Ann. Inst. Statist. Math.
Vol. 47, No. 1, 81–97 (1995)

THE CONVERGENCE RATES OF EMPIRICAL BAYES ESTIMATION IN A MULTIPLE LINEAR REGRESSION MODEL*

LAISHENG WEI[1] AND SHUNPU ZHANG[2]

[1] *Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, 230026, China*
[2] *Department of Mathematics, Hangzhou Normal College, Zhejiang, 310036, China*

(Received January 17, 1994; revised July 27, 1994)

Abstract. Empirical Bayes (EB) estimation of the parameter vector $\theta = (\beta', \sigma^2)'$ in a multiple linear regression model $Y = X\beta + \epsilon$ is considered, where β is the vector of regression coefficient, $\epsilon \sim N(0, \sigma^2 I)$ and σ^2 is unknown. In this paper, we have constructed the EB estimators of θ by using the kernel estimation of multivariate density function and its partial derivatives. Under suitable conditions it is shown that the convergence rates of the EB estimators are $O(n^{-(\lambda k-1)(k-2)/k(2k+p+1)})$, where the natural number $k \geq 3$, $\frac{1}{3} < \lambda < 1$, and p is the dimension of vector β.

Key words and phrases: Empirical Bayes estimation, multiple linear regression model, convergence rates.

1. Introduction

Since the empirical Bayes (EB) procedure was first suggested by Robbins (1955), the EB estimation problems have been studied in a great deal of literature. Suppose there is a pair (X, θ) of random variables, where r.v. X is observable and parameter (vector) θ is unobservable. The conditional distribution of X given θ is specified by density f_θ, and θ has an unknown and unspecified distribution G on parameter space Θ. Based on an observation on X (which could be a sufficient statistics for θ), the problem is to decide about θ under nonnegative loss function. If the prior distribution G were known, we could use the Bayes estimator $\hat{\theta}_G$ which achieves the minimum Bayes risk $R(G)$ relative to G. But since G is not known, and therefore the optimal estimator $\hat{\theta}_G$ is not directly available. In the EB decision problem, we assume that the above problem has occurred independently in the past, say n times. Hence there are $n+1$ independent pairs $(X_1, \theta_1), \dots, (X_n, \theta_n)$ and (X, θ). Our purpose is to use the information contained in the past observation $(X_1, \dots, X_n)$ and the present observation X to obtain the estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n; X)$ for the present parameter θ. This estimator

* The project is supported by the National Natural Science Foundation of China.

is called EB estimator. So that for large n, this estimator is "nearly" as good as the unavailable optimal estimator $\hat{\theta}_G$ in the following sense: the overall risk, say R_n, of $\hat{\theta}_n$ approximates the minimum Bayes risk $R(G)$ achieved by $\hat{\theta}_G$. If $\lim_{n\to\infty} R_n = R(G)$, then the estimator is called asymptotically optimal (a.o.). If for some $\delta > 0$, $R_n - R(G) = O(n^{-\delta})$, we will say that the EB estimator is a.o. with convergence rate $O(n^{-\delta})$. These are two large sample properties of EB estimators. From the above statement we know that the EB approach to statistical decision problems is applicable when the same decision problem presents itself repeatedly and independently with a fixed but unknown prior distribution of the parameters.

The EB estimation problems for exponential families have been discussed. Singh (1976, 1979) considered the EB estimation problem for noncontinuous and continuous Lebesgue exponential families respectively. Chen (1983) studied the asymptotic optimality (a.o.) of the EB estimators for one-dimensional discrete exponential family. Singh and Wei (1992) considered the EB problem for scale exponential family. About the multi-parameter exponential families, Tao (1986) and Zhang (1985) discussed the EB estimation problems for parameters in a normal distribution family. Wei (1985, 1987) considered the EB estimation problems for continuous type multi-parameter exponential family. Recently, Yang and Wei (1993, 1994) studied the EB estimation problems for multi-parameter discrete exponential family. Singh (1985) and Wei (1990) considered the EB estimation and test problem in a multiple linear regression model for the regression coefficient under known error-variance. In this paper we will consider the EB estimation problem of regression coefficient and the unknown error-variance for the following multiple linear regression model

$$Y_{l\times 1} = X_{l\times p}\beta_{p\times 1} + \epsilon_{l\times 1} \tag{1.1}$$

where $\epsilon \sim N_l(0, \sigma^2 I)$ with unknown σ^2 and $l \geq p+2$, denote $\theta = (\beta', \sigma^2)' \in \Theta$, $(\Theta, \mathcal{B}_\Theta)$ is the parametric space. Let the loss function be as follows

$$L(\theta, d) = \|d_1 - \beta\|^2/\sigma^2 + (d_{p+1} - \sigma^2)^2/\sigma^4 \tag{1.2}$$

where $d = (d_1', d_{p+1})' \in \mathcal{D}$, $(\mathcal{D}, \mathcal{B}_\mathcal{D})$ is the decision space and $\|t\|^2 = \sum_{i=1}^p t_i^2$ for a vector $t = (t_1, \dots, t_p)'$.

In this paper we will consider the EB estimation problem for $\theta = (\beta', \sigma^2)'$ in the model (1.1) with unknown σ^2.

Suppose that the prior distribution G of θ belongs to the following family

$$\mathcal{F}_k = \left\{ G(\theta) : \int_\Theta \sigma^{-(2k+l)} dG(\theta) < \infty \right\} \tag{1.3}$$

where $k \geq 3$ is an integer.

It is well known that the Least Square Estimators of β and σ^2 are

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)' = \Sigma X'Y, \tag{1.4}$$

$$\hat{\sigma}^2 = \|Y - \hat{\beta}\|^2/(l-p) \tag{1.5}$$

where $\Sigma = (X'X)^{-1}$. Since $(\hat{\beta}', \hat{\sigma}^2)$ is sufficient for (β', σ^2), we may substitute $Z = (\hat{\beta}', \hat{\sigma}^2)'$ for the original sample Y, where $Z \in \mathcal{Z}, (\mathcal{Z}, \mathcal{B}_{\mathcal{Z}})$ can be seen a sample space.

Let $\psi_G(Y) = (\psi_G^1, \psi_G^2, \ldots, \psi_G^p)'$ be the Bayes estimator of β and $\phi_G(Y)$ be the Bayes estimator of σ^2 under the unknown prior distribution G. By the Corollary 1.1 in Chapter 4 of Lehmann (1983), we have

$$\psi_G = (\psi_G^1, \psi_G^2, \ldots, \psi_G^p)' = E(\sigma^{-2}\beta \mid Z)/E(\sigma^{-2} \mid Z), \tag{1.6}$$

$$\phi_G = E(\sigma^{-2} \mid Z)/E(\sigma^{-4} \mid Z). \tag{1.7}$$

From (1.4), (1.5) we know that the conditional distributions of $\hat{\beta}$ and $\hat{\sigma}^2$ with given β and σ^2 are

$$\hat{\beta} \mid \theta \sim N(\beta, \sigma^2\Sigma), \qquad \frac{(l-p)\hat{\sigma}^2}{\sigma^2} \mid \theta \sim \chi^2_{l-p}.$$

Since $\hat{\beta}, \hat{\sigma}^2$ are independent when $\theta = (\beta', \sigma^2)'$ is given, the conditional density function of Z given θ is

$$\begin{aligned} f(z \mid \theta) &= f_1(\hat{\beta} \mid \theta) \cdot f_2(\hat{\sigma}^2 \mid \theta) \\ &= c\sigma^{-l}\hat{\sigma}^{l-p-2} \exp\left\{-\frac{(l-p)\hat{\sigma}^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}(\hat{\beta}-\beta)'\Sigma^{-1}(\hat{\beta}-\beta)\right\} \end{aligned} \tag{1.8}$$

where $l \geq p+2$, $c = [(l-p)/2]^{(l-p)/2}/[(2\pi)^{p/2}|\Sigma|^{1/2}\Gamma((l-p)/2)]$. The marginal density of $Z = (\hat{\beta}', \hat{\sigma}^2)'$ is

$$\begin{aligned} f(z) &= \int f(z \mid \theta)dG(\theta) \\ &= c\int \sigma^{-l}\hat{\sigma}^{l-p-2} \exp\left\{-\frac{(l-p)\hat{\sigma}^2}{2\sigma^2}\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}(\hat{\beta}-\beta)'\Sigma^{-1}(\hat{\beta}-\beta)\right\} dG(\theta). \end{aligned} \tag{1.9}$$

Let $g(z \mid \theta) = f(z \mid \theta)/\hat{\sigma}^{l-p-2}$, then

$$g(z \mid \theta) = c\sigma^{-l} \exp\left\{-\frac{(l-p)\hat{\sigma}^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}(\hat{\beta}-\beta)'\Sigma^{-1}(\hat{\beta}-\beta)\right\}, \tag{1.10}$$

$$\begin{aligned} g(z) &= \int g(z \mid \theta)dG(\theta) \\ &= c\int \sigma^{-l} \exp\left\{-\frac{(l-p)\hat{\sigma}^2}{2\sigma^2}\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}(\hat{\beta}-\beta)'\Sigma^{-1}(\hat{\beta}-\beta)\right\} dG(\theta). \end{aligned} \tag{1.11}$$

In the context of this paper $c, c_1, c_2, \ldots$ always stand for positive constants. They may denote different values even within the same expression.

LEMMA 1.1. *Let $G \in \mathcal{F}_k$, $\mathcal{F}_k$ is defined by (1.3), then $g(z)$ has continuous r-th order mixed partial derivatives*

$$g^{(r)}(z) = \frac{\partial^r g(z)}{\partial \hat{\beta}_1^{r_1} \cdots \partial \hat{\beta}_p^{r_p} \partial (\hat{\sigma}^2)^{r_{p+1}}},$$
$$\left(0 \le r_i, \quad i = 1, \ldots, p+1, \qquad r = \sum_{i=1}^{p+1} r_i, \quad 0 \le r \le k\right)$$

which satisfies

$$g^{(r)}(z) = \int g^{(r)}(z \mid \theta) dG(\theta) \qquad \text{and} \qquad |g^{(r)}(z)| \le \alpha \tag{1.12}$$

where α is a constant independent of G and Z.

PROOF. This lemma is a direct extension of Lemma 2 of Tao (1986).

From Lemma 1.1 and formula (1.11), we have

$$\begin{aligned}\frac{\partial g}{\partial \hat{\beta}} &= -\Sigma^{-1} \int \sigma^{-2}(\hat{\beta} - \beta) g(z \mid \theta) dG \\ &= -g(z)\Sigma^{-1} E(\sigma^{-2}(\hat{\beta} - \beta) \mid Z)\end{aligned}$$

therefore

$$E(\sigma^{-2}(\hat{\beta} - \beta) \mid Z) = -g^{-1}(z)\Sigma \frac{\partial g}{\partial \hat{\beta}}. \tag{1.13}$$

Analogously we have

$$E(\sigma^{-2} \mid Z) = -\frac{2}{l-p} g^{-1}(z) \frac{\partial g}{\partial \hat{\sigma}^2}, \tag{1.14}$$

$$E(\sigma^{-4} \mid Z) = \frac{4}{(l-p)^2} g^{-1}(z) \frac{\partial^2 g}{(\partial \hat{\sigma}^2)^2}. \tag{1.15}$$

Let

$$g_{\hat{\beta}_i}^{(1)} = \frac{\partial g}{\partial \hat{\beta}_i}, \quad i = 1, 2, \ldots, p; \qquad g_{\hat{\sigma}^2}^{(1)} = \frac{\partial g}{\partial \hat{\sigma}^2}, \qquad g_{\hat{\sigma}^2}^{(2)} = \frac{\partial^2 g}{\partial (\hat{\sigma}^2)^2}$$

and

$$q(z) = \frac{\partial g}{\partial \hat{\beta}} \Big/ \frac{\partial g}{\partial \hat{\sigma}^2} = (q_1, \ldots, q_p)', \qquad q_i = \frac{\partial g^{(1)}}{\partial \hat{\beta}_i} \Big/ \frac{\partial g^{(1)}}{\partial \hat{\sigma}^2}.$$

From (1.13) to (1.15) and (1.6), (1.7) we get the Bayes estimators of β and σ^2 as follows respectively

$$\psi_G = E(\sigma^{-2}\beta \mid Z)/E(\sigma^{-2} \mid Z) = \hat{\beta} - \frac{(l-p)}{2} \Sigma q(z), \tag{1.16}$$

$$\phi_G = E(\sigma^{-2} \mid Z)/E(\sigma^{-4} \mid Z) = -\frac{l-p}{2} g_{\hat{\sigma}^2}^{(1)} / g_{\hat{\sigma}^2}^{(2)}. \tag{1.17}$$

Let $\hat{\theta} = (\psi_G', \phi_G)'$ be the Bayes estimator of θ, then the minimum Bayes risk with respect to (w.r.t.) G is

$$(1.18)\quad \begin{aligned} R(G) = R(\hat{\theta}, G) &= E_{(Z,\theta)}[\|\psi_G - \beta\|^2/\sigma^2] + E_{(Z,\theta)}[(\phi_G - \sigma^2)^2/\sigma^4] \\ &= R_1(G) + R_2(G) \end{aligned}$$

where $E_{(Z,\theta)}$ denotes the expectation w.r.t. the joint distribution of random vectors Z and θ.

We know that $R(G) = \inf_{\theta^*} R(\theta^*, G)$, where the inf is taken over the set of all estimators θ^* for which $R(\theta^*, G)$ is finite. The estimator which achieves the minimum Bayes risk (i.e., Bayes envelope) $R(G)$ is the Bayes estimator, also called optimal estimator (o.e.) $\hat{\theta} = (\psi_G', \phi_G)'$ given by (1.6) and (1.7). Thus $R(\hat{\theta}, G) = R(G)$. Notice that $R(G)$ can be exactly achieved only if the prior distribution G is known and θ is estimated by o.e. $\hat{\theta}$. Unfortunately G is completely unknown and hence $\hat{\theta}$ is unavailable to us. This leads us to use EB approach to exhibit estimators whose risks are close to $R(G)$ achieved by $\hat{\theta}$.

In Section 2 of this paper we will construct the EB estimators of $\theta = (\beta', \sigma^2)'$ and in Section 3 we will give several lemmas and get the convergence rates. Finally we will give an example in Section 4.

2. Proposed EB estimators of $\theta = (\beta', \sigma^2)'$

In the EB framework, we make the following assumptions: Let $\{Y_{(1)}, \beta_{(1)}, \sigma_1^2\}$, $\ldots, \{Y_{(n)}, \beta_{(n)}, \sigma_n^2\}$ be independent random pairs from the past experiments and $\{Y_{(n+1)}, \beta_{(n+1)}, \sigma_{n+1}^2\} = \{Y, \beta, \sigma^2\}$ be the present sample, with $Y_{(i)} = X\beta_{(i)} + \epsilon_{(i)}$, $i = 1, \ldots, n+1$. The vectors $Y_{(i)}, \beta_{(i)}, \epsilon_{(i)}$, $i = 1, \ldots, n$ behave like Y, β, ϵ described above. $(\beta_{(1)}', \sigma_1^2)', \ldots, (\beta_{(n)}', \sigma_n^2)'$ and $(\beta', \sigma^2)'$ are i.i.d. and have the common unknown prior distribution G. $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ are called the historical samples, Y is the present sample. Let $Z_i = (\hat{\beta}_{(i)}', \hat{\sigma}_i^2)'$, $i = 1, \ldots, n+1$, where $\hat{\beta}_{(i)} = \Sigma X' Y_{(i)}$, $\hat{\sigma}_i^2 = \|Y - X\hat{\beta}_{(i)}\|^2/(l-p)$, $\Sigma = (X'X)^{-1}$. Then $Z_1, \ldots, Z_n$, $Z_{n+1} = Z$ are i.i.d., and have the same marginal density (1.9). It is easy to know that $Z_1, \ldots, Z_{n+1}$ are seperately sufficient statistics for $\theta_{(i)} = (\beta_{(i)}', \sigma^2)'$, $i = 1, 2, \ldots, n+1$. Therefore we may substitute Z_i for $Y_{(i)}$, $i = 1, 2, \ldots, n+1$, and $Z_1, Z_2, \ldots, Z_n$ can be said the historical samples, Z is called the present sample.

In order to get the EB estimator of θ, we use a class of kernel function defined as follows to make the kernel estimation of multiple density and its derivatives.

Let $P_i(x_i)$, $x_i \in R^1$, $i = 0, 1, \ldots, k-1$ be a class of Borel measurable functions, satisfying the following conditions:

i) when $x_i \notin (0,1)$, $P_i(x_i) = 0$, $i = 0, 1, \ldots, k-1$,
ii) $P_i(x_i)$ is bounded in $(0,1)$,
iii) for each $0 \le i \le k-1$,

$$\frac{1}{l_0!} \int_0^1 y^{l_0} P_i(y)\,dy = \begin{cases} 1 & l_0 = i \\ 0 & l_0 \ne i,\ l_0 = 0, \ldots, k-1 \end{cases}$$

where $k \geq 3$ is an integer. Obviously $K_r(u) = \prod_{i=1}^{p+1} P_{r_i}(u_i)$ satisfies

$$\frac{1}{l_1! \cdots l_{p+1}!} \int_{R^{p+1}} K_r(u) u_1^{l_1} \cdots u_{p+1}^{l_{p+1}} du = \begin{cases} 1 & \text{if } l_i = r_i,\ i = 1, \ldots, p+1 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

where $u = (u_1, \ldots, u_{p+1})' \in R^{p+1}$, $r = \sum_{i=1}^{p+1} r_i$, $r_i \geq 0$, $0 \leq r \leq k-1$ and $0 \leq l_i \leq k-1$, $(i = 1, 2, \ldots, p+1)$, $0 \leq \sum_{i=1}^{p+1} l_i \leq k-1$. Since $g(z) = f(z)/\hat{\sigma}^{l-p-2}$, we estimate

$$g^{(r)}(z) = \frac{\partial^r g(z)}{\partial \hat{\beta}_1^{r_1} \cdots \partial \hat{\beta}_p^{r_p} \partial (\hat{\sigma}^2)^{r_{p+1}}},$$
$$\left(0 \leq r_i, \quad i = 1, 2, \ldots, p+1, \quad r = \sum_{i=1}^{p+1} r_i, \quad r \leq k-1 \right)$$

by

$$g_n^{(r)} = g^{(r)}(Z_1, \ldots, Z_n, Z) = \frac{1}{nh^{r+p+1}} \sum_{i=1}^{p+1} \frac{K_r\left(\dfrac{Z_i - Z}{h}\right)}{\hat{\sigma}_i^{l-p-2}} I_{(\hat{\sigma}^2 > 0)} \tag{2.2}$$

where $h > 0$ and $h \to 0$ as $n \to \infty$. Let

$$\psi_n = \hat{\beta} - \frac{l-p}{2} \Sigma q_n(z) = \hat{\beta} - \frac{l-p}{2} \Sigma q_n, \tag{2.3}$$

$$\phi_n = -\frac{l-p}{2} [g_{\hat{\sigma}^2,n}^{(1)} / \hat{g}_{\hat{\sigma}^2,n}^{(2)}]_{n^\nu} \tag{2.4}$$

where

$$q_n = (q_{1n}, \ldots, q_{pn})', \qquad q_{in} = [g_{\hat{\beta}_i,n}^{(1)} / \hat{g}_{\hat{\sigma}^2,n}^{(1)}]_{n^\nu}, \qquad [a]_L = \begin{cases} a & \text{if } |a| \leq L \\ 0 & \text{if } |a| > L \end{cases} \tag{2.5}$$

and $g_{\hat{\beta}_i,n}^{(1)}$ and $g_{\hat{\sigma}^2,n}^{(r)}$, defined by (2.2), is the kernel estimator of $g_{\hat{\beta}_i}^{(1)}$ $(i = 1, \ldots, p)$ and $g_{\hat{\sigma}^2}^{(r)}$ $(r = 1, 2)$, respectively.

We define $\hat{\theta}_n = (\psi_n', \phi_n)'$ as the EB estimator of $\theta = (\beta', \sigma^2)'$. Let E_* and E be the expectation w.r.t. the joint distribution of $(Z_1, \ldots, Z_n; (Z, \theta))$ and $(Z_1, \ldots, Z_n)$ respectively in this paper. Then the "overall" Bayes risk of $\hat{\theta}_n$ is

$$\begin{aligned} R_n = R_n(\hat{\theta}_n, G) &= E_*[\|\psi_n - \beta\|^2/\sigma^2] + E_*[|\phi_n - \sigma^2|^2/\sigma^4] \\ &= R_{1n} + R_{2n}. \end{aligned} \tag{2.6}$$

By definition, if for some $\delta > 0$, $R_n - R(G) = O(n^{-\delta})$, then the convergence rates of $\{\hat{\theta}_n\}$ are $O(n^{-\delta})$.

3. The convergence rates of $\hat{\theta}_n$

In order to get the convergence rates of EB estimators, we need the following lemmas

LEMMA 3.1. *Suppose Y, Y' are random variables, y, y' are real numbers. $L > 0$ is a constant, then for $0 < r \le 2$ we have*

$$E\left|\left[\frac{Y'}{Y} - \frac{y'}{y}\right]_L\right|^r \le 2|y|^{-r}\left\{E|Y' - y'|^r + \left(\left|\frac{y'}{y}\right| + L\right)^r E|Y - y|^r\right\}. \tag{3.1}$$

PROOF. See Lemma 3 of Zhao (1981). It is similar to the Lemma 4.1 of Singh (1979).

LEMMA 3.2. *Suppose that $R(G) < \infty$. Let R_n, R_{1n}, R_{2n} be defined by (2.6) and $R(G)$, $R_1(G)$, $R_2(G)$ be given by (1.18), then we have*

$$R_{1n} - R_1(G) = E_*[\|\psi_n - \psi_G\|^2/\sigma^2], \tag{3.2}$$
$$R_{2n} - R_2(G) = E_*[\|\phi_n - \phi_G\|^2/\sigma^4], \tag{3.3}$$
$$R_n - R(G) = E_*[\|\psi_n - \psi_G\|^2/\sigma^2] + E_*[\|\phi_n - \phi_G\|^2/\sigma^4]. \tag{3.4}$$

PROOF. We only prove (3.2). The proof of the others is analogous.

$$\begin{aligned} R_{1n} &= E_*[\|\psi_n - \beta\|^2/\sigma^2] \\ &= R_1(G) + E_*[\|\psi_n - \psi_G\|^2/\sigma^2] + 2E_*[(\psi_n - \psi_G)'(\psi_G - \beta)/\sigma^2] \end{aligned}$$

where

$$\begin{aligned} &E_*[(\psi_n - \psi_G)'(\psi_G - \beta)/\sigma^2] \\ &\quad = E_{(Z_1,\ldots,Z_n,Z)}\{(\psi_n - \psi_G)'[E(\sigma^{-2}\psi_G \mid Z) - E(\sigma^{-2}\beta \mid Z)]\} \end{aligned}$$

since $\psi_G = E(\sigma^{-2}\beta \mid Z)/E(\sigma^{-2} \mid Z)$ and $E(\sigma^{-2}\psi_G \mid Z) = \psi_G E(\sigma^{-2} \mid Z)$, we get $E_*[(\psi_n - \psi_G)'(\psi_G - \beta)/\sigma^2] = 0$. Thus

$$R_{1n} - R_1(G) = E_*[\|\psi_n - \psi_G\|^2/\sigma^2].$$

This lemma is proved.

LEMMA 3.3. *Let $g_n^{(r)}$ be defined by (2.2) with $h = n^{-1/(2k+p+1)}$, $r = 0, 1, 2$. Suppose $G \in \mathcal{F}_k$ and $0 < \lambda \le 2$, then*

$$E|g_n^{(r)} - g^{(r)}|^\lambda \le c \cdot n^{-\lambda(k-r)/(2k+p+1)}(\hat{\sigma}^{-\lambda(l-p-2)/2} + 1). \tag{3.5}$$

PROOF. Since

$$E|g_n^{(r)} - g^{(r)}|^\lambda \le 2[(\operatorname{Var} g_n^{(r)})^{\lambda/2} + |E(g_n^{(r)}) - g^{(r)}|^\lambda] \\ = 2[I_1^{\lambda/2} + I_2^\lambda]. \tag{3.6}$$

First notice that $Z_1, \ldots, Z_n$ are i.i.d. and $f(z) = 0$ when $\hat\sigma = 0$, then we have

$$I_1 = \operatorname{Var} g_n^{(r)} = \frac{1}{nh^{2(r+p+1)}} \operatorname{Var}\left[\frac{K_r\left(\frac{Z_1 - z}{h}\right)}{\hat\sigma^{l-p-2}} I_{(\hat\sigma_1^2>0)}\right]$$

$$\le \frac{1}{nh^{2(r+p+1)}} \int K_r^2\left(\frac{z_1 - z}{h}\right) f(z_1)/\sigma_1^{2(l-p-2)} dz_1$$

$$\le \frac{1}{nh^{2(r+p+1)}} \int K_r^2\left(\frac{z_1 - z}{h}\right) g(z_1)/\hat\sigma_1^{l-p-2} dz_1.$$

Let $(z_1 - z)/h = u$ (obviously $\hat\sigma_1^2 = \hat\sigma^2 + u_{p+1}h$). Since $0 < u_i < 1$, $i = 1, \ldots, p+1$, $u = (u_1, \ldots, u_{p+1})'$ and $h > 0$, we have

$$I_1 \le \frac{1}{nh^{2r+p+1}} \int K_r^2(u) g(z + hu)(\hat\sigma^2 + u_{p+1}h)^{-(l-p-2)/2} du$$

$$\le \frac{\hat\sigma^{-(l-p-2)}}{nh^{2r+p+1}} \int K_r^2(u) g(z + hu) du.$$

Since $G \in \mathcal{F}_k$, by Lemma 1.1 we know that $|g(z)| \le \alpha$. Take $h = n^{-1/(2k+p+1)}$, then we have

$$I_1 \le cn^{-2(k-r)/(2k+p+1)} \cdot \hat\sigma^{-(l-p-2)}. \tag{3.7}$$

Secondly since

$$Eg_n^{(r)} = \frac{1}{h^{r+p+1}} E\left[\frac{K_r\left(\frac{Z_1 - z}{h}\right)}{\hat\sigma^{l-p-2}} I_{(\hat\sigma^2>0)}\right]$$

$$= \frac{1}{h^{r+p+1}} \int K_r\left(\frac{z_1 - z}{h}\right) g(z_1) dz_1.$$

Let $(z_1 - z)/h = u$, then $Eg_n^{(r)} = \frac{1}{h^r} \int K_r(u) g(z + hu) du$. Since $G \in \mathcal{F}_k$, $g(u)$ has the k-th order mixed continuous partial derivatives and $|g^{(k)}(z)| \le \alpha$. Instead of $g(z + hu)$, we use its k-th order Taylor expansion about z with Lagrange-form of the remainder at the k-th term and make use of the orthogonality properties of $K_r(u)$ and the fact that $K_r(u)$ vanishes outside $(0, 1)$, then

$$I_2 = |E(g_n^{(r)}) - g^{(r)}| = \left|\frac{1}{h^r} \int K_r(u) g(z + hu) du - g^{(r)}\right| \tag{3.8}$$

$$= h^{k-r} \left|\sum_{r_1 + \cdots + r_{p+1} = k} \int \frac{g^{(k)}(z + \xi hu)}{r_1! \cdots r_{p+1}!} K_r(u) u_1^{r_1} \cdots u_{p+1}^{r_{p+1}} du\right|$$

$$\le ch^{k-r} \le cn^{-(k-r)/(2k+p+1)} \qquad r = 1, 2.$$

Substituting (3.7) and (3.8) into (3.6), this lemma is proved.

Let

$$
\begin{aligned}
&\tilde{f}(z) = \int_\Theta f(z \mid \theta)\sigma^{-2} dG(\theta), \qquad \tilde{g}(z) = \int_\Theta g(z \mid \theta)\sigma^{-2} dG(\theta), \\
&\tilde{f}_1(\hat{\beta}) = \int_\Theta f_1(\hat{\beta} \mid \theta)\sigma^{-2} dG(\theta), \qquad \tilde{f}_2(\hat{\sigma}^2) = \int_\Theta f_2(\hat{\sigma}^2 \mid \theta)\sigma^{-2} dG(\theta)
\end{aligned} \tag{3.9}
$$

and

$$
\begin{aligned}
&\bar{f}(z) = \int_\Theta f(z \mid \theta)\sigma^{-4} dG(\theta), \qquad \bar{g}(z) = \int_\Theta g(z \mid \theta)\sigma^{-4} dG(\theta), \\
&\bar{f}_1(\hat{\beta}) = \int_\Theta f_1(\hat{\beta} \mid \theta)\sigma^{-4} dG(\theta), \qquad \bar{f}_2(\hat{\sigma}^2) = \int_\Theta f_2(\hat{\sigma}^2 \mid \theta)\sigma^{-4} dG(\theta)
\end{aligned} \tag{3.10}
$$

where $f(z \mid \theta)$, $f_1(\hat{\beta} \mid \theta)$, $f_2(\hat{\sigma}^2 \mid \theta)$ are defined by (1.8) and $g(z \mid \theta)$ is defined by (1.10). From the fact that $f(z \mid \theta) = \hat{\sigma}^{(l-p-2)} g(z \mid \theta)$, it is obvious that $\tilde{f}(z) = \hat{\sigma}^{(l-p-2)}\tilde{g}(z)$ and $\bar{f}(z) = \hat{\sigma}^{(l-p-2)}\bar{g}(z)$. Correspondingly we have $\tilde{E}(\cdot)$, $\bar{E}(\cdot)$ by substituting $\sigma^{-2}dG$ and $\sigma^{-4}dG$ for dG in the expression of $E_*(\cdot)$ respectively. Similarly we can define $\tilde{E}_1(\cdot)$ and $\tilde{E}_2(\cdot)$ and etc., then we have the following lemma.

LEMMA 3.4. *Suppose that $G \in \mathcal{F}_k$, $\frac{1}{3} < \lambda < 1$ and $\epsilon > 0$ is an arbitrarily small number, $\xi = p - 1 + \epsilon$, for $d = 2$ or $2 - \lambda$ we have*

(i) *If $E_*[\sigma^{-2}\|\beta\|^{(1+\xi)(1+\lambda)/(1-\lambda)}] < \infty$ and $E_*[\sigma^{(2\xi(1+\lambda)+2\lambda(l-p))/(1-\lambda)}] < \infty$ then*

$$\int \hat{\sigma}^{d(l-p-2)/2}[\tilde{g}(z)]^{1-\lambda} dz < \infty,$$

(ii) *If $E_*[\sigma^{-4}\|\beta\|^{(1+\xi)(1+\lambda)/(1-\lambda)}] < \infty$ and $E_*[\sigma^{(2\xi(1+\lambda)+2\lambda(l-p+1)-2)/(1-\lambda)}] < \infty$ then*

$$\int \hat{\sigma}^{d(l-p-2)/2}[\bar{g}(z)]^{1-\lambda} dz < \infty.$$

PROOF. Since the conditional distribution of $(l-p)\hat{\sigma}^2/\sigma^2$ and $\hat{\beta}$ with give $\theta = (\beta', \sigma^2)'$ is χ^2_{l-p} and $N(\beta, \sigma^2\Sigma)$ respectively, therefore we have

$$E(\hat{\sigma}^a \mid \theta) \le c\sigma^a \qquad \text{for} \quad a > -(l-p) \tag{3.11}$$

by the fact that $\int_{-\infty}^{\infty} |x|^r \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \le c\sigma^r$, we get

$$
\begin{aligned}
E[\|\hat{\beta} - \beta\|^r \mid \theta] &= \int \|\hat{\beta} - \beta\|^r c\sigma^{-p} \exp\left\{-\frac{1}{2\sigma^2}(\hat{\beta} - \beta)'\Sigma^{-1}(\hat{\beta} - \beta)\right\} d\hat{\beta} \\
&\le \int \|\hat{\beta} - \beta\|^r c\sigma^{-p} \exp\left\{-\frac{\lambda_0}{2\sigma^2}\|\hat{\beta} - \beta\|^2\right\} d\hat{\beta} \\
&\le c\sigma^r \qquad \text{for} \quad r > 0
\end{aligned} \tag{3.12}
$$

where $0 < \lambda_0 = \min(\text{root}\,\Sigma^{-1})$.

By Hölder's inequality and from (1.8) and (3.9) we have

$$\tag{3.13} \begin{aligned} \tilde{f}(z) &= \int_\Theta f(z \mid \theta)\sigma^{-2}dG(\theta) = \int_\Theta [\sigma^{-1} f_1(\hat{\beta} \mid \theta)][\sigma^{-1} f_2(\hat{\sigma}^2 \mid \theta)]dG(\theta) \\ &\le \left[\int_\Theta f_1^2(\hat{\beta} \mid \theta)\sigma^{-2}dG(\theta)\right]^{1/2} \left[\int_\Theta f_2^2(\hat{\sigma}^2 \mid \theta)\sigma^{-2}dG(\theta)\right]^{1/2} \\ &\le c[\tilde{f}_1(\hat{\beta})]^{1/2}[\tilde{f}_2(\hat{\sigma}^2)]^{1/2}. \end{aligned}$$

Similarly we have

$$\tag{3.14} \bar{f}(z) \le c[\bar{f}_1(\hat{\beta})]^{1/2}[\bar{f}_2(\hat{\sigma}^2)]^{1/2}.$$

Using the above facts we can prove this lemma as follows.

Let $I = \int \hat{\sigma}^{d(l-p-2)/2}[\tilde{g}(z)]^{1-\lambda}dz = \int \hat{\sigma}^{(d+2\lambda-2)(l-p-2)/2}[\tilde{f}(z)]^{1-\lambda}dz$, $r = (d + 2\lambda - 2)(l - p - 2)/2$, and by (3.13) we have

$$\begin{aligned} I &\le c \iint \hat{\sigma}^r [\tilde{f}_1(\hat{\beta})]^{(1-\lambda)/2} \cdot [\tilde{f}(\hat{\sigma}^2)]^{(1-\lambda)/2}dz \\ &= c \int [\tilde{f}_1(\hat{\beta})]^{(1-\lambda)/2}d\hat{\beta} \cdot \int \hat{\sigma}^r [\tilde{f}_2(\hat{\sigma}^2)]^{(1-\lambda)/2}d\hat{\sigma}^2 = cI_1 \cdot I_2 \end{aligned}$$

where

$$\begin{aligned} I_1 &= \int [\tilde{f}_1(\hat{\beta})]^{(1-\lambda)/2}d\hat{\beta} \\ &= \int_{\|\hat{\beta}\| \le p} [\tilde{f}_1(\hat{\beta})]^{(1-\lambda)/2}d\hat{\beta} + \int_{\|\hat{\beta}\| > p} [\tilde{f}_1(\hat{\beta})]^{(1-\lambda)/2}d\hat{\beta} = I_{11} + I_{12} \end{aligned}$$

obviously $I_{11} < \infty$ and by Hölder's inequality we get

$$\begin{aligned} I_{12} &= \int_{\|\hat{\beta}\| > p} [\|\hat{\beta}\|^{-((1+\xi)(1+\lambda))/2}] \cdot [\|\hat{\beta}\|^{(1+\xi)(1+\lambda)/2} \cdot (\tilde{f}_1(\hat{\beta}))^{(1-\lambda)/2}]d\hat{\beta} \\ &\le \left[\int_{\|\hat{\beta}\| > p} \|\hat{\beta}\|^{-(1+\xi)}d\hat{\beta}\right]^{(1+\lambda)/2} \left[\int_{\|\hat{\beta}\| > p} \|\hat{\beta}\|^{(1+\xi)(1+\lambda)/(1-\lambda)} \cdot \tilde{f}_1(\hat{\beta})d\hat{\beta}\right]^{(1-\lambda)/2} \\ &= [I_{12}^{(1)}]^{(1+\lambda)/2} \cdot [I_{12}^{(2)}]^{(1-\lambda)/2}. \end{aligned}$$

It is obvious that $I_{12}^{(1)} < \infty$ and by (3.12) we have

$$\begin{aligned} I_{12}^{(2)} &= \int_{\|\hat{\beta}\| > p} \|\hat{\beta}\|^{(1+\xi)(1+\lambda)/(1-\lambda)} \tilde{f}_1(\hat{\beta})d\hat{\beta} = \tilde{E}_1[\|\hat{\beta}\|^{(1+\xi)(1+\lambda)/(1-\lambda)} I_{[\|\hat{\beta}\| > p]}] \\ &\le c\tilde{E}_{(\theta)}[E_{(\hat{\beta}|\theta)}(\|\hat{\beta} - \beta\|^{(1+\xi)(1+\lambda)/(1-\lambda)})] + c\tilde{E}_{(\theta)}[\|\beta\|^{(1+\xi)(1+\lambda)/(1-\lambda)}] \\ &\le cE_*[\sigma^{(\xi(1+\lambda)+3\lambda-1)/(1-\lambda)}] + cE_*[\sigma^{-2}\|\beta\|^{(1+\xi)(1+\lambda)/(1-\lambda)}] < \infty. \end{aligned}$$

Therefore $I_{12} \leq [I_{12}^{(1)}]^{(1+\lambda)/2} \cdot [I_{12}^{(2)}]^{(1-\lambda)/2} < \infty$. Thus

$$I_1 = I_{11} + I_{12} < \infty \tag{3.15}$$

and

$$\begin{aligned} I_2 &= \int \hat{\sigma}^r [\tilde{f}_2(\hat{\sigma}^2)]^{(1-\lambda)/2} d\hat{\sigma}^2 \\ &= \int_{\hat{\sigma}^2 \leq 1} \hat{\sigma}^r [\tilde{f}_2(\hat{\sigma}^2)]^{(1-\lambda)/2} d\hat{\sigma}^2 + \int_{\hat{\sigma}^2 > 1} \hat{\sigma}^r [\tilde{f}_2(\hat{\sigma}^2)]^{(1-\lambda)/2} d\hat{\sigma}^2 \\ &= I_{21} + I_{22}. \end{aligned}$$

It is easy to see that

$$I_{21} < \infty$$

and by Hölder's inequality we have

$$\begin{aligned} I_{22} &= \int_{\hat{\sigma}^2 > 1} [(\hat{\sigma}^2)^{-((1+\epsilon)(1+\lambda))/2}] \cdot [\hat{\sigma}^{(1+\epsilon)(1+\lambda)+r} (\tilde{f}_2(\hat{\sigma}^2))^{(1-\lambda)/2}] d\hat{\sigma}^2 \\ &\leq \left[\int_{\hat{\sigma}^2 > 1} (\hat{\sigma}^2)^{-(1+\epsilon)} d\hat{\sigma}^2 \right]^{(1+\lambda)/2} \\ &\quad \cdot \left[\int_{\hat{\sigma}^2 > 1} \hat{\sigma}^{2(1+\epsilon)(1+\lambda)/(1-\lambda)+2r/(1-\lambda)} \cdot \tilde{f}_2(\hat{\sigma}^2) d\hat{\sigma}^2 \right]^{(1-\lambda)/2} \\ &= [I_{22}^{(1)}]^{(1+\lambda)/2} \cdot [I_{22}^{(2)}]^{(1-\lambda)/2}. \end{aligned}$$

It is obvious that $I_{22}^{(1)} < \infty$ and by (3.11) we have

$$\begin{aligned} I_{22}^{(2)} &= \tilde{E}_2[\hat{\sigma}^{2(1+\epsilon)(1+\lambda)/(1-\lambda)+2r/(1-\lambda)}] \leq \tilde{E}_{(\theta)}[E_{(\hat{\sigma}|\theta)}(\hat{\sigma}^{2(1+\epsilon)(1+\lambda)/(1-\lambda)+2r/(1-\lambda)})] \\ &\leq c\tilde{E}_{(\theta)}[\sigma^{2(1+\epsilon)(1+\lambda)/(1-\lambda)+2r/(1-\lambda)}] = cE_*[\sigma^{2(1+\epsilon)(1+\lambda)/(1-\lambda)+2r/(1-\lambda)-2}] \\ &\leq cE_*[\sigma^{(2\epsilon(1+\lambda)+2\lambda(l-p))/(1-\lambda)}] < \infty \end{aligned}$$

the last two inequalities are held since $d = 2$ or $2 - \lambda$, which implies $d \leq 2$, and $r = (d + 2\lambda - 2)(l - p - 2)/2 \leq \lambda(l - p - 2)$. Therefore

$$I_{22} \leq [I_{22}^{(1)}]^{(1+\lambda)/2} \cdot [I_{22}^{(2)}]^{(1-\lambda)/2} < \infty.$$

Thus

$$I_2 = I_{21} + I_{22} < \infty. \tag{3.16}$$

From (3.15) and (3.16) we know that

$$I = \int \hat{\sigma}^{d(l-p-2)/2} (\tilde{g}(z))^{1-\lambda} dz \leq cI_1 \cdot I_2 < \infty.$$

The part (i) of this lemma is proved. Similarly we can prove (ii) of this lemma.

THEOREM 3.1. *Let ψ_n and ϕ_n be defined by* (2.3), (2.4) *with* $h = n^{-1/(2k+p+1)}$ *and* $\nu = (k-2)/[2k(2k+p+1)]$, ψ_G *and* ϕ_G *are given by* (1.16) *and* (1.17). *If the following conditions are held*:

(i) $G \in \mathcal{F}_k$,

(ii) $E_*[\sigma^{-2}\|\beta\|^r] < \infty$, $E_*[\sigma^{-4}\|\beta\|^r] < \infty$ *with* $r = [\frac{(1+\xi)(1+\lambda)}{1-\lambda}] \vee (2k\lambda)$,

(iii) $E_*(\sigma^\tau) < \infty$, *with* $\tau = [\frac{2\xi(1+\lambda)+2\lambda(l-p)}{1-\lambda}] \vee (4k\lambda - 4)$

where $\frac{1}{3} < \lambda < 1$, $\xi = p - 1 + \epsilon$, $\epsilon > 0$ *is an arbitrarily small number and* $k \geq 3$ *is a given positive integer, then*

$$\begin{aligned} R_n - R(G) &= E_*\left[\frac{\|\psi_n - \psi_G\|^2}{\sigma^2}\right] + E_*\left[\frac{(\psi_n - \phi_G)^2}{\sigma^4}\right] \\ &= O(n^{-((\lambda k-1)(k-2))/k(2k+p+1)}). \end{aligned}$$

PROOF. By Lemma 3.2 and (1.16), (1.17), (2.3) and (2.4) we have

$$\begin{aligned} R_n - R(G) &\leq c\{E_*\|\Sigma(q_n(z) - q(z))\|^2/\sigma^2] + E_*[(\phi_n - \phi_G)^2/\sigma^4]\} \\ &\leq c\left\{\bar{\lambda}^2 \sum_{i=1}^p E_*(q_{in} - q_i)^2/\sigma^2] + E_*[(q_{p+1,n} - q_{p+1})^2/\sigma^4]\right\} \\ &\leq c\left\{\sum_{i=1}^p E_*[(q_{in} - q_i)^2/\sigma^2] + E_*[(q_{p+1,n} - q_{p+1})^2/\sigma^4]\right\} \\ &= c\left(\sum_{i=1}^p I_i + J\right) \end{aligned}$$

where $\bar{\lambda} = \max(\text{root}\,\Sigma)$, $q_{p+1,n} = g^{(1)}_{\hat{\sigma}^2,n}(z)/g^{(2)}_{\hat{\sigma}^2,n}(z)$, $q_{p+1} = g^{(1)}_{\hat{\sigma}^2}(z)/g^{(2)}_{\hat{\sigma}^2}(z)$ and q_{in}, q_i are defined by (2.5) and (1.16) respectively.

First consider $I_1 = E_*\left[(q_{1n} - q_1)^2/\sigma^2\right] = E_{(z,\theta)}\left\{\sigma^{-2}\left[E(q_{1n} - q_1)^2\right]\right\}$. Suppose $A_{1n} = \{z = (\hat{\beta}', \hat{\sigma}^2)' \in \mathcal{Z}, |q_1(z)| \leq \frac{1}{2}n^\nu\}$, $B_{1n} = \mathcal{Z} - A_{1n}$. If $z \in A_{1n}$ then $|q_{1n} - q_1| \leq \frac{3}{2}n^\nu$, therefore by Lemmas 3.1 and 3.3 we have

$$\begin{aligned} E(q_{1n} - q_1)^2 &\leq \left(\frac{3}{2}n^\nu\right)^{2-\lambda} \cdot E\left|\left[\frac{g^{(1)}_{\hat{\beta}_1,n}(z)}{g^{(1)}_{\hat{\sigma}^2,n}(z)} - \frac{g^{(1)}_{\hat{\beta}_1}(z)}{g^{(1)}_{\hat{\sigma}^2}(z)}\right]_{(3/2)n^\nu}\right|^\lambda \\ &\leq cn^{\nu(2-\lambda)}|g^{(1)}_{\hat{\sigma}^2}(z)|^{-\lambda}\{E|g^{(1)}_{\hat{\beta}_1,n}(z) - g^{(1)}_{\hat{\beta}_1}(z)|^\lambda + (2n^\nu)^\lambda E|g^{(1)}_{\hat{\sigma}^2,n}(z) - g^{(1)}_{\hat{\sigma}^2}(z)|^\lambda\} \\ &\leq cn^{-[\lambda(k-1)/(2k+p+1)-2\nu]} \cdot |g^{(1)}_{\hat{\sigma}^2}(z)|^{-\lambda}(\hat{\sigma}^{-(\lambda(l-p-2))/2} + 1). \end{aligned}$$

Since $g^{(1)}_{\hat{\sigma}^2}(z) = c\int_\Theta \sigma^{-2}g(z \mid \theta)dG(\theta) = c\tilde{g}(z)$ and $\tilde{f}(z) = \hat{\sigma}^{l-p-2} \cdot \tilde{g}(z)$, therefore

$$\begin{aligned} E_*[\sigma^{-2}|g^{(1)}_{\hat{\sigma}^2}(z)|^{-\lambda}\hat{\sigma}^{-(\lambda(l-p-2))/2}] &= \tilde{E}[|g^{(1)}_{\hat{\sigma}^2}(z)|^{-\lambda}\hat{\sigma}^{-(\lambda(l-p-2))/2}] \\ &= c_1 \int \hat{\sigma}^{(2-\lambda)(l-p-2)/2}(\tilde{g}(z))^{1-\lambda}dz \end{aligned}$$

and

$$\begin{aligned}E_*[\sigma^{-2}|g^{(1)}_{\hat\sigma^2}(z)|^{-\lambda}] &= \tilde E[|g^{(1)}_{\hat\sigma^2}(z)|^{-\lambda}]\\ &\le c_2\int(\tilde g(z))^{-\lambda}\cdot\tilde f(z)dz = c_2\int\hat\sigma^{l-p-2}(\tilde g(z))^{1-\lambda}dz.\end{aligned}$$

By Lemma 3.4 and the conditions (ii) and (iii) of theorem 3.1, we have

$$\begin{aligned}(3.17)\quad & E_*\{[\sigma^{-2}E(q_{1n}-q_1)^2]I_{[z\in A_{1n}]}\}\\ &\le cn^{-[\lambda(k-1)/(2k+p+1)-2\nu]}E_*[\sigma^{-2}|g^{(1)}_{\hat\sigma^2}(z)|^{-\lambda}(\hat\sigma^{-\lambda(l-p-2)/2}+1)]\\ &\le cn^{-[\lambda(k-1)/(2k+p+1)-2\nu]}\\ &\quad\cdot\left[\int\hat\sigma^{(2-\lambda)(l-p-2)/2}(\tilde g(z))^{1-\lambda}dz+\int\hat\sigma^{l-p-2}(\tilde g(z))^{1-\lambda}dz\right]\\ &\le cn^{-[\lambda(k-1)/(2k+p+1)-2\nu]}.\end{aligned}$$

If $z\in B_{1n}$ then $|q_1(z)|>\frac12 n^\nu$, therefore we get $(q_{1n}-q_1)^2\le 2q_1^2+2n^{2\nu}\le 10q_1^2$, hence by Hölder's and Markov's inequalities we have

$$\begin{aligned}(3.18)\quad & E_*\{\sigma^{-2}E(q_{1n}-q_1)^2I_{[z\in B_{1n}]}\}\\ &\le 10E_*\{\sigma^{-2}q_1^2(z)I_{[|q_1(z)|>n^\nu/2]}\}\\ &= 10E_*\{[q_1^2(z)\sigma^{-4/\delta}]\cdot[\sigma^{-(2\delta-4)/\delta}I_{[|q_1(z)|>n^\nu/2]}]\}\\ &\le 10\{E_*(|q_1(z)|^\delta\sigma^{-2})\}^{2/\delta}\cdot\{E_*(\sigma^{-2}I_{[|q_1(z)|>n^\nu/2]})\}^{(\delta-2)/\delta}\\ &\le 10\{\tilde E[\|q(z)\|^\delta]\}^{2/\delta}\cdot\{2^\delta n^{-\delta\nu}\tilde E(\|q(z)\|^\delta)\}^{(\delta-2)/\delta}\end{aligned}$$

where $\delta>2$. By (1.6) and (1.16) we get

$$\begin{aligned}\|q(z)\|^2 &= \left(\frac{2}{l-p}\right)^2(\hat\beta-\psi_G)'\Sigma^{-2}(\hat\beta-\psi_G)\\ &\le c\lambda_*^2\|\hat\beta-\psi_G\|^2\le c[\|\hat\beta\|^2+\|\psi_G\|^2],\\ \psi_G &= \frac{E(\sigma^{-2}\beta\mid z)}{E(\sigma^{-2}\mid z)}=\tilde E(\beta\mid z)\end{aligned}$$

where $\lambda_*=\max(\mathrm{root}\,\Sigma^{-1})$, therefore by Jensen's and C_r-inequality, formula (3.12) and the conditions of Theorem 3.1, for $\delta>2$ we have

$$\begin{aligned}(3.19)\quad \tilde E\|q(z)\|^\delta &\le c[\tilde E\|\hat\beta\|^\delta+\tilde E\|\psi_G\|^\delta]\\ &\le c[(\tilde E\|\hat\beta-\beta\|^\delta+\tilde E\|\beta\|^\delta)+\tilde E\|\tilde E(\beta\mid z)\|^\delta]\\ &\le c[E_*(\sigma^{\delta-2})+E_*(\sigma^{-2}\|\beta\|^\delta)]<\infty.\end{aligned}$$

Substituting (3.19) into (3.18) we have

$$(3.20)\qquad E_*\{\sigma^{-2}E(q_{1n}-q_1)^2I_{[z\in B_{1n}]}\}\le cn^{-\nu(\delta-2)}.$$

Let $\delta=2\lambda k$ with $\frac13<\lambda<1$, $k\ge3$, then $\delta>2$. Since $\nu=\frac{k-2}{2k(2k+p+1)}$, then from (3.17) and (3.20) we have

$$\begin{aligned}(3.21)\quad I_1 = E_*[(q_{1n}-q_1)^2/\sigma^2] &= E_*\{\sigma^{-2}E(q_{1n}-q_1)^2I_{[z\in A_{1n}]}\}\\ &\quad+E_*\{\sigma^{-2}E(q_{1n}-q_1)^2I_{[z\in B_{1n}]}\}\\ &\le cn^{-((k\lambda-1)(k-2))/k(2k+p+1)}.\end{aligned}$$

Similar to the proof of I_1, we can obtain

$$I_i = E_*[(q_{in} - q_i)^2/\sigma^2] \leq cn^{-((\lambda k-1)(k-2))/k(2k+p+1))}, \qquad i = 2, \ldots, p. \tag{3.22}$$

Secondly we consider the J,

$$J = E_*[(q_{p+1,n} - q_{p+1})^2/\sigma^4] = E_{(Z,\theta)}\{\sigma^{-4}[(q_{p+1,n} - q_{p+1})^2]\}.$$

Similar to the proof of I_1, let $C_{1n} = \{z = (\hat{\beta}', \hat{\sigma}^2)' \in \mathcal{Z}, |q_{p+1,n}(z)| \leq \frac{1}{2}n^\nu\}$, and $D_{1n} = \mathcal{Z} - C_{1n}$. If $z \in C_{1n}$, then $|q_{p+1,n} - q_{p+1}| \leq \frac{3}{2}n^\nu$, therefore by Lemmas 3.1 and 3.3 we have

$$\begin{aligned} &E(q_{p+1,n} - q_{p+1})^2 \\ &\quad \leq cn^{-[\lambda(k-2)/(2k+p+1)-2\nu]} \cdot |g_{\hat{\sigma}^2}^{(2)}(z)|^{-\lambda}(\hat{\sigma}^{-(\lambda(l-p-2))/2} + 1). \end{aligned}$$

Since $g_{\hat{\sigma}^2}^{(2)}(z) = c\int_\Theta \sigma^{-4} g(z \mid \theta) dG(\theta) = c\bar{g}(z)$ and $\bar{f}(z) = \hat{\sigma}^{l-p-2}\bar{g}(z)$, therefore

$$E_*[\sigma^{-4}|g_{\hat{\sigma}^2}^{(2)}(z)|^{-\lambda} \cdot \hat{\sigma}^{-(\lambda(l-p-2))/2}] \leq c_1 \int \hat{\sigma}^{(2-\lambda)(l-p-2)/2}(\bar{g}(z))^{1-\lambda} dz$$

and

$$E_*[\sigma^{-4}|g_{\hat{\sigma}^2}^{(2)}(z)|^{-\lambda}] \leq c_2 \int \hat{\sigma}^{l-p-2}(\bar{g}(z))^{1-\lambda} dz.$$

By Lemma 3.4 and the conditions (ii) and (iii) of Theorem 3.1, we have

$$\begin{aligned} (3.17')\quad &E_*\{\sigma^{-4}E(q_{p+1,n} - q_{p+1})^2 \cdot I_{[z \in C_{1n}]}\} \\ &\leq cn^{-[\lambda(k-2)/(2k+p+1)-2\nu]} \\ &\quad \cdot E_*[\sigma^{-4} \cdot |g_{\hat{\sigma}^2}^{(2)}(z)|^{-\lambda}(\hat{\sigma}^{-\lambda(l-p-2)/2} + 1)] \\ &\leq cn^{-[\lambda(k-2)/(2k+p+1)-2\nu]} \\ &\quad \cdot \left[\int \hat{\sigma}^{(2-\lambda)(l-p-2)/2}(\bar{g}(z))^{1-\lambda} dz + \int \hat{\sigma}^{l-p-2}(\bar{g}(z))^{1-\lambda} dz\right] \\ &\leq cn^{-[\lambda(k-2)/(2k+p+1)-2\nu]}. \end{aligned}$$

If $z \in D_{1n}$, then $(q_{p+1,n} - q_{p+1})^2 \leq 10q_{p+1}^2(z)$, and by (1.7) and (1.17) we have

$$|q_{p+1}(z)| = \frac{2}{l-p} \cdot \frac{E(\sigma^{-2} \mid z)}{E(\sigma^{-4} \mid z)} = \frac{2}{l-p}\bar{E}(\sigma^2 \mid z)$$

then by Jensen's inequality and the condition (iii) of Theorem 3.1, for $\delta > 2$, we obtain

$$\bar{E}|q_{p+1}(z)|^\delta = \bar{E}\left[\frac{2}{l-p}\bar{E}(\sigma^2 \mid z)\right]^\delta \leq c\bar{E}(\sigma^{2\delta}) = cE_*(\sigma^{2\delta-4}) < \infty. \tag{3.19'}$$

Similar to the proof of the first part of this theorem and by (3.19'), for $\delta > 2$ we have

$$(3.20') \quad \begin{aligned} &E_*\{\sigma^{-4}\cdot E(q_{p+1,n}-q_{p+1})^2 I_{[z\in D_{1n}]}\}\\ &\le 10E_*\{\sigma^{-4}q_{p+1}^2(z)I_{[|q_{p+1}(z)|>n^\nu/2]}\}\\ &= 10E_*\{[\sigma^{-8/\delta}q_{p+1}^2(z)]\cdot[\sigma^{-(4\delta-8)/\delta}I_{[|q_{p+1}(z)|>n^\nu/2]}]\}\\ &\le c\{\bar{E}(|q_{p+1}(z)|^\delta)\}^{2/\delta}\cdot\{2^\delta n^{-\delta\nu}\bar{E}(|q_{p+1}(z)|^\delta)\}^{(\delta-2)/\delta}\\ &\le cn^{-\nu(\delta-2)}. \end{aligned}$$

Let $\delta = 2\lambda k$ with $\frac{1}{3} < \lambda < 1$, $k \ge 3$, then $\delta > 2$. Since $\nu = \frac{k-2}{2k(2k+p+1)}$ such that $\frac{\lambda(k-2)}{2k+p+1} - 2\nu = \nu(\delta-2) = \nu(2\lambda k - 2)$, then from (3.17') and (3.20') we have

$$(3.23) \quad \begin{aligned} J = E_*[(q_{p+1,n}-q_{p+1})^2/\sigma^4] &= E_*[\sigma^{-4}E(q_{p+1,n}-q_{p+1})^2 I_{[z\in C_{1n}]}]\\ &+ E_*\{\sigma^{-4}E(q_{p+1,n}-q_{p+1})^2 I_{[z\in D_{1n}]}\}\\ &\le cn^{-((\lambda k-1)(k-2))/k(2k+p+1)}. \end{aligned}$$

From (3.21), (3.22) and (3.23), we have

$$R_n - R_G = c\left(\sum_{i=1}^p I_i + J\right) \le cn^{-((\lambda k-1)(k-2))/k(2k+p+1)}.$$

This theorem is proved.

Corollary 3.1. *Let ψ_G and ψ_n be defined by* (1.16) *and* (2.3) *with $h = n^{-1/(2k+p+1)}$ and $\nu = \frac{k-1}{2k(2k+p+1)}$. If the conditions* (i), (ii) *and* (iii) *with $\tau = [\frac{2\xi(1+\lambda)+2\lambda(l-p)}{1-\lambda}] \vee (2\lambda k-2)$ are held, then the convergence rates of* EB *estimators of the regression coefficient β, is*

$$R_{1n} - R_1(G) = E_*[\|\psi_n - \psi_G\|^2/\sigma^2] = O(n^{-((\lambda k-1)(k-1))/k(2k+p+1)}).$$

Proof. In the proof of the first part in Theorem 3.1, take $\nu = \frac{k-1}{2k(2k+p+1)}$ instead of $\nu = \frac{k-2}{2k(2k+p+1)}$, which satisfies $\frac{\lambda(k-1)}{2k+p+1} - 2\nu = \nu(\delta-2) = \nu(2\lambda k-2)$, then from (3.17) and (3.20) we get the conclusion.

4. An example

Consider the following model $Y_{l\times 1} = X_{l\times p}\beta_{p\times 1} + \epsilon_{l\times 1}$, where $\epsilon \sim N_l(0, \sigma^2 I)$. Suppose that the prior distribution has the form $G(\theta) = G_1(\beta \mid \sigma^2)\cdot G_2(\sigma^2)$, where $dG_1(\beta \mid \sigma^2) = c\sigma^{-p}\exp\{-\frac{1}{2\sigma^2}\beta'\beta\}d\beta$ and $dG_2(\sigma^2) = c(\sigma^{-2})^{\rho+1}\cdot\exp\{-\frac{m}{\sigma^2}\}d\sigma^2$ with $m > 0$, $\rho > \frac{\tau}{2}\vee(\frac{r}{2}-1)$, r and τ appeared in Theorem 3.1, then the distribution density of $\theta = (\beta', \sigma^2)'$ is

$$dG(\theta) = c\sigma^{-(2\rho+p+2)}\exp\left\{-\frac{m}{\sigma^2}\right\}\cdot\exp\left\{-\frac{1}{2\sigma^2}\beta'\beta\right\}d\beta d\sigma^2.$$

We can verify that the conditions of Theorem 3.1 are satisfied by simple calculation as follows.

(i) For any fixed integer $k > 0$

$$\begin{aligned}\int_\Theta \sigma^{-(2k+l)} dG(\theta) &= \int_\Theta \sigma^{-(2k+l+2\rho+p+2)} \exp\left\{-\frac{m}{\sigma^2}\right\} \cdot \exp\left\{-\frac{1}{2\sigma^2}\beta'\beta\right\} d\beta d\sigma^2 \\ &\le c\int_0^\infty y^{(k+\rho+l/2)-1} e^{-my} dy < \infty,\end{aligned}$$

(ii)

$$\begin{aligned}\int_\Theta \sigma^\tau dG(\theta) &= \int_\Theta \sigma^{\tau-(2\rho+p+2)} \exp\left\{-\frac{m}{\sigma^2}\right\} \cdot \exp\left\{-\frac{1}{2\sigma^2}\beta'\beta\right\} d\beta d\sigma^2 \\ &\le c\int_0^\infty y^{(\rho-\tau/2)-1} e^{-my} dy < \infty,\end{aligned}$$

(iii)

$$\begin{aligned}\int_\Theta \sigma^{-2}\|\beta\|^r dG(\theta) &= \int_\Theta \sigma^{-(2\rho+p+4)}\|\beta\|^r \exp\left\{-\frac{m}{\sigma^2}\right\} \cdot \exp\left\{-\frac{1}{2\sigma^2}\beta'\beta\right\} d\beta d\sigma^2 \\ &= c\int_0^\infty \sigma^{-(2\rho+4)} \exp\left\{-\frac{m}{\sigma^2}\right\} \\ &\quad \cdot \left[\int c_1\|\beta\|^r \sigma^{-p} \exp\left\{-\frac{1}{2\sigma^2}\beta'\beta\right\} d\beta\right] d\sigma^2 \\ &\le c\int_0^\infty y^{(\rho-r/2+1)-1} e^{-my} dy < \infty.\end{aligned}$$

Similarly

$$\int_\Theta \sigma^{-4}\|\beta\|^r dG(\theta) \le c\int_0^\infty y^{(\rho-r/2+2)-1} e^{-my} dy < \infty.$$

Therefore the conditions of Theorem 3.1 are satisfied. This example indicates that there exists the prior distribution $G(\theta)$ satifying the conditions of Theorem 3.1.

References

Chang, S. and Wei, L. (1994). The asymptotically optimal empirical Bayes estimation in multiple linear regression model, *Appl. Math.—A Journal of Chinese Universities*, **9**:B, 245–258.

Chen, X. (1983). Asymptotically optimal empirical Bayes estimation for parameter of one-dimensional discrete exponential families, *Chinese Ann. Math. Ser. B*, **4**(1), 41–50.

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

Robbins, H. (1955). An empirical Bayes approach to statistics, *Proc. Third Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 157–163, Univ. California Press, Berkeley.

Singh, R. S. (1976). Empirical Bayes estimation with convergence rates in non-continuous Lebesgue exponential families, *Ann. Statist.*, **4**, 431–439.

Singh, R. S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate, *Ann. Statist.*, **7**, 890–902.

Singh, R. S.(1985). Empirical Bayes estimation in a multiple linear regression model, *Ann. Inst. Statist. Math.*, **37**, 71–86.

Singh, R. S. and Wei, L. (1992). Empirical Bayes with rates and best rates of convergence in $u(x)c(\theta)\exp\{-x/\theta\}$-family: estimation case, *Ann. Inst. Statist. Math.*, **44**, 435–449.

Tao, B. (1986). Asymptotically optimal empirical Bayes estimators for the parameters of normal distribution family, *J. Math. Res. Exposition*, **6**, 157–162.

Wei, L. (1985). Asymptotically optimal empirical Bayes estimation of the parameters about continuous type multi-parameter exponential families, *Chinese J. Appl. Probab. Statist.*, **1**, 127–133.

Wei, L. (1987). Empirical Bayes estimation of parameters with convergence rates about continous type multi-parameter exponential families, *Acta Math. Sinica*, **30**, 272–279.

Wei, L. (1990). Empirical Bayes test of regression coefficient in a multiple linear regression model, *Acta Math. Appl. Sinica*, **6**, 251–262.

Yang, Y. and Wei, L. (1993). Asymptotically optimal empirical Bayes estimation for the parameters of multi-parameter discrete exponential family (to appear).

Yang, Y. and Wei, L. (1994). The convergence rates of empirical Bayes estimation for the parameters of multi-parameter discrete exponential family (to appear).

Zhang, P. (1985). Convergence rate of the asymptotic optimal empirical Bayes estimator of the parameters in normal distribution family, *J. Systems Sci. Math. Sci.*, **5**, 185–191.

Zhao, L. (1981). Empirical Bayes estimation with convergence rates about a class of discrete distribution families, *J. Math. Res. Exposition*, **1**, 59–69.