} dt_j \right) \left(\int_N^\infty e^{-\sigma^2 t^2/2} dt \right)^{m-1} \\
& \leq c_3 \left(\frac{1}{N} \right)^{m-1} (e^{-\gamma_0 N^2/2})^m,
\end{aligned} \tag{A.8}$$

where c_i ($i = 1, 2, 3$) are positive constants independent of y , n , and N . The first term on the right-hand side of (A.7) can be bounded as follows:

$$\begin{aligned}
& \left| \int_{[-N, N]^m} (g_N(\mathbf{t}) - 1) (-it_j) e^{-it'y} \phi(t) dt \right| \\
& \leq c_4 \int_{[-N, N]^m} |g_N(\mathbf{t}) - 1| |t_j| dt \\
& \leq c_5 N \int_{[-N, N]^m} |g_N(\mathbf{t}) - 1| dt \\
& \leq c_6 N \left(\frac{2}{N} e^{-M^2 \gamma_0/2} \right)^N \\
& = c_6 \left(\frac{1}{N} \right)^{m-1} (e^{-\gamma_0 M^2/2})^N,
\end{aligned} \tag{A.9}$$

where c_i ($i = 4, 5, 6$) are positive constants independent of y , n , and N . Now the variance of $\hat{f}_{N,j}^{(j)}(y)$ is bounded as follows:

$$\begin{aligned}
\text{Var}(\hat{f}_{N,j}^{(1)}(y)) & \leq c_7 \text{Var} \left(\int_{[-N, N]^m} g_N(\mathbf{t}) (-it_j) e^{-it'y} e^{it'Y_1} dt \right) \\
& \leq c_7 E \left(\int_{[-N, N]^m} g_N(\mathbf{t}) (-it_j) e^{-it'y} e^{it'Y_1} dt \right)^2 \\
& \leq c_8 \frac{1}{n} (N^{m+1})^2,
\end{aligned} \tag{A.10}$$

for some positive constants c_7 and c_8 independent of n , y , and N . Now by combining (A.7) to (A.10) and choosing $N^2 = (\gamma_0 m)^{-1} \log n$, we complete the proof of (3.12). \square

In order to prove Theorem 4.1, we now state and prove two lemmas.

LEMMA A.1. *Let $f(y)$ be given by (3.2), and let $\hat{f}_M(y)$ be given by (3.5) with $M = ((\gamma_0 m)^{-1} \log n)^{1/2}$. Then*

$$E_n(\hat{f}_n(y) - f(y))^4 = O(n^{-2}(\log n)^{2m}). \quad (\text{A.11})$$

LEMMA A.2. *Let $\partial f(y)/\partial y_j$ be defined by (3.9), and let $\hat{f}_{N,j}^{(1)}(y)$ ($j = 1, \dots, m$) be defined by (3.11) with $N = ((\gamma_0 m)^{-1} \log n)^{1/2}$. Then*

$$E_n\left(\hat{f}_{N,j}^{(1)}(y) - \frac{\partial f(y)}{\partial y_j}\right)^4 = O(n^{-2}(\log n)^{4m+1}), \quad j = 1, \dots, m. \quad (\text{A.12})$$

Proof of Lemma A.1. First, observe that by Hölder's inequality,

$$\begin{aligned} E(\hat{f}_M(y) - E\hat{f}_M(y))^4 &= \frac{1}{(2\pi)^{2m}} E\left(\int_{[-M, M]^m} g_M(\mathbf{t}) e^{-it'y} (\hat{\phi}(t) - \phi(t)) dt\right)^4 \\ &\leq CE\left(\int_{[-M, M]^m} |\hat{\phi}(t) - \phi(t)| dt\right)^4 \\ &\leq C\left(\int_{[-M, M]^m} E|\hat{\phi}(t) - \phi(t)|^4 dt\right)\left(\int_{[-M, M]^m} dt\right)^3, \end{aligned} \quad (\text{A.13})$$

here and in what follows, expectation “ E ” denotes the expectation as defined in (4.7), and C denotes a constant independent of n , y , M , and N (which, in different positions, may take different values). Now write $\hat{\phi}(t) - \phi(t) = n^{-1} \sum_{j=1}^n T_j$, where $T_j = (e^{it'Y_j} - \phi(t))$. Denote $S_n = \sum_{j=1}^n T_j$. Then, since T_j s are i.i.d., we have

$$E(S_n^4) = \kappa_4(t) + 3\kappa^2(t) = n\kappa'_4(t) + 3n^2\kappa_2'^2(t), \quad (\text{A.14})$$

where $\kappa_r(t)$ and $\kappa'_r(t)$ denote the r th cumulant of S_n and T_1 , respectively (see, e.g., Barndorff-Nielson and Cox [1]). Therefore,

$$E(\hat{\phi}(t) - \phi(t))^4 = n^{-4}(n\kappa'_4(t) + 3n^2\kappa_2'^2(t)). \quad (\text{A.15})$$

Note that $\kappa_2'(t)$ and $\kappa_4'(t)$ are bounded functions of t , since T_j s are bounded. From (A.13) and (A.15), we obtain

$$\begin{aligned} E(\hat{f}_M(y) - E\hat{f}_M(y))^4 &\leq Cn^{-4}(2M)^{3m} \int_{[-M, M]^m} (n\kappa'_4(t) + 3n^2\kappa_2'^2(t)) dt \\ &\leq Cn^{-2}M^{4m}. \end{aligned} \quad (\text{A.16})$$

But from (A.2) to (A.5), we have

$$|E\hat{f}_M(y) - f(y)| \leq CM^{-m}e^{-m\gamma_0 M^2/2}. \quad (\text{A.17})$$

Now by combining (A.16) and (A.17) together with $M^2 = (\gamma_0 m)^{-1} \log n$, and the C_r -inequality (Loève [7], page 157), we obtain

$$\begin{aligned} E(\widehat{f}_M(y) - f(y))^4 &\leq C \left\{ n^{-2} M^{4m} + (M^{-m} e^{-m\gamma_0 M^2/2})^4 \right\} \\ &= C \left\{ n^{-2} (\log n)^{2m} + n^{-2} (\log n)^{2m} \right\} \\ &\leq C n^{-2} (\log n)^{2m}. \end{aligned} \quad (\text{A.18})$$

This completes the proof of Lemma A.1. \square

Proof of Lemma A.2. Again by Hölder's inequality, we have

$$\begin{aligned} E(\widehat{f}_{N,j}^{(1)}(y) - E\widehat{f}_{N,j}^{(1)}(y))^4 &\leq CE \left(\int_{[-N,N]^m} |t_j| |\widehat{\phi}(t) - \phi(t)| dt \right)^4 \\ &\leq C \left(\int_{[-N,N]^m} E |\widehat{\phi}(t) - \phi(t)|^4 dt \right) \left(\int_{[-N,N]^m} |t_j|^{4/3} dt \right)^3 \\ &\leq C n^{-4} \left\{ \int_{[-N,N]^m} (n\kappa'_4(t) + n^2\kappa'_2(t)) dt \right\} \{2^{3m} N^{4+3(m-1)}\} \\ &\leq C n^{-2} N^{4m+1}. \end{aligned} \quad (\text{A.19})$$

From (A.7) to (A.9), we have

$$\left| E\widehat{f}_{N,j}^{(1)}(y) - \frac{\partial f(y)}{\partial y_j} \right| \leq C N^{-(m-1)} e^{-m\gamma_0 N^2/2}. \quad (\text{A.20})$$

Now by combining inequalities (A.19) and (A.20) together with $N^2 = (\gamma_0 m)^{-1} \log n$, and the C_r -inequality, we obtain

$$\begin{aligned} E \left(\widehat{f}_{N,j}^{(1)}(y) - \frac{\partial f(y)}{\partial y_j} \right)^4 &\leq C \left\{ n^{-2} N^{4m+1} + (N^{-(m-1)} e^{-\gamma_0 m N^2/2})^4 \right\} \\ &\leq C \left\{ n^{-2} (\log n)^{4m+1} + n^{-2} \right\} \\ &\leq C n^{-2} (\log n)^{4m+1}. \end{aligned} \quad (\text{A.21})$$

This completes the proof of Lemma A.2. \square

Proof of Theorem 4.1. Let $\lambda_1 = \max[\text{root}(X(X'X)^{-1}X')]$, where $\max[\text{root}(A)]$ denotes the largest eigenvalue of a matrix A . Recall that $\gamma_0 \leq \widehat{\sigma}^2 \leq \gamma_1$ by construction. Then by (2.7)

and (4.6), we obtain

$$\begin{aligned}
\|\widehat{\phi}(y) - \delta_G(y)\|^2 &= \|\widehat{\sigma}^2(X'X)^{-1}X'\widehat{\psi}(y) - \sigma^2(X'X)^{-1}X'\psi(y)\|^2 \\
&\leq C\left\{\|(X'X)^{-1}X'\psi(y)(\widehat{\sigma}^2 - \sigma^2)\|^2 + \|(X'X)^{-1}X'\widehat{\sigma}^2(\widehat{\psi}(y) - \psi(y))\|^2\right\} \\
&\leq C\left\{\lambda_1\|(\widehat{\sigma}^2 - \sigma^2)\psi(y)\|^2 + \lambda_1\gamma_1^2\|\widehat{\psi}(y) - \psi(y)\|^2\right\} \\
&\leq C\left\{(\widehat{\sigma}^2 - \sigma^2)^2\|\psi(y)\|^2 + \|\widehat{\psi}(y) - \psi(y)\|^2\right\} \\
&= C\left\{(\widehat{\sigma}^2 - \sigma^2)^2\|\psi(y)\|^2 + \sum_{j=1}^m |\widehat{\psi}_j(y) - \psi_j(y)|^2\right\},
\end{aligned} \tag{A.22}$$

where $\psi_j(y)$ and $\widehat{\psi}_j(y)$ are as given in (2.8) and (4.3), respectively. From (4.4) and (4.5) we observe that $(\widehat{\sigma}^2 - \sigma^2)^2 \leq (\widetilde{\sigma}^2 - \sigma^2)^2$ and $\text{Var}(\widetilde{\sigma}_{(\ell)}^2) = 2\sigma^4/s$, see circa (4.5). Thus, for each y ,

$$\begin{aligned}
E(\widehat{\sigma}^2 - \sigma^2)^2\|\psi(y)\|^2 &\leq \|\psi(y)\|^2 E(\widetilde{\sigma}^2 - \sigma^2)^2 \\
&= \|\psi(y)\|^2 \text{Var}(\widetilde{\sigma}^2) \\
&= \frac{2\sigma^4\|\psi(y)\|^2}{ns}.
\end{aligned} \tag{A.23}$$

Now denote

$$d_n^2(y) = E\left(\widehat{f}_{N,j}^{(1)}(y) - f_j^{(1)}(y)\right)^2 + E(\widehat{f}_M(y) - f(y))^2, \tag{A.24}$$

where $f_j^{(1)}(y) = \partial f(y)/\partial y_j$. Then from (3.8) and (3.12), we have

$$d_n^2(y) = O(n^{-1}(\log n)^{m+1}). \tag{A.25}$$

In order to find the convergence rate of

$$E|\widehat{\psi}_j(y) - \psi_j(y)|^2 = E\left|\Delta\left(\widehat{f}_{j,N}^{(1)}(y), \widehat{f}_M(y), \alpha_n(y)\right) - \frac{f_j^{(1)}(y)}{f(y)}\right|^2, \tag{A.26}$$

we will use [10, Theorem 2]. According to the preceding theorem, for $n \rightarrow \infty$,

$$E\left(\Delta\left(\widehat{f}_{N,j}^{(1)}(y), \widehat{f}_M(y), \alpha_n(y)\right) - \frac{f_j^{(1)}(y)}{f(y)}\right)^2 = O(d_n^2(y) + \alpha_n^2(y)) \tag{A.27}$$

provided $\lim_{n \rightarrow \infty} \alpha_n(y) = 0$ and

$$\alpha_n^{-1/\tau} \left\{E\left(\widehat{f}_{N,j}^{(1)}(y) - f_j^{(1)}(y)\right)^4 + E(\widehat{f}_M(y) - f(y))^4\right\} = o(\alpha_n^2(y) + d_n^2(y)). \tag{A.28}$$

The preceding equality may be written in the form

$$\lim_{n \rightarrow \infty} \alpha_n^{-2-1/\tau} \left\{E\left(\widehat{f}_{N,j}^{(1)}(y) - f_j^{(1)}(y)\right)^4 + E(\widehat{f}_M(y) - f(y))^4\right\} = 0. \tag{A.29}$$

From Lemmas A.1 and A.2, we see that (A.29) is satisfied if $\alpha_n^{-2-1/\tau} = n^2(\log n)^{-(4m+1)-\varepsilon}$ for some positive number ε such that $0 < \varepsilon < 1$. That is, $\alpha_n = n^{-2\tau/(2\tau+1)}(\log n)^{\tau(4m+1+\varepsilon)/(2\tau+1)}$. With the preceding choice of α_n , we now obtain from (A.25) and (A.27) that

$$E|\hat{\psi}_j(y) - \psi_j(y)|^2 = O\left(n^{-1}(\log n)^{m+1} + n^{-4\tau/(2\tau+1)}(\log n)^{2\tau(4m+1+\varepsilon)/(2\tau+1)}\right). \quad (\text{A.30})$$

The proof is now completed by (A.22)-(A.23), and (A.30). \square

Acknowledgments

The research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors acknowledge gratefully many useful suggestions by the referees.

References

- [1] O. E. Barndorff-Nielsen and D. R. Cox, *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, 1989.
- [2] H. Bunke and J. Gladitz, *Empirical linear Bayes decision rules for a sequence of linear models with different regressor matrices*, *Mathematische Operationsforschung und Statistik* **5** (1974), no. 3, 235–244.
- [3] B. Efron and C. Morris, *Empirical Bayes on vector observations: an extension of Stein's method*, *Biometrika* **59** (1972), no. 2, 335–347.
- [4] ———, *Combining possibly related estimation problems*, *Journal of the Royal Statistical Society. Series B* **35** (1973), 379–421, with discussion.
- [5] ———, *Stein's estimation rule and its competitors—an empirical Bayes approach*, *Journal of the American Statistical Association* **68** (1973), no. 341, 117–130.
- [6] H. Fairfield Smith, *A discriminant function for plant selection*, *Annals of Eugenics (London)* **7** (1936), 240–253.
- [7] M. Loève, *Probability Theory I*, 4th ed., Springer-Verlag, New York, 1963.
- [8] H. F. Martz and R. G. Krutchkoff, *Empirical Bayes estimators in a multiple linear regression model*, *Biometrika* **56** (1969), no. 2, 367–374.
- [9] T. E. O'Bryan and V. Susarla, *Rates in the empirical Bayes estimation problem with non-identical components*, *Annals of the Institute of Statistical Mathematics* **28** (1976), no. 3, 389–397.
- [10] M. Y. Penskaya, *Mean square consistent estimation of a ratio*, *Scandinavian Journal of Statistics* **22** (1995), no. 1, 129–137.
- [11] M. Pensky, *Empirical Bayes estimation of a location parameter*, *Statistics & Decisions* **15** (1997), no. 1, 1–16.
- [12] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, New York, 1952.
- [13] ———, *Discriminant functions for genetic differentiation and selection*, *Sankhyā* **12** (1953), 229–246.
- [14] ———, *Simultaneous estimation of parameters in different linear models and applications to biometric problems*, *Biometrics* **31** (1975), 545–554.
- [15] H. Robbins, *An empirical Bayes approach to statistics*, *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I*, University of California Press, California, 1956, pp. 157–163.
- [16] ———, *The empirical Bayes approach to statistical decision problems*, *Annals of Mathematical Statistics* **35** (1964), 1–20.

- [17] R. S. Singh, *Empirical Bayes estimation in a multiple linear regression model*, Annals of the Institute of Statistical Mathematics **37** (1985), no. 1, 71–86.
- [18] R. S. Singh and M. Pensky, *Non-parametric estimation of prior densities of multidimensional location and scale parameters with rates and best possible rates of convergence*, The Journal of Mathematical Sciences. New Series **1** (2002), 86–105.
- [19] L. Wei, *Convergence rates of empirical Bayesian estimation in a class of linear models*, Statistica Sinica **8** (1998), no. 2, 589–605.
- [20] L. Wei and S. P. Zhang, *The convergence rates of empirical Bayes estimation in a multiple linear regression model*, Annals of the Institute of Statistical Mathematics **47** (1995), no. 1, 81–97.
- [21] S. L. Wind, *An empirical Bayes approach to multiple linear regression*, The Annals of Statistics **1** (1973), 93–103.
- [22] S. P. Zhang and L. Wei, *The asymptotically optimal empirical Bayes estimation in multiple linear regression model*, Applied Mathematics. A Journal of Chinese Universities **9** (1994), no. 3, 245–258.

R. J. Karunamuni: Department of Mathematical and Statistical Sciences, University of Alberta,
Edmonton, AB, Canada T6G 2G1
E-mail address: r.j.karunamuni@ualberta.ca

L. Wei: Department of Statistics and Finance, University of Science and Technology of China, Hefei,
Anhui 230026, China
E-mail address: lwei@ustc.edu.cn