This is to certify that the

dissertation entitled

Strong Consistency And Bahadur Type Expansions Of A Class Of Minimum Distance Estimators In Linear Regression

presented by

Zhiwei Zhu

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Statistics

H. Koul
Major professor

Date August 17, 1993

STRONG CONSISTENCY AND BAHADUR TYPE EXPANSIONS OF A CLASS OF MINIMUM DISTANCE ESTIMATORS IN LINEAR REGRESSION

By

Zhiwei Zhu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability
1993

ABSTRACT

STRONG CONSISTENCY AND BAHADUR TYPE EXPANSIONS OF A CLASS OF MINIMUM DISTANCE ESTIMATORS IN LINEAR REGRESSION

by

Zhiwei Zhu

Let $p \geq 1$ be an integer, F be a distribution function (d.f.) on the real line R and $\{\varepsilon_i\}$, $1 \leq i \leq n$, be independent and identically distributed (i.i.d.) F random variables (r.v.'s). Consider the linear regression model

$$Y_{ni} = \boldsymbol{x}'_{ni}\boldsymbol{\beta} + \varepsilon_i, \qquad 1 \leq i \leq n,$$

where $\boldsymbol{x}'_{ni}$ is the ith row of the known $n \times p$ design matrix $\boldsymbol{X}_n$, $1 \leq i \leq n$, and $\boldsymbol{\beta}$ is the regression parameter vector of interest of dimension $p \times 1$.

For a nondecreasing right continuous function H from R to R, *Koul & DeWet* (1983) defined a minimum distance estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ as

$$\tilde{\boldsymbol{\beta}} = \text{argmin}_b M(\boldsymbol{b}),$$

where, for $\boldsymbol{b} \in R^p$,

$$T(\boldsymbol{b}) = \int \| \sum_{i=1}^{n} \boldsymbol{x}'_{ni}\{I(Y_{ni} \leq y + \boldsymbol{x}'_{ni}\boldsymbol{b}) - F(y)\} \|^2 \, dH(y).$$

When F is unknown but symmetric around 0, *Koul* (1985) defined a similar estimator $\boldsymbol{\beta}^+$ of $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta}^+ = \text{argmin}_b \; M^+(\boldsymbol{b}),$$

where, for $\boldsymbol{b} \in R^p$,

$$T^+(\boldsymbol{b}) = \int \| \sum_{i=1}^{n} \boldsymbol{x}'_{ni}\{I(Y_{ni} \leq y + \boldsymbol{x}'_{ni}\boldsymbol{b}) - I(-Y_{ni} < y - \boldsymbol{x}'_{ni}\boldsymbol{b})\} \|^2 \, dH(y).$$

In both papers, the authors discussed the asymptotic normality of these estimators. The estimator $\tilde{\boldsymbol{\beta}}$ provides the right extension of the one sample minimum distance estimation methodology of Wolfowitz (1957) to the linear regression setup.

This thesis analyzes the strong asymptotic behavior of these estimators.

In the first part (Chapter 2), some inequalities about weighted and centered empirical processes are developed. In the second part (Chapter 3), strong consistency of the above mentioned estimators is proved under different sets of conditions. Finally, in the third part (Chapter 4), a *Bahadur* type expansion of $\boldsymbol{\beta}^+$ is given, using the results from the first part.

To my parients and my wife

ACKNOWLEDGMENTS

I wish to express my sincere thanks to my advisor Professor Hira L. Koul for his patient guidance and continuous encouragement during the preparation of this dissertation. I would also like to thank Professors Habib Salehi, James Hannan, and James Stapleton for their serving on my guidance committee. My special thanks are due to Professor James Hannan for his very useful comments which lead to the improvement of this dissertation.

Finally, I express my deep appreciation to my wife for her stupendous encouragement and support to my study.

Contents

1 Introduction **1**

2 Tail Probability Inequalities **6**

2.1 Introduction 6

2.2 Results under sup-norm 7

2.3 Results under L_2-norm 16

3 Strong Consistency **21**

3.1 Introduction 21

3.2 Main results and proofs 24

4 Bahadur Expansion **36**

4.1 Main result and proof 36

Chapter 1

Introduction

The study of minimum distance (MD) estimation of a parameter can be traced back to that of the least square (LS) estimation. Other examples of classical MD estimation are, for instance, the least absolute deviation (LAD) and the least chi-square (LCS) estimations. In these methods, estimators are obtained by minimizing some types of distance functions related to the data and the parameters to be estimated. However, it was in 1950's that *Wolfowitz* first explicitly employed the concept of MD estimation when estimating a parameter by minimizing a distance between an empirical distribution function (d.f.) and the modeled parametric family of d.f.'s. As *Millar* (1981) commented that *Wolfowitz* took *Neyman*'s idea on minimum chi-square and elevated it to a general principle.

In his work, *Wolfowitz* (1953, 1954, 1957) demonstrated that MD method not only could be used in a wide range of problems but also yielded strongly consistent estimators even when sometimes classical methods, like the maximum likelihood method, failed to give a consistent estimator. *Wolfowitz*'s work drew people's attention to MD estimation. *Blackman* (1955) and *Bolthausen* (1977) studied the asymptotic normality of some MD estimators. *Pollard* (1980) worked on testing hypothesis problems with MD estimators. *Beran* (1977, 1978, 1982), *Parr & Schucany* (1979), *Millar* (1981, 1982, 1984), and *Donoho & Liu* (1988a,1988b) investigated various robustness and local asymptotic minimaxity of a large class of MD estimators. Most of these authors worked on the one sample or the two sample location models and found that the MD estimators corresponding to L_2-distances are generally more robust against certain gross errors than the ones corresponding to the supremum distance. A bibliography about the work on MD estimation up to 1980 can be found in *Parr* (1981).

The above mentioned MD methodology was extended to estimating parameters in linear regression models by *Koul* (1979, 1980, 1985a, 1985b), *Williamson* (1979, 1982), and *Koul & DeWet* (1983). These authors successfully established the asymptotic distributions of a class of MD estimators which minimize some *Cramer-Von Mises* type distances. Systematic presentation of the work in this field can be found in *Koul* (1992).

This thesis is concerned with the strong consistency, the rates of convergence, and the *Bahadur* type representations of the class of MD estimators defined by *Koul & DeWet* (1983) and by *Koul* (1985a, 1985b). A special case of the study can be seen in *Koul & Zhu* (1991). It is known that the almost sure convergence rate and *Bahadur* type expansion of an estimator provide a deeper understanding of its large sample behavior. They are also important in using the given estimator in sequential analysis.

We shall now describe the estimators studied in this thesis in more detail. Let $X_n,\ n \geq 1$ be a sequence of r.v.'s and $a_n,\ n \geq 1$, a sequence of real numbers. We write '$X_n = O(a_n)$' if $\limsup_{n\to\infty} |X_n| / |a_n| \leq M$ *a.s.* for some $0 < M < \infty$ and '$X_n = o(a_n)$' if $\limsup_{n\to\infty} |X_n| / |a_n| = 0$ *a.s.* Further, we write '$X_n < a_n \ \ wp1n$' if $\limsup_{n\to\infty} X_n/a_n < 1$ *a.s.*

Let $p \geq 1$ be an integer, F be a d.f. on the real line R, and $\{\varepsilon_i, 1 \leq i \leq n\}$ be independent and identically distributed (i.i.d.) random variables (r.v.'s) with the common d.f. F. Consider the p-dimensional linear regression model

$$Y_{ni} = \boldsymbol{x}_{i\cdot}\boldsymbol{\beta} + \varepsilon_i, \qquad 1 \leq i \leq n, \tag{1.1}$$

where $\boldsymbol{x}_{i\cdot}$, $1 \leq i \leq n$, is the ith row of a known real $n \times p$ design matrix $\boldsymbol{X}_n$ and $\boldsymbol{\beta}$ is the parameter p-vector to be estimated.

With respect to (1.1), define a weighted empirical process corresponding to an

$n \times p$ real weight matrix $\boldsymbol{D}_n$, of which $\boldsymbol{d}_{i\cdot}$ is the ith row, as

$$\boldsymbol{V}_{\mathbf{D}}(y, \boldsymbol{b}) = \sum_{i=1}^{n} \boldsymbol{d}_{i\cdot} I(Y_{ni} \leq y + \boldsymbol{x}_{i\cdot}\boldsymbol{b}), \qquad y \in R,\ \boldsymbol{b} \in R^p, \tag{1.2}$$

where I is the standard zero-one valued indicator function. Further, define a *Cramer-von Mises* type distance $(T(\cdot))^{1/2}$ between $V_{\mathbf{D}}(y, \boldsymbol{b})$ and the expectation of $V_{\mathbf{D}}(y, \boldsymbol{\beta})$ as

$$\begin{aligned} T(\boldsymbol{b}) &= \int \| \boldsymbol{V}_{\mathbf{D}}(y, \boldsymbol{b}) - E\boldsymbol{V}_{\mathbf{D}}(y, \boldsymbol{\beta}) \|^2 \, dH(y) \\ &= \int \| \sum_{i=1}^{n} \boldsymbol{d}_{i\cdot}\{I(Y_{ni} \leq y + \boldsymbol{x}_{i\cdot}\boldsymbol{b}) - F(y)\} \|^2 \, dH(y), \end{aligned} \tag{1.3}$$

where E is the expectation under (1.1), $\| \,.\, \|$ is the *Euclidean* norm, and H is a given nondecreasing right continuous function. When $p = 1$, *Koul & DeWet* (1983) defined a MD estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ as a minimizer of the function T assuming F is known. Their motivation of this definition is similar to that of the LS estimator: in the integrand of $T(\cdot)$, $V_{\mathbf{D}}(y, \boldsymbol{b}) - EV_{\mathbf{D}}(y, \boldsymbol{\beta})$, has mean $\mathbf{0}$ when $\boldsymbol{b}$ equals the true parameter $\boldsymbol{\beta}$. Observe that (1.3) actually defines a class of T functions, one corresponding to each H and $\boldsymbol{D}_n$. Therefore, a class of estimators $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained upon chosing different H's and $\boldsymbol{D}_n$'s in (1.3).

For the one dimensional case, i.e. when $p = 1$, and when F is known, *Koul & DeWet* (1983) studied some finite sample properties, asymptotic distribution, and asymptotic efficiency of $\tilde{\boldsymbol{\beta}}$. This study was later extended by *Koul* (1985a, b) to multiple linear regression models in which the errors could be either i.i.d. F with F being a known d.f., or independent with unknown d.f.'s F_i's which are symmetric around a common point. When F is known, the definition of $\tilde{\boldsymbol{\beta}}$ is as above. In the case when error d.f.'s F_i are unknown but symmetric about a common point, assuming that the common point of symmetry of F_i's is 0 without loss of generality, *Koul* (1985a, b) defined a MD estimator $\boldsymbol{\beta}^+$ of $\boldsymbol{\beta}$ as a minimizer of $T^+(\cdot)$, where, for $\boldsymbol{b} \in R^p$,

$$T^+(\boldsymbol{b}) = \int \| \sum_{i=1}^{n} \boldsymbol{d}_{i\cdot}\{I(Y_{ni} \leq y + \boldsymbol{x}_{i\cdot}\boldsymbol{b}) - I(-Y_{ni} < y - \boldsymbol{x}_{i\cdot}\boldsymbol{b})\} \|^2 \, dH(y). \tag{1.4}$$

The robustness of both $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^+$ was also discussed in *Koul*'s papers.

According to *Koul* (1985a), among the estimators obtained from choosing certain type of weight matrices $\boldsymbol{D}_n$, the one corresponding to $\boldsymbol{D}_n = \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1/2}$ is asymptotically most efficient. Therefore, in the sequel, we take $\boldsymbol{D}_n = \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1/2}$ and consider the cases when $F_i = F$ with F known and unknown. We use $\boldsymbol{\beta}^*$ for either $\tilde{\boldsymbol{\beta}}$ or $\boldsymbol{\beta}^+$.

According to *Wolfowitz* (1953, 1954, 1957), it is desirable to prove the strong consistency of the MD estimators $\boldsymbol{\beta}^*$. The first goal of this thesis is to give appropriate conditions under which $\boldsymbol{\beta}^*$ is strongly consistent. We in fact prove, under certain conditions, that

$$\| \boldsymbol{\beta}^* - \boldsymbol{\beta} \| = O(\gamma_n),$$

where $\{\gamma_n\}$ is a sequence of real numbers which depend only on the design matrices $\boldsymbol{X}_n$'s and converges to 0 for a wide choice of $\boldsymbol{X}_n$'s.

In 1966, *Bahadur* obtained linear expansions for sample quantiles as estimators of the population quantiles. It is known that *Bahadur* expansion is very useful in sequential analysis. *Ghosh* (1971) and *Ghosh & Sukathme* (1974) weakened *Bahadur*'s conditions and obtained similar expansions for sample quantiles in term of convergence in probability. *Haan & Taconis-Haantjes* (1979) further extended *Ghosh & Sukathme*'s work and also obtained *Bahadur*'s result under slightly weaker conditions than *Bahadur*'s. Others also obtained results similar to *Bahadur*'s for other estimators. An important example is in *Babu* (1989) where it is shown that the least absolute deviation estimator of linear regression parameter has *Bahadur* expansion. The second goal of this thesis is to obtain *Bahadur* type expansions for the MD estimators $\boldsymbol{\beta}^*$defined above. We shall prove that

$$\boldsymbol{\beta}^* - \boldsymbol{\beta} - \sum_{i=1}^{n} \boldsymbol{\phi}_i = O(R_n),$$

where $\boldsymbol{\phi}_i$'s are independent random vectors and $\{R_n\}$ is a sequence of real numbers which converge to 0 at a rate depending on the choice of $\boldsymbol{X}_n$'s. We call our expansions

'*Bahadur* type expansions' because the convergence rate R_n is different from what *Bahadur* obtained in the one sample problem. See Chapter 4 for the details.

To reach our goals, we first study in Chapter 2 some properties of the weighted empiricals defined in (1.2). Some tail probability inequalities related to the sup- and L_2-norms of these empirical processes are obtained in Section 2.2. The inequality pertaining to the sup-norm extends an inequality of *Ghosh & Sen* (1972) from a simple linear regression model to the multiple linear regression model. These inequalities are the fundmental tools used in the proofs of strong consistency and the *Bahadur* expansions of $\boldsymbol{\beta}^*$. In Chapter 3, strong consistency and convergence rates of $\boldsymbol{\beta}^*$ are discussed. According to our conclusions, many frequently used designs yield strongly consistent MD estimators $\boldsymbol{\beta}^*$. Finally, in Chapter 4, we present the *Bahadur* type expansions of $\boldsymbol{\beta}^*$ in detail.

Chapter 2

Tail Probability Inequalities

2.1 Introduction

In this chapter, we develop in Theorem 2.1 an exponential inequality for the tail probabilities of the centered weighted empirical processes (1.2). Then some large sample probability inequalities are obtained. *Ghosh & Sen* (1972) also derived a similar inequality involving the sup-norm of certain weighted and centered empirical processes for their study on bounded length confidence intervals. In the following, we give a brief description of *Ghosh & Sen*'s inequality because we are going to show that their result is actually covered by our Corollary 2.1.

For a sequence of real numbers $c_1, c_2, \cdots$, let

$$c_{ni}^* = (c_i - \bar{c}_n)/C_n, \qquad 1 \le i \le n,$$

where $\bar{c}_n = n^{-1}\sum_{i=1}^n c_i$ and $C_n^2 = \sum_{i=1}^n (c_i - \bar{c}_n)^2$. Also, let $\{Y_1, Y_2, \cdots\}$ be a sequence of i.i.d. r.v.'s having uniform distribution over (0,1) and let F be a d.f. on R. For $1 \le i \le n$, $0 < t < 1$, and $-\infty < b < \infty$, define

$$d_{ni}(t,b) = F(F^{-1}(t) + bc_{ni}^*) - t,$$

$$G^*(t,b) = \sum_{i=1}^n c_{ni}^*\Big[I(Y_i \le t + d_{ni}(t,b)) - d_{ni}(t,b)\Big]. \tag{2.1}$$

Then, under the assumptions

1. $\max_{1\le i\le n} |c_{ni}^*| = O(n^{-1/2})$,

2. $\liminf_{n\to\infty} n^{-1}C_n^2 \geq K_0 > 0$,

3. F is an absolutely continuous d.f. for which the density function f and its first derivative f' are bounded a.e. under *Lebesgue* measure,

Ghosh & Sen proved that for every $h > 0$, there exist positive constants K_1, K_2 and n^* (all of which may depend on h) such that for $n \geq n^*$, $k \geq 1$ and $0 < \delta < 1/4$,

$$P\Big(\sup_{0<t<1}\sup_{b\in A_n} |G_n^*(t,b) - G_n^*(t,0)| \geq K_1 n^{-\delta}(\ln n)^k\Big) \leq K_2 n^{-h}, \tag{2.2}$$

where $A_n = \{b: \ |b| \leq C(\ln n)^k\}$ and C a nonnegative real number.

In the next two sections, we present two similar probability inequalities with the r.v.'s Y_i's being generally i.i.d. instead of uniformly distributed. Of the two inequalities, one is for the $\widetilde{\text{sup}}$-norm (Corollary 2.1) and the other for the L_2-norm generated by a nondecreasing right continuous function (Theorem 2.2). Remark 2.1 shows that (2.2) is implied by our Corollary 2.1.

2.2 Results under sup-norm

The following notation is used in the sequel.

For any $n \times p$ real matrix $\boldsymbol{B} = \{b_{ij}\}$, $\boldsymbol{b}_{i\cdot}$ and $\boldsymbol{b}_{\cdot j}$ denote the ith row and the jth column of $\boldsymbol{B}$, respectively, for $1 \leq i \leq n$, $1 \leq j \leq p$, $\boldsymbol{B}'$ its transpose, and

$$|\boldsymbol{B}| := \max_{1\leq i\leq n} \parallel \boldsymbol{b}_{i\cdot} \parallel .$$

Similarly, for a vector $\boldsymbol{t} \in R^n$, define $\parallel \boldsymbol{t} \parallel$ to be the *Euclidean* norm and

$$|\boldsymbol{t}| := \max_{1\leq i\leq n}\{|t_i|\}.$$

Now, we describe the assumptions used in this section.

(A) $\{a_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$ are two sequences of real numbers such that $b_n \uparrow \infty$, $a_n \geq n^{-k_0}$ for some $0 < k_0 < \infty$, and $a_n b_n = o(1)$.

(B) The d.f. F satisfies that

$$\sup_{y \in R} |F(y+\delta) - F(y)| \le M_0 |\delta|,$$

for $\delta \in R$ and some $M_0 < \infty$.

(C) $\{\boldsymbol{C}_n, n \ge 1\}$ is a sequence of $n \times p$ real matrices, with some fixed $p \ge 1$, such that $\boldsymbol{C}_n' \boldsymbol{C}_n = \mathbf{I}_{p \times p}$, the identity matrix, $|\boldsymbol{C}_n| \ge n^{-k_0}$ for some $0 < k_0 < \infty$, and $|\boldsymbol{C}_n| \, b_n = o(1)$ for some $b_n \uparrow \infty$.

For any $y, u \in R$, $\boldsymbol{s}, \boldsymbol{t} \in R^n$, and $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ whose components are i.i.d. r.v's with d.f. F, let

$$\begin{aligned} I^*(Y_i \le y) &= I(Y_i \le y) - F(y), \\ W_{\boldsymbol{t}}(y, u\boldsymbol{s}) &= \sum_{i=1}^{n} t_i I^*(Y_i \le y + us_i). \end{aligned} \tag{2.3}$$

Note that the process $W_{\boldsymbol{t}}(\cdot, u\boldsymbol{s})$ reduces to the ordinary centered empirical process when $\boldsymbol{s} = \boldsymbol{0}$ and $\boldsymbol{t} = (1/n, \cdots, 1/n)'$. We are now ready to state our results.

Theorem 2.1 *Assume that (B) holds. Given $b > 1, h_1 < \infty,\ h_2 < \infty,\ \gamma > 0$,*

$$P\Big(\sup_{|y|<\infty} \sup_{0 \le u \le b} |W_{\boldsymbol{t}}(y, u\boldsymbol{s}) - W_{\boldsymbol{t}}(y, \boldsymbol{0})| \ge 2\gamma + 2M_0 \parallel \boldsymbol{s} \parallel \parallel \boldsymbol{t} \parallel \sigma\Big)$$

$$\le 32 \frac{b}{M_0 \sigma^2} \Big(\frac{n}{h_1}\Big)^{1/2} \exp\Big(-\frac{\gamma^2}{2[M_0 h_1 h_2 b + \frac{1}{3} h_2 \gamma]}\Big), \tag{2.4}$$

holds for all $n \ge 1$, all σ satisfying

$$0 < \sigma \le b^{-1} \wedge (2M_0 h_1^{1/2})^{-1},$$

and all $\boldsymbol{t} \in R^n$, and $\boldsymbol{s} \in R^n$ satisfying

$$\parallel \boldsymbol{s} \parallel^2 \vee \parallel \boldsymbol{t} \parallel^2 \le h_1, \quad |\boldsymbol{s}| \vee |\boldsymbol{t}| \le h_2. \tag{2.5}$$

The *Bernstein* inequality and some elementary facts about nondecreasing functions are used in the proof of this lemma. For the sake of self containment, they are stated in the following two lemmas, respectively.

Lemma 2.1 *(Bernstein). Let $X_1, \cdots, X_n$ be independent random variables satisfying $EX_i = 0$ and $|X_i| \leq m$ a.s., for each i, where $m < \infty$. Then, for all $\tau > 0$ and $n \geq 1$,*

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \geq \tau\right) \leq 2\exp\left(-\frac{\tau^2}{2[\sum_{i=1}^{n} Var(X_i) + \frac{1}{3}m\tau]}\right).$$

(See Serfling 1980, P95) □

Lemma 2.2 . *Let ϕ_1, ϕ_2, ψ_1 and ψ_2 be nondecreasing functions from R to R. Let $\Phi = \phi_1 - \phi_2$ and $\Psi = \psi_1 - \psi_2$. Then for any $x \in [a, b] \subset R$,*

$$|\Phi(x) - \Psi(0)| \leq \max(|\Phi(a) - \Psi(0)|, |\Phi(b) - \Psi(0)|) + \phi_2(b) - \phi_2(a), \tag{2.6}$$

and

$$\begin{aligned} |\Phi(x) - \Psi(x)| \;\; \leq \;\; & \max(|\Phi(a) - \Psi(b)|, |\Phi(b) - \Psi(a)|) \\ & + \phi_2(b) + \psi_2(b) - \phi_2(a) - \psi_2(a). \end{aligned} \tag{2.7}$$

□

Proof of Theorem 2.1. For a real number a, use a^+ and a^- to refer to the positive and negative part of it, respectively. Similarly, for a vector $\boldsymbol{t} = (t_1, \cdots, t_n) \in R^n$, let $\boldsymbol{t}^+ = (t_1^+, \cdots, t_n^+)$ and $\boldsymbol{t}^- = (t_1^-, \cdots, t_n^-)$. Then, $\boldsymbol{t} = \boldsymbol{t}^+ - \boldsymbol{t}^-$. According to (2.3) and by the triangle inequality, it is enough to prove (2.4) with the 32 on the RHS replaced by 16 for the case when $\boldsymbol{t}$ has nonnegative components.

Now, observe that

$$|W_{\boldsymbol{t}}(y, u\boldsymbol{s}) - W_{\boldsymbol{t}}(y, \boldsymbol{0})| = \left|\sum_{i=1}^{n} t_i I^*(-y - u(-s_i) \leq -Y_i < -y)\right|,$$

where I^* is defined as in (2.3). Thus, we can and hanceforth further restrict our proof of (2.4) with the 32 on the RHS replaced by 8 for the case that both $\boldsymbol{t}$ and $\boldsymbol{s}$ have nonnegative components.

To simplify the notation, in this proof, we write

$$I(y, u) = \sum_{i=1}^{n} t_i I(Y_i \leq y + us_i), \qquad F(y, u) = \sum_{i=1}^{n} t_i F(y + us_i)$$

$$W(y,u) = W_{\mathbf{t}}(y, u\boldsymbol{s}).$$

Hence,

$$W(y,u) = I(y,u) - F(y,u).$$

First, we prove that for any $y \in R$, $0 \le u \le b$, and $\tau > 0$,

$$P\Big(|W(y,u) - W(y,0)| \ge \tau\Big) \le 2\exp\Big(-\frac{\tau^2}{2(M_0\,|\boldsymbol{s}|\,\|\,\boldsymbol{t}\,\|^2\, b + \frac{1}{3}\,|\boldsymbol{t}|\,\tau)}\Big), \tag{2.8}$$

and for any $y_1 < y_2$,

$$P\Big(|W(y_2,0) - W(y_1,0)| \ge \tau\Big) \le 2\exp\Big(-\frac{\tau^2}{2(\|\,\boldsymbol{t}\,\|^2\,\delta + \frac{1}{3}\,|\boldsymbol{t}|\,\tau)}\Big), \tag{2.9}$$

where $\delta = F(y_2) - F(y_1)$.

Let

$$X_i \equiv t_i[I^*(Y_i \le y + us_i) - I^*(Y_i \le y)].$$

Then, $W(y,u) - W(y,0) = \sum_{i=1}^n X_i$ and

$$EX_i = 0, \qquad |X_i| \le t_i \le |\boldsymbol{t}|\,, \;\; 1 \le i \le n.$$

Further, by (B),

$$\sum_{i=1}^n \mathrm{Var}(X_i) \le \sum_{i=1}^n t_i^2[F(y + us_i) - F(y)] \le M_0\,|\boldsymbol{s}|\,\|\,\boldsymbol{t}\,\|^2\, b.$$

Hence, the *Bernstein* bound of Lemma 2.1 gives (2.8).

Similarly, if we let

$$X_i \equiv t_i[I^*(Y_i \le y_2) - I^*(Y_i \le y_1)],$$

then, $W(y_2,0) - W(y_1,0) = \sum_{i=1}^n X_i$ and

$$EX_i = 0, \qquad |X_i| \le t_i \le |\boldsymbol{t}|\,, \qquad 1 \le i \le n,$$

$$\sum_{i=1}^n \mathrm{Var}(X_i) \le \sum_{i=1}^n t_i^2(F(y_2) - F(y_1)) = \delta\,\|\,\boldsymbol{t}\,\|^2\,.$$

Again, the *Bernstein* bound of Lemma 2.1 gives (2.9).

Next, for a fixed a σ, $0 < \sigma \leq b^{-1} \wedge (2M_0 h_1^{1/2})^{-1}$, construct a partition $0 = \eta_0 < \eta_1 < \cdots < \eta_{r_n} = b$ of the range of u such that

$$\eta_r - \eta_{r-1} \leq \sigma, \qquad r = 1, 2, \cdots, r_n,$$

$$r_n \leq b\sigma^{-1}. \tag{2.10}$$

The assumption (B) and (2.10) imply that, for $1 \leq r \leq r_n$,

$$\begin{aligned} |F(y, \eta_r) - F(y, \eta_{r-1})| &= \left| \sum_{i=1}^{n} t_i \Big(F(y + s_i \eta_r) - F(y + s_i \eta_{r-1}) \Big) \right| \\ &\leq M_0 \sum_{i=1}^{n} t_i s_i (\eta_r - \eta_{r-1}) \\ &\leq M_0 \parallel \boldsymbol{s} \parallel \parallel \boldsymbol{t} \parallel \sigma \\ &:= M, \qquad \text{say.} \end{aligned} \tag{2.11}$$

Therefore, by the nondecreasing property of $I(y, u)$ and $F(y, u)$ in u and (2.6) of Lemma 2.2, it follows from (2.11) that

$$\sup_{0 \leq u \leq b} |W(y, u) - W(y, 0)| \leq \max_{1 \leq r \leq r_n} |W(y, \eta_r) - W(y, 0)| + M. \tag{2.12}$$

Now, for a fixed η_r, define

$$g_r(y) = F(y, \eta_r) + F(y, 0), \qquad y \in R.$$

Then, $g_r(\cdot)$ is nondecreasing and $0 \leq g_r \leq 2 \sum_{i=1}^{n} t_i \leq 2n^{1/2} \parallel \boldsymbol{t} \parallel$. Choose a partition $\{-\infty = \xi_0 < \xi_1 < \cdots < \xi_{v_n} = \infty\}$ of $(-\infty, \infty)$ such that

$$\Delta(r, v) := g_r(\xi_v) - g_r(\xi_{v-1}) \leq M,$$

$$v_n \leq \frac{2 \parallel \boldsymbol{t} \parallel n^{1/2}}{M} \leq \frac{2n^{1/2}}{M_0 \parallel \boldsymbol{s} \parallel \sigma}. \tag{2.13}$$

By the nondecreasing property of $F(y, \eta_r)$ and $I(y, \eta_r)$ in y, (2.7) of Lemma 2.2, and (2.13), when $y \in [\xi_{v-1}, \xi_v]$, we obtain

$$|W(y, \eta_r) - W(y, 0)|$$

$$\le \max\Big(|W(\xi_v,\eta_r) - W(\xi_{v-1},0)|\,,|W(\xi_{v-1},\eta_r) - W(\xi_v,0)|\Big) + \Delta(r,v)$$

$$\le \max_{i=v-1,v} |W(\xi_i,\eta_r) - W(\xi_i,0)| + |W(\xi_{v-1},0) - W(\xi_v,0)| + M, \tag{2.14}$$

where the last inequality follows from the triangle ineqality.

Hence,

$$\sup_{|y|<\infty} |W(y,\eta_r) - W(y,0)| \le \max_{1\le v\le v_n}\Big\{|W(\xi_v,\eta_r) - W(\xi_v,0)|\Big\} + \max_{1\le v\le v_n}\Big\{|W(\xi_v,0) - W(\xi_{v-1},0)|\Big\} + M. \tag{2.15}$$

Combining (2.12) with (2.15) obtains

$$\sup_{|y|<\infty}\sup_{0\le u\le b} |W(y,u) - W(y,0)| \le \max_{1\le v\le v_n}\max_{1\le r\le r_n}\Big\{|W(\xi_v,\eta_r) - W(\xi_v,0)|\Big\} + \max_{1\le v\le v_n}\max_{1\le r\le r_n}\Big\{|W(\xi_v,0) - W(\xi_{v-1},0)|\Big\} + 2M. \tag{2.16}$$

Now, from (2.5) and (2.8) we obtain

$$P\Big(\max_{0\le r\le r_n}\max_{1\le v\le v_n} |W(\xi_v,\eta_r) - W(\xi_v,0)| \ge \gamma\Big) \le 2r_n \times v_n \exp\Big(-\frac{\gamma^2}{2(M_0\,|\boldsymbol{s}|\,\|\,\boldsymbol{t}\,\|^2\, b + \frac{1}{3}\,|\boldsymbol{t}|\,\gamma)}\Big) \le 2r_n \times v_n \exp\Big(-\frac{\gamma^2}{2(M_0 h_1 h_2 b + \frac{1}{3}h_2\gamma)}\Big). \tag{2.17}$$

Next, observe that

$$\begin{aligned} \boldsymbol{t}^2[F(\xi_v) - F(\xi_{v-1})] &\le |\boldsymbol{t}|\,\Delta(0,v) \\ &\le M_0\,|\boldsymbol{t}|\,\|\,\boldsymbol{t}\,\|\|\,\boldsymbol{s}\,\|\,\sigma \\ &\le M_0 h_1 h_2 \sigma, \quad 1 \le v \le v_n. \end{aligned} \tag{2.18}$$

Hence, by (2.5), (2.9), and (2.18), we obtain

$$P\Big(\max_{0\le v\le v_n}\max_{1\le r\le r_n} |W(\xi_v,0) - W(\xi_{v-1},0)| \ge \gamma\Big) \le 2r_n \times v_n \exp\Big(-\frac{\gamma^2}{2(M_0 h_1 h_2 \sigma + \frac{1}{3}h_2\gamma)}\Big). \tag{2.19}$$

Finally, (2.4) follows from (2.17), (2.19), the upper bounds of r_n, v_n given in (2.10) and (2.13), and the fact that $\sigma \le b^{-1} \le b$. $\square$

In Theorem 2.1, if we properly take the values of b, σ, h_1, h_2, and γ, we have

Corollary 2.1 . *Assume that (A) and (B) hold and $b_n \ge (\ln n)^{1/2}$. For any $0 < k < \infty$, there exists a constant $K < \infty$ such that*

$$P\Big(\sup_{|y|<\infty}\sup_{0\le u\le b_n} |W_{\mathbf{t}}(y, u\mathbf{s}) - W_{\mathbf{t}}(y, \mathbf{0})| \ge K a_n^{1/2} b_n^{3/2}\Big) \le n^{-k} \tag{2.20}$$

holds for all $n \ge 1$ and for all $\mathbf{t} \in R^n$, $\mathbf{s} \in R^n$ satisfying

$$\| \mathbf{s} \| \vee \| \mathbf{t} \| \le 1, \quad |\mathbf{s}| \vee |\mathbf{t}| \le a_n. \tag{2.21}$$

Consequently, there exists a constant $K < \infty$ such that for any sequences $\mathbf{s}_n, \mathbf{t}_n \in R^n$ satisfying (2.21),

$$\sup_{|y|<\infty}\sup_{0\le u\le b_n} |W_{\mathbf{t}_n}(y, u\mathbf{s}_n) - W_{\mathbf{t}_n}(y, \mathbf{0})| < K a_n^{1/2} b_n^{3/2} \qquad wp1n.$$

Proof. We only need to prove (2.20). In Theorem 2.1, take $h_1 = 1$, $h_2 = a_n$, $b = b_n$, $\sigma = (2M_0)^{-1} \wedge b_n^{-1} \wedge a_n^{1/2} b_n^{3/2}$, and $\gamma = K_0 a_n^{1/2} b_n^{3/2}$ with K_0 to be determined. Then, $2M_0 \| \mathbf{t} \| \| \mathbf{s} \| \sigma \le 2M_0 a_n^{1/2} b_n^{3/2}$.

Now, under (2.21), the exponent in the RHS of (2.4) is

$$\begin{aligned}\frac{\gamma^2}{2[M_0 a_n b_n + \frac{1}{3} a_n \gamma]} &= \frac{K_0^2 a_n b_n^3}{2[M_0 a_n b_n + \frac{1}{3} a_n K_0 a_n^{1/2} b_n^{3/2}]} \\ &= \frac{K_0^2}{2[M_0 + \frac{1}{3} K_0 a_n^{1/2} b_n^{1/2}]} b_n^2.\end{aligned}$$

Also, by (A) and the choice of σ, we have $b_n = o(a_n^{-1}) = o(n^{k_0})$ and $(a_n b_n)^{-1} \le a_n^{-1} \le n^{k_0}$. Thus, there exists a constant $0 < M < \infty$ such that the coefficient part $32 \frac{b_n}{M_0 \sigma} (\frac{n}{h_1})^{1/2}$ on the RHS(2.4) is bounded above by $M n^{2k_0 + 1/2}$.

Then, (2.20) follows for $K = K_0 + 2M_0$ from the facts that $a_n b_n = o(1)$, $b_n \ge (\ln n)^{1/2}$, and

$$\lim_{K_0 \to \infty} \frac{K_0^2}{2\big(M_0 + \frac{1}{3} K_0 a_n^{1/2} b_n^{1/2}\big)} = \infty.$$

$\square$

Remark 2.1 *The continuity of F implies that $F(Y_i)$ are i.i.d. uniformly distributed on (0,1). Observe that $G_n^*(t,b) - G_n^*(t,0)$ defined in (2.1) equals to our $W_{\mathbf{t}_n}(y, b\boldsymbol{s}_n) - W_{\mathbf{t}_n}(y, \mathbf{0})$ at the $y = F^{-1}(t)$ when $\boldsymbol{t} = \boldsymbol{s} = \mathbf{c}_n^*$. Hence,*

$$\begin{aligned} &\sup_{|y|<\infty, |b|\le b_n} |W_{\mathbf{t}_n}(y, b\boldsymbol{s}_n) - W_{\mathbf{t}_n}(y, \mathbf{0})| \\ &\qquad \ge \sup_{0<t<1} \sup_{b \in I_n^*} |G_n^*(t,b) - G_n^*(t,0)| \,. \end{aligned}$$

If we further take $a_n = n^{-\delta}$ and $b_n = (\ln n)^{\sigma}$, where $0 < \delta < 1/2$ and $\sigma \ge 1/2$, all conditions in Corollary 2.1 are satisfied so that (2.20) holds. Therefore, our corollary does imply the result of Ghosh & Sen appearing at the inequality (2.2). $\square$

Based on Corollary 2.1, we can further take supermum over another variable to obtain the following corollary. This allows us to use these results in multi-dimensional regression problems. See Chapter 3 for details. To state this corollary and the next theorem, we further define

$$\begin{aligned} \mathcal{E} &= \{\boldsymbol{e} \in R^p : \|\, \boldsymbol{e} \,\| \le 1\}, \\ \mathcal{D}_n &= \{\boldsymbol{d} \in R^n : \boldsymbol{d} = \boldsymbol{C}_n \boldsymbol{e},\ \boldsymbol{e} \in \mathcal{E}\}, \end{aligned} \tag{2.22}$$

where $\boldsymbol{C}_n$ is a $n \times p$ real matrix.

For simplicity, in the following we shall not exhibit the dependence of $\mathcal{D}_n$ and $\boldsymbol{C}_n$ on n, i.e. we write $\mathcal{D}$ and $\boldsymbol{C}$ for $\mathcal{D}_n$ and $\boldsymbol{C}_n$, respectively.

Corollary 2.2 . *Assume that (B) holds and that the $\boldsymbol{C}_n$'s used to define $\mathcal{D}_n$ satisfy assumption (C) with $b_n \ge (\ln n)^{1/2}$. Then, there exists a constant $K < \infty$ such that*

$$\sup_{|y|<\infty, 0\le b\le b_n, \mathbf{d}\in\mathcal{D}} |W_{\mathbf{t_n}}(y, b\boldsymbol{d}) - W_{\mathbf{t_n}}(y, \mathbf{0})| < K\,|C|^{1/2}\, b_n^{3/2}, \quad wp1n, \tag{2.23}$$

for any sequence $\boldsymbol{t}_n \in R^n$, $\|\, \boldsymbol{t}_n \,\| \le 1$, $|\boldsymbol{t}_n| \le |C|$.

Proof. As argued in the proof of Theorem 2.1, assume that all components of $\boldsymbol{t}_n$ are nonnegative and write $W(y, b\boldsymbol{d})$ for $W_{\mathbf{t}_n}(y, b\boldsymbol{d})$.

For each n, split $\mathcal{D}$ into m_n different parts: $\mathcal{D}_1, \cdots, \mathcal{D}_{m_n}$ such that

1. the diameter of each $\mathcal{D}_k$ is no larger than n^{-1}, $1 \leq k \leq m_n$,

2. $m_n \leq (p\,n)^p$.

(Split the cubic $[-1, 1]^p$ into equal volume small cubics, say.)

Fix a point $\boldsymbol{d}^k$ in each $\mathcal{D}_k$. Let $\boldsymbol{\delta}_n = (n^{-1}, \cdots, n^{-1})' \in R^n$ and let

$$\mathcal{D}_0 = \{\boldsymbol{d}^k \pm \boldsymbol{\delta}_n, k = 1, \cdots, m_n\}.$$

Then $\mathcal{D}_0$ contains $2m_n$ different points.

Now, for any $\boldsymbol{d} \in \mathcal{D}$, $\;\boldsymbol{d} \in \mathcal{D}_k$ for some k. By the fact that the ith summand in $W(y, b\,\boldsymbol{d})$ is a difference of two nondecreasing functions of $b\,d_i$, $1 \leq i \leq n$, and (2.6) of Lemma 2.2 we obtain that for every $y \in R$,

$$\begin{aligned}
&\sup_{\mathbf{d}\in\mathcal{D}} |W(y, b\,\boldsymbol{d}) - W(y, 0)| \\
&\qquad \leq \max_{\mathbf{d}\in\mathcal{D}_0} |W(y, b\,\boldsymbol{d}) - W(y, 0)| \\
&\qquad\qquad + \max_{1\leq k\leq m_n} \sum_{i=1}^{n} t_i \Big(F(y + b(d_i^k + n^{-1})) - F(y + b(d_i^k - n^{-1}))\Big) \\
&\qquad \leq \max_{\mathbf{d}\in\mathcal{D}_0} |W(y, b\,\boldsymbol{d}) - W(y, 0)| + 2b\,|C|,
\end{aligned}$$

where we used assumption (B) to obain the last step.

Observe that when $0 \leq b \leq b_n$, $b\,|C| = o(b_n^{3/2}\,|C|^{1/2})$. Thus, to complete the proof of (2.23), it suffices to show that there exists a constant $0 < K < \infty$ such that

$$\sup_{|y|<\infty} \sup_{0\leq b\leq b_n} \sup_{\mathbf{d}\in\mathcal{D}_0} |W(y, b\,\boldsymbol{d}) - W(y, 0)| \leq K\,|C|^{1/2}\,b_n^{3/2}, \qquad wp1n. \tag{2.24}$$

By Corollary 2.1 applied with $a_n = |C|$ and $k = p + 2$, there exists a constant $K < \infty$ such that

$$P\Big(\sup_{|y|<\infty} \sup_{0\leq b\leq b_n} \sup_{\mathbf{d}\in\mathcal{D}_0} |W(y, b\,\boldsymbol{d}) - W(y, 0)| \geq K\,|C|^{1/2}\,b_n^{3/2}\Big)$$

$$\leq \sum_{d \in \mathcal{D}_0} P\Big(\sup_{|y|<\infty} \sup_{0 \leq b \leq b_n} |W(y, b\boldsymbol{d}) - W(y, \boldsymbol{0})| \geq K\,|C|^{1/2}\, b_n^{3/2}\Big)$$

$$\leq 2m_n n^{-(p+2)}$$

$$\leq 2p^p n^{-2}.$$

Now, (2.24) follows from the *Borel-Cantelli* Lemma. This also completes the proof of (2.23). □

2.3 Results under L_2-norm

Let H be a nondecreasing right continuous real function and $\|\cdot\|_H$ the L_2-norm induced by H. We further assume

(D) The d.f. F and the integrating measure H satisfy the following: there exist $a_0 > 0$ and $M_1 < \infty$ such that as $\delta \to 0$

$$\sup_{|a|<a_0} \int \Big(F(y+\delta) - F(y)\Big)^2 dH(y+a) \leq M_1 \delta^2.$$

(E) There exist $\sigma > 1$, $A < \infty$, $0 < B < \infty$, $n_0 \geq 1$, and $\Lambda < \infty$ such that for all $\lambda \geq \Lambda$ and $n \geq n_0$,

$$P\Big((a_n b_n)^{-1}\,|H(Y) - H_-(Y - a_n b_n)| > \lambda\Big) \leq A\exp(-B\lambda^\sigma),$$

where Y is a r.v. with d.f. F and H_- is the left limit of H. Also,

$$\int F(1-F)dH < \infty. \tag{2.25}$$

The next theorem gives an analog of (2.20) for the L_2-norm $\|\,.\,\|_H$. This theorem holds for large n only.

Theorem 2.2 . *Let assumptions (A), (D), and (E) hold and* $b_n \geq (\ln n)^{1/2}$. *Then, for any* $k > 0$ *there exist a constant* $K < \infty$ *and* $N_0 \geq 1$ *such that*

$$P\Big(\sup_{0\leq u \leq b_n} \|\, W_{\boldsymbol{t}}(\cdot, u\boldsymbol{s}) - W_{\boldsymbol{t}}(\cdot, \boldsymbol{0})\,\|_H \geq K a_n^{1/2} b_n^{3/2}\Big) \leq n^{-k}, \tag{2.26}$$

holds for all $\boldsymbol{s}$, $\boldsymbol{t} \in R^n$ *satistying (2.21) and for all* $n \geq N_0$.

To prove Theorem 2.2, the following lemma due to *Bychkova* (1986) is used.

Lemma 2.3 . *Let $\{X_k,\ k \geq 1\}$ be a sequence of random elements taking values in a Hilbert space such that $EX_k = 0,\ k \geq 1$, with 0 being the 0-element of the Hilbert space. If*

$$P\Big(\| X_k \|_h \geq x\Big) \leq A\exp(-Bx^{q/(q-1)}),$$

where $1 < q < 2,\ A > 0$, and $B > 0$, then for any sequence of real numbers $\{v_n,\ n \geq 1\}$ satisfying

$$\sum_{n=1}^{\infty} |v_n|^{\alpha} < \infty$$

for an $\alpha \in (q, 2]$, we have

$$P\Big(\| \sum_{k=1}^{\infty} v_k X_k \|_h \geq x\Big) \leq \exp(-A_\alpha x^{\alpha/(\alpha-1)}),$$

where $A_\alpha > 0$ is a constant depending only on α and $\| \ . \ \|_h$ is the norm defined on the Hilbert space. □

Proof of Theorem 2.2. Similar to the proof of Theorem 2.1, it suffices to prove (2.26) for $\boldsymbol{t}$ and $\boldsymbol{s}$ having nonnegative components.

Let $\mathcal{H}$ be the *Hilbert* space defined by $\| \ . \ \|_H$. For $0 \leq u \leq b_n,\ y \in R$, and $1 \leq i \leq n$, let

$$X_i(y) = (a_n b_n)^{-1/2}\Big\{I^*(Y_i \leq y + us_i) - I^*(Y_i \leq y)\Big\}. \tag{2.27}$$

Then

$$\begin{aligned} \| X_i \|_H^2 = (a_n b_n)^{-1}\Big\{\int & I(y < Y_i \leq y + us_i)dH(y) \\ & + \int \Big(F(y + us_i) - F(y)\Big)^2 dH(y) \\ & -2\int I(y < Y_i \leq y + us_i)\Big(F(y + us_i) - F(y)\Big)dH(y)\Big\}. \end{aligned}$$

Clearly, for all $1 \leq i \leq n$,

$$\int I(y < Y_i \leq y + us_i)\,dH(y) = H_-(Y_i) - H_-(Y_i - us_i). \tag{2.28}$$

Because $0 < u \le b_n$ and $|s| \le a_n$, by (A) and (D) it follows that there exists an $1 \le N_1 < \infty$ such that

$$\begin{aligned}\max_{1\le i\le n} \int \Big(F(y+us_i) - F(y)\Big)^2 dH(y) &\le M_1(us_i)^2 \\ &\le M_1(a_n b_n)^2, \qquad (2.29)\end{aligned}$$

for all $1 \le i \le n$ and $n \ge N_1$.

Further, the *Cauchy-Schwarz* inequality, (2.28), and (2.29) imply that

$$\begin{aligned}&\int I(y < Y_i \le y + us_i)\Big(F(y+us_i) - F(y)\Big)\, dH(y) \\ &\qquad \le M_1^{1/2} a_n b_n (H_-(Y_i) - H_-(Y_i - us_i))^{1/2}, \qquad (2.30)\end{aligned}$$

for all $1 \le i \le n$ and $n \ge N_1$. These imply that $X_i \in \mathcal{H}$, $1 \le i \le n$ and $n \ge N_1$.

Now, by (2.29),

$$\max_{1\le i\le n} (a_n b_n)^{-1} \int \Big(F(y+us_i) - F(y)\Big)^2 dH(y) = O(a_n b_n) = o(1),$$

and by (2.28), (2.30), and (E), there exist $A < \infty$, $0 < B < \infty$, $\sigma > 1$, and $\Lambda < \infty$ such that for $1 \le i \le n$, $\lambda > \Lambda$, and $n \ge N_0 = n_0 \vee N_1$,

$$\begin{aligned}&P\Big((a_n b_n)^{-1} \int I(y < Y_i \le y + us_i)\, dH(y) > \lambda^2\Big) \\ &\qquad \le A \exp(-B\lambda^{2\sigma}),\end{aligned}$$

$$\begin{aligned}&P\Big((a_n b_n)^{-1} \int I(y < Y_i \le y + us_i)\Big(F(y+us_i) - F(y)\Big)\, dH(y) > \lambda^2\Big) \\ &\qquad \le A \exp(-B\lambda^{2\sigma}).\end{aligned}$$

These inequalities imply that there exist $A_1 < \infty$, and $0 < B_1 < \infty$ such that for $1 \le i \le n$, $n \ge N_0$,

$$\begin{aligned}P(\| X_i \|_H > \lambda) &= P(\| X_i \|_H^2 > \lambda^2) \\ &\le A_1 \exp(-B_1\lambda^{2\sigma}).\end{aligned}$$

Since $\sigma > 1$, there is a $1 < q < 2$ such that $2\sigma = q/(q-1)$. Apply Lemma 2.3 to the $\{X_i\}$ defined in (2.27) with $\alpha = 2,\ v_i = t_i,\ 1 \le i \le n,\ v_i = 0,\ i > n$, to obtain

$$P\Big((a_n b_n)^{-1/2} \parallel W(\cdot, u\mathbf{s}_n) - W(\cdot, \mathbf{0}) \parallel_H > \lambda\Big) \le \exp(-A\lambda^2),$$

for some $A < \infty$ and $n \ge N_0$. Taking $\lambda = Kb_n$ gives that for $0 \le u \le b$ and $n \ge N_0$,

$$P\Big(\parallel W(\cdot, u\mathbf{s}) - W(\cdot, \mathbf{0}) \parallel_H > K a_n^{1/2} b_n^{3/2}\Big) \le \exp(-AK^2 b_n^2). \tag{2.31}$$

Now, take a partition on $[0, b_n]$ as in the proof of Theorem 2.1 and take the configuration of $b,\ h_1,\ h_2$, and σ as in Corollary 2.1. Then, $\sigma \le a_n^{1/2} b_n^{3/2}$. The discussion similar to that for (2.12) in the proof of Theorem 2.1 leads to

$$\begin{aligned} &\sup_{0 \le u \le b_n} \parallel W(\cdot, u\mathbf{s}_n) - W(\cdot, \mathbf{0}) \parallel_H \\ &\quad \le \max_{0 \le r \le r_n} \parallel W(\cdot, \eta_r \mathbf{s}_n) - W(\cdot, \mathbf{0}) \parallel_H \\ &\qquad + \max_{0 \le r \le r_n} \parallel F(\cdot, \eta_r) - F(\cdot, \eta_{r-1}) \parallel_H . \end{aligned} \tag{2.32}$$

By (A), (D), and the *Cauchy-Schwarz* inequality,

$$\begin{aligned} &\parallel F(\cdot, \eta_r) - F(\cdot, \eta_{r-1}) \parallel_H^2 \\ &\quad = \parallel \sum_{i=1}^{n} t_i \Big(F(\cdot + s_i \eta_r) - F(\cdot + s_i \eta_{r-1})\Big) \parallel_H^2 \\ &\quad \le \parallel \mathbf{t} \parallel^2 \sum_{i=1}^{n} \parallel F(\cdot + s_i \eta_r) - F(\cdot + s_i \eta_{r-1}) \parallel_H^2 \\ &\quad \le M_1 (a_n^{1/2} b_n^{3/2})^2, \qquad n \ge N_1. \end{aligned}$$

Therefore, for $n \ge N_1$, (2.32) can be rewritten as

$$\begin{aligned} &\sup_{0 \le u \le b_n} \parallel W(\cdot, u\mathbf{s}) - W(\cdot, \mathbf{0}) \parallel_H \\ &\quad \le \max_{0 \le r \le r_n} \parallel W(\cdot, \eta_r \mathbf{s}) - W(\cdot, \mathbf{0}) \parallel_H + M_1^{1/2} a_n^{1/2} b_n^{3/2}. \end{aligned}$$

To prove (2.26), it thus suffice to show that there exists a constant $K < \infty$ such that for all $n \ge N_0$,

$$P\Big(\max_{0 \le r \le r_n} \parallel W(\cdot, \eta_r \mathbf{s}) - W(\cdot, \mathbf{0}) \parallel_H > K a_n^{1/2} b_n^{3/2}\Big) \le n^{-k}. \tag{2.33}$$

By (2.31),

$$\begin{aligned} LHS(2.33) &\leq \sum_{0\leq r\leq r_n} P\Big(\| W(\cdot,\eta_r \mathbf{s}_n) - W(\cdot,\mathbf{0}) \|_H > K a_n^{1/2} b^{3/2}\Big) \\ &\leq r_n \exp(-AK^2 b_n^2), \qquad n \geq N_0. \end{aligned}$$

Since $b_n \geq (\ln n)^{1/2}$ and $r_n \leq b_n \sigma^{-1} \leq (a_n b_n)^{-1/2} \leq a_n^{-1/2} \leq n^{k_0/2}$ by (A), we can select K large enough so that (2.33) holds. This completes our proof. □

The following corollary is analogous to Corollary 2.2.

Corollary 2.3 . *Define $\mathcal{E}$ and $\mathcal{D}_n$ as in (2.22). In the assumptions of Theorem 2.2, replace (A) by (C). Then there exists a constant $K < \infty$ such that, for any $\boldsymbol{t}_n \in R^n$, $\| \boldsymbol{t}_n \| \leq 1$, $|\boldsymbol{t}_n| \leq |C|$,*

$$\sup_{0\leq b\leq b_n, \mathbf{d}\in\mathcal{D}} \| W_{\mathbf{t}_n}(\cdot, b\boldsymbol{d}) - W_{\mathbf{t}_n}(\cdot,\mathbf{0}) \|_H < K\,|C|^{1/2}\, b_n^{3/2}, \quad wp1n. \tag{2.34}$$

□

The proof of Corollary 2.3 is similar to that of Corollary 2.2 and hence is not given.

Chapter 3

Strong Consistency

3.1 Introduction

In this chapter, we first recall the MD estimators defined in Chapter 1. Asymptotic distributions of these estimators have been studied by *Koul & DeWet* (1983) and *Koul* (1985a, 1985b). We will present in the next section some strong consistency results of these estimators.

Consider the linear regression model (1.1). As in Chapter 2, we will not exhibit the dependence of Y_{ni} and $\boldsymbol{X}_n$ on n and use $\boldsymbol{x}_{i\cdot}$ and $\boldsymbol{x}_{\cdot j}$ for the ith row and jth column of $\boldsymbol{X}$, respectively.

Let $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X}$ and assume that $\boldsymbol{S}^{-1}$ exist for all $n \geq p$. Let

$$\boldsymbol{C} = \boldsymbol{X}\boldsymbol{S}^{-1/2}. \tag{3.1}$$

Then the model (1.1) is equivalent to

$$Y_i = \mathbf{c}_{i\cdot}\boldsymbol{\Delta} + \varepsilon_i, \qquad 1 \leq i \leq n, \tag{3.2}$$

where $\boldsymbol{c}_{i\cdot}$ is the ith row of the $\boldsymbol{C}$ and

$$\boldsymbol{\Delta} = \boldsymbol{S}^{1/2}\boldsymbol{\beta}. \tag{3.3}$$

Observe that the design matrices $\boldsymbol{C}$'s of (3.2) satisfies $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{I}_{p\times p}$. Our study is conducted based on model (3.2). Any conclusions obtained can also be translated to the forms with respect to model (1.1) according to (3.3).

Given a nondecreasing right continuous function H from R to R, let

$$T(\boldsymbol{b}) = \int \| \boldsymbol{U}(y, \boldsymbol{C}\boldsymbol{b}) \|^2 \, dH(y), \qquad \boldsymbol{b} \in R^p,$$

where

$$\boldsymbol{U}(y, \boldsymbol{C}\boldsymbol{b}) = \sum_{i=1}^{n} \boldsymbol{c}_{i\cdot}\{I(Y_i \le y + \boldsymbol{c}_{i\cdot}\boldsymbol{b}) - F(y)\}. \tag{3.4}$$

Note that under (3.2), $E\,\boldsymbol{U}(y, \boldsymbol{C}\boldsymbol{\Delta}) = \boldsymbol{0}$. This motivates one to define a MD estimator $\tilde{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta}$, in the case F is known, as a minimizer of $T(\cdot)$:

$$\tilde{\boldsymbol{\Delta}} = \text{argmin}_b\, T(\boldsymbol{b}).$$

Similarly, in the case that F is unknown but symmetric around 0, a MD estimator $\boldsymbol{\Delta}^+$ of $\boldsymbol{\Delta}$ is defined as a minimizer of $T^+(\cdot)$:

$$\boldsymbol{\Delta}^+ = \text{argmin}_b\, T^+(\boldsymbol{b}),$$

where, for $\boldsymbol{b} \in R^p$,

$$T^+(\boldsymbol{b}) = \int \| \boldsymbol{U}^+(y, \boldsymbol{C}\boldsymbol{b}) \|^2 \, dH(y),$$

$$\boldsymbol{U}^+(y, \boldsymbol{C}\boldsymbol{b}) = \sum_{i=1}^{n} \boldsymbol{c}_{i\cdot}\{I(Y_i \le y + \boldsymbol{c}_{i\cdot}\boldsymbol{b}) - I(-Y_i < y - \boldsymbol{c}_{i\cdot}\boldsymbol{b})\}. \tag{3.5}$$

Note that for $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^+$ defined in Chapter 1 with $\boldsymbol{D} = \boldsymbol{C}$, we have

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{S}^{-1/2}\tilde{\boldsymbol{\Delta}}, \qquad \boldsymbol{\beta}^+ = \boldsymbol{S}^{-1/2}\boldsymbol{\Delta}^+. \tag{3.6}$$

See *Koul & DeWet* (1983) and *Koul* (1985a) for more motivation and other properties of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^+$.

In this chapter, we give the strong consistency results for $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^+$ along with rates. To this effect, besides the assumptions (B)–(E) in Chapter 1, we shall also use the following assumptions.

(F) The d.f. F satisfies (2.25) of (E) and has a density f which satisfies

$$0 < \| f \|_H^2 < \infty, \tag{3.7}$$

$$\int_{[-1 \le H \le 1]} f dH > 0. \tag{3.8}$$

(G) There exist $\sigma > 1,\ A < \infty,\ B > 0$, and $\Lambda < \infty$ such that for all $\lambda \geq \Lambda$,

$$P\Big(H(|\varepsilon|) - H_-(-|\varepsilon|) > \lambda\Big) \leq A\exp(-B\lambda^\sigma),$$

where ε has distribution F.

(H) There exists $0 < \alpha < 2$ such that

$$\sum_{n=1}^{\infty} |C|^{2\alpha} < \infty \qquad \text{and} \qquad |C|^\alpha = o((\ln n)^{-1}).$$

(I) The function F and H are such that

$$\iint_{x\leq y} F(x)\big(1 - F(y)\big)\, dH(x)\, dH(y) < \infty.$$

The following lemma demonstrates some facts related to assumptions (F) and (I).

Lemma 3.1 . *Let F be a distribution function and H be a nondecreasing right continuous real function. Then, the following hold.*

(1) $\quad$ $\int F(1-F)dH < \infty$ *if and only if* $\int H_- dF < \infty$.

(2) $\quad$ $\int H_-^2 dF < \infty$ *if and only if*

$$\iint_{x\leq y} F(x)(1 - F(y))dH(x)dH(y) < \infty.$$

Proof. By the *Fubini* Theorem,

$$\begin{aligned}\int F(1-F)dH &= \iiint_{s\leq x<t} dF(s)dF(t)dH(x)\\ &= \iint_{s<t} [H_-(t) - H_-(s)]dF(s)dF(t)\\ &= \frac{1}{2}\iint |H_-(t) - H_-(s)|\, dF(s)dF(t).\end{aligned}$$

Now, (1) follows from the fact that for two independent r.v.'s X and Y, $E\,|X - Y| < \infty$ if and only if $E\,|X| < \infty$ and $E\,|Y| < \infty$.

To prove (2), by the *Fubini* Theorem,

$$\begin{aligned}
&\iint_{x\le y} F(x)(1-F(y))dH(x)dH(y) \\
&= \iiiint_{s\le x\le y<t} dF(s)dF(t)dH(x)dH(y) \\
&\le \frac{1}{2}\iint_{s<t}[H_-(t)-H_-(s)]^2 dF(s)dF(t),
\end{aligned}$$

and

$$\begin{aligned}
&\iint_{x\le y} F(x)(1-F(y))dH(x)dH(y) \\
&\ge \frac{1}{2}\iiiint_{s\le x,y<t} dF(s)dF(t)dH(x)dH(y) \\
&= \frac{1}{4}\iint [H_-(t)-H_-(s)]^2 dF(s)dF(t).
\end{aligned}$$

Hence (2) follows from a fact similar to the one used in the proof of (1), i.e., for two independent r.v.'s X and Y, $E(X-Y)^2<\infty$ if and only if $EX^2<\infty$ and $EY^2<\infty$.

□

3.2 Main results and proofs

Throughout the rest of the thesis, the $\boldsymbol{C}_n$ and $\boldsymbol{b}_n$ in the assumption (C) are taken to be $\boldsymbol{C}$ of (3.1) and $(\ln n)^{1/2}$, respectively. Also because the estimators $\tilde{\boldsymbol{\Delta}}$ and $\boldsymbol{\Delta}^+$ are translation invariant (*Koul* (1985b, 1992)), throughout the following, the true parameter $\boldsymbol{\Delta}$ will be assumed to be $\mathbf{0}$. For simplifying our notation, we also assume that H is continuous.

We first present the strong consistency of $\boldsymbol{\beta}^+$ under assumptions (B), (C), (F), (G), and the symmetry of F around 0. Then the strong consistency of $\tilde{\boldsymbol{\beta}}$ under (B), (F), (H), and (I) is given.

Theorem 3.1 *In addition to (B), (C), (F), and (G), assume that F is symmetric around 0. Then, there exists a constant $K_1<\infty$ such that*

$$\|\boldsymbol{\Delta}^+\| < (K_1 \ln n)^{\frac{1}{2}}, \qquad wp1n,$$

and (recalling that $\boldsymbol{\beta}^+ = \boldsymbol{S}^{-1/2}\boldsymbol{\Delta}^+$*),*

$$\| \boldsymbol{\beta}^+ \| < (K_1 \| \boldsymbol{S}^{-1/2} \|^2 \ln n)^{1/2}, \qquad wpln.$$

□

Remark 3.1 . *This theorem implies that if the design matrix* $\boldsymbol{X}$ *is such that*

$$\| \boldsymbol{S}^{-1/2} \|^2 = o((\ln n)^{-1}),$$

the Koul *estimators* $\boldsymbol{\beta}^+$ *are strongly consistent for* $\boldsymbol{\beta}$ *in (1.1). This condition is satisfied by a large class of designs. Examples include the one sample location model where* $\boldsymbol{X} = (1, \cdots, 1)'$*, so that* $\boldsymbol{S}^{-1/2} = n^{-1/2}$ *and* $|\boldsymbol{C}| = n^{-1/2}$*; and the first order polynomial through the origin where* $\boldsymbol{X} = (1, 2, \cdots, n)'$*, so that* $\boldsymbol{S}^{-1/2} \leq n^{-3/2}$ *and* $|\boldsymbol{C}| \leq n^{-1/2}$. □

Here is our next theorem.

Theorem 3.2 *Under the assumptions (B), (F), (H), and (I), there exists a constants* $K < \infty$ *such that*

$$\| \tilde{\boldsymbol{\Delta}} \| < (K |\boldsymbol{C}|^{-\alpha})^{\frac{1}{2}}, \qquad wpln,$$

and

$$\| \tilde{\boldsymbol{\beta}} \| < (K \| \boldsymbol{S}^{-1/2} \|^2 |\boldsymbol{C}|^{-\alpha})^{1/2}, \qquad wpln.$$

□

Now, we proceed to prove these two theorems. Theorem 3.1 is a consequence of the following three Lemmas.

Lemma 3.2 . *Under assumptions (B), (C), (F), and (G), there exists a constant* $K_0 > 0$ *such that*

$$T^+(\mathbf{0}) < K_0 \ln n \qquad wpln. \tag{3.9}$$

Proof. By the definition,

$$\begin{aligned} T^+(\mathbf{0}) &= \int \| \boldsymbol{U}^+(y,\mathbf{0}) \|^2 \, dH(y) \\ &= \sum_{j=1}^{p} \int \Big\{ \sum_{i=1}^{n} c_{ij} \Big(I(Y_i \le y) - I(-Y_i < y) \Big) \Big\}^2 dH(y) \\ &:= \sum_{j=1}^{p} T_j(\mathbf{0}) \end{aligned} \tag{3.10}$$

with

$$T_j(\mathbf{0}) = \int \Big\{ \sum_{i=1}^{n} c_{ij} \Big(I(Y_i \le y) - I(-Y_i < y) \Big) \Big\}^2 dH(y), \qquad 1 \le j \le p.$$

Since p is fixed, it thus suffices to show that (3.9) holds for each $T_j(\mathbf{0})$, $1 \le j \le p$.

Let

$$\begin{aligned} X_i(y) &= I(Y_i \le y) - I(-Y_i < y), \;\; 0 \le i \le n, \;\; y \in R, \\ \mathcal{H} &= \Big\{ g : \| g \|_H^2 < \infty \Big\}. \end{aligned}$$

Observe that

$$\begin{aligned} \| X_1 \|_H^2 &= \int \Big[I(Y_1 \le y) - I(-Y_1 < y) \Big]^2 dH(y) \\ &= H(|Y_1|) - H(-|Y_1|). \end{aligned}$$

From (1) of Lemma 3.1, (2.25) of (E), and the symmetry of F, it follows that $\{X_i\}$ is a sequence of i.i.d. $\mathcal{H}$-valued random elements satisfying $E\,X_i = 0$, $i \ge 1$. Condition (G) implies that when $\lambda \ge \Lambda$

$$\begin{aligned} P(\| X_1 \|_H \ge \lambda) &= P(\| X_1 \|_H^2 > \lambda^2) \\ &= P\Big(H(|Y_1|) - H(-|Y_1|) > \lambda^2 \Big) \\ &\le A \exp(-B\lambda^{2\sigma}). \end{aligned}$$

Since $2\sigma > 2$, there is q, $1 < q < 2$, such that $2\sigma = q/(q-1)$.

Now for each fixed $1 \leq j \leq p$, take $\alpha = 2$ and $a_i = c_{ij}$ when $0 \leq i \leq n$, $a_i = 0$ when $i > n$. Then

$$\sum_{i=1}^{\infty} |a_i|^2 = 1 < \infty.$$

By Lemma 2.3,

$$\begin{aligned} P\Big(T_j(0) \geq \lambda^2\Big) &= P(\| \sum_{i=1}^{\infty} c_{ij} X_i \|_H > \lambda) \\ &\leq \exp(-A_j \lambda^2) \\ &\leq \exp(-A\lambda^2), \end{aligned}$$

where $A_j > 0$, $1 \leq j \leq p$, are constants independent of $\mathbf{c}_{\cdot j}$ and $A = \min_{1 \leq j \leq p}(A_j)$.

Now, take $\lambda = (K \ln n)^{1/2}$ with the K such that $AK > 1$ in the above inequality. Then the *Borel-Cantelli* Lemma and (3.10) imply that (3.9) holds with $K_0 = pK$. □

Lemma 3.3 . *If H is a nondecreasing right continuous real function, then there exists a nonnegative real function g such that*

(a) $\quad 0 < g \leq 1.$

(b) $\quad \int g dH < \infty.$

Proof. We just construct such a function g. Let

$$g(x) = \begin{cases} 1 & |H(x)| \leq 1, \\ \frac{1}{H^2(x)} & |H(x)| > 1. \end{cases}$$

This g satisfies (a). To prove (b), we only need to prove $\int_{[H>1]} g dH < \infty$. By *Fatou* Lemma,

$$\int_{[H>1]} g dH = \lim_{n\to\infty} \int_{[1<H\leq n]} g dH \leq \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{i^2} < \infty.$$

This completes the proof. □

Lemma 3.4 . *Assume that (B), (C) and (F) hold. Then there exists a constant* $0 < K_1 < \infty$, *such that*

$$\inf_{\|\mathbf{b}\| \geq (K_1 \ln n)^{1/2}} T^+(\boldsymbol{b}) > K_0 \ln n, \qquad wp1n, \tag{3.11}$$

with K_0 *as in Lemma 3.2.*

Proof. Let $h_n = (K_1 \ln n)^{1/2}$ with $K_1 < \infty$ to be determined and $\mathcal{E}$ and $\mathcal{D}$ be as in (2.22) with $\boldsymbol{C}_n$ equal to $\boldsymbol{C}$ of (3.1). Let g be as in Lemma 3.3 and define

$$\mu(x) = \int_{-\infty}^{x} g dH. \qquad x \in R. \tag{3.12}$$

Then μ is a bounded nondecreasing function and so is $\| g \|_H$. We assume that $\| g \|_H = 1$, without loss of generality, and denote $\mu_0 = \int g dH$.

For any $\boldsymbol{b} \in R^p, \boldsymbol{b} \neq \mathbf{0}$, there is unique $\boldsymbol{e} \in \mathcal{E}$ such that

$$\boldsymbol{b} = \| \boldsymbol{b} \| \, \boldsymbol{e} = b\, \boldsymbol{e},$$

where $b = \| \boldsymbol{b} \|$. Therefore (3.11) is equivalent to

$$\inf_{b \geq h_n, \boldsymbol{e} \in \mathcal{E}} T^+(b\,\boldsymbol{e}) > K_0 \ln n, \qquad wp1n. \tag{3.13}$$

By the *Cauchy-Schwarz* inequality,

$$\begin{aligned} T^+(\boldsymbol{b}) &= T^+(b\,\boldsymbol{e}) \\ &= \int \| \boldsymbol{e} \|^2 \| \boldsymbol{U}^+(y, b\boldsymbol{C}\boldsymbol{e}) \|^2 \, dH(y) \\ &\geq \int \left[\mathbf{e}'\boldsymbol{U}^+(y, b\,\boldsymbol{d})\right]^2 dH(y) \\ &\geq \left[\int U_d^*(y, b\,\boldsymbol{d})\, d\mu(y)\right]^2, \end{aligned} \tag{3.14}$$

where $\boldsymbol{d} = (d_1, d_2, \cdots, d_n)' = \boldsymbol{C}\boldsymbol{e}$ and

$$\begin{aligned} U_d^*(y, b\,\boldsymbol{d}) &:= \boldsymbol{e}'\boldsymbol{U}^+(y, b\boldsymbol{d}) \\ &= \sum_{i=1}^{n} d_i \{ I(Y_i \leq y + bd_i) - I(-Y_i < y - bd_i) \}. \end{aligned}$$

Observe that for any fixed $\boldsymbol{d} \in \mathcal{D}$ and $y \in R$, $U_d^*(y, b\,\boldsymbol{d})$ is a nondecreasing function of b. Therefore, when $b \geq h_n$,

$$U_d^*(y, b\,\boldsymbol{d}) \geq U_d^*(y, h_n \boldsymbol{d}), \qquad y \in R,\ \boldsymbol{d} \in \mathcal{D}.$$

Hence, to prove (3.13), it suffices to show that

$$\inf_{\mathbf{d} \in \mathcal{D}} \int U_d^*(y, h_n \boldsymbol{d})\, d\mu(y) > (K_0 \ln n)^{1/2}, \qquad wp1n,$$

or

$$\sup_{\mathbf{d} \in \mathcal{D}} - \int U_d^*(y, h_n \boldsymbol{d})\, d\mu(y) < -(K_0 \ln n)^{1/2}, \qquad wp1n. \tag{3.15}$$

Now, divide $\mathcal{D}$ into m_n pieces, say, $\mathcal{D}_1, \cdots, \mathcal{D}_{m_n}$ such that

1. The diameter of $\mathcal{D}_k$ is no larger than n^{-2}, $k = 1, \cdots, m_n$.

2. $m_n \leq (pn^2)^p$.

By the *Fatou* lemma and (3.8) of (F),

$$\liminf_{\delta \to 0} \int \frac{1}{\delta}[F(x+\delta) - F(x)]d\mu \geq \int f d\mu \geq \int_{[-1 \leq H \leq 1]} f dH > 0. \tag{3.16}$$

Therefore, we can select a $0 < \eta < \infty$ such that

$$LHS(3.16) > \eta. \tag{3.17}$$

Let K_1 be such that $\eta\sqrt{K_1} - 2\sqrt{K_0} > 0$. We first prove that there exists an $1 \leq N < \infty$ such that

$$\begin{aligned} P\Big(- \int U_{d^k}^*(y, h_n \boldsymbol{d}^k)\, d\mu(y) \geq -2(K_0 \ln n)^{1/2}\Big) \\ \leq \exp\Big\{-\frac{1}{2}(\eta\sqrt{K_1} - 2\sqrt{K_0})^2 \ln n\Big\}, \end{aligned} \tag{3.18}$$

for every $\boldsymbol{d}^k \in \mathcal{D}_k$ and $n \geq N$.

From (C) and (3.17) it follows that, there exists $1 \leq N < \infty$ such that, for all $n \geq N$,

$$\begin{aligned}
&E \int U^*_{d^k}(y, b_n \mathbf{d}^k)\, d\mu(y) \\
&= \int \sum_{i=1}^{n} d_i^k \Big(F(y + h_n d_i^k) - F(y - h_n d_i^k)\Big)\, d\mu(y) \\
&= \int \sum_{i=1}^{n} d_i^k h_n d_i^k \frac{1}{h_n d_i^k} \Big(F(y + h_n d_i^k) - F(y - h_n d_i^k)\Big)\, d\mu(y) \\
&\geq \eta (K_1 \ln n)^{1/2}. \qquad (3.19)
\end{aligned}$$

Now observe that $\int U^*_{d^k}(y, b_n \mathbf{d}^k)\, d\mu(y)$ is the sum of n independent bounded r.v.'s

$$X_i = d_i^k \int \Big(I(Y_i \leq y + d_i^k h_n) - I(-Y_i < y - d_i^k h_n)\Big)\, d\mu(y)$$

with the bounds given as

$$-\left|d_i^k\right| \leq X_i \leq \left|d_i^k\right|, \quad 1 \leq i \leq n.$$

By the exponential inequality of Hoeffding (Theorem 2, 1963) and (3.19), for $n \geq N$,

$$\begin{aligned}
&P\Big(- \int U^*_{d^k}(y, b_n \mathbf{d}^k)\, d\mu(y) \geq -2(K_0 \ln n)^{1/2}\Big) \\
&\quad \leq P\Big(- \int U^*_{d^k}(y, b_n \mathbf{d}^k)\, d\mu(y) + E \int U^*_{d^k}(y, b_n \mathbf{d}^k)\, d\mu(y) \\
&\qquad \geq \big(\eta \sqrt{K_1} - 2\sqrt{K_0}\big)(\ln n)^{1/2}\Big) \\
&\quad \leq \exp\Big(-\frac{1}{2}\big(\eta\sqrt{K_1} - 2\sqrt{K_0}\big)^2 \ln n\Big).
\end{aligned}$$

This proves (3.18).

Note that the RHS in (3.18) does not depend on k. Hence,

$$\begin{aligned}
&P\Big(\max_{1 \leq k \leq m_n} \Big\{- \int U^*_{d^k}(y, h_n \mathbf{d}^k)\, d\mu(y)\Big\} \geq -2(K_0 \ln n)^{1/2}\Big) \\
&\quad \leq m_n \exp\Big\{-\frac{1}{2}\big(\eta\sqrt{K_1} - 2\sqrt{K_0}\big)^2 \ln n\Big\} \\
&\quad \leq p^p n^{-\frac{1}{2}(\eta\sqrt{K_1} - 2\sqrt{K_0})^2 + 2p}, \quad n \geq N. \qquad (3.20)
\end{aligned}$$

Clearly, there exists a positive constant K_1 such that the RHS of (3.20) is summable in n. Then, the *Borel-Cantelli* lemma gives

$$\max_{1\le k\le m_n}\left\{-\int U^*_{d^k}(y, h_n \boldsymbol{d}^k)\, d\mu(y)\right\} < -2(K_0 \ln n)^{1/2}, \qquad wp1n. \tag{3.21}$$

Next, for any $\boldsymbol{d} \in \mathcal{D}$, $\boldsymbol{d} \in \mathcal{D}_k$ for some k, and

$$\begin{aligned}
&\left|\int \left[U^*_d(y, h_n\boldsymbol{d}) - U^*_{d^k}(y, h_n\boldsymbol{d}^k)\right] d\mu(y)\right| \\
&= \left|\int \left[U^*_d(y, h_n\boldsymbol{d}) - U^*_{d^k}(y, h_n\boldsymbol{d})\right] d\mu(y)\right. \\
&\qquad \left. + \int \left[U^*_{d^k}(y, h_n\boldsymbol{d}) - U^*_{d^k}(y, h_n\boldsymbol{d}^k)\right] d\mu(y)\right| \\
&\le \Big(\sup_{|y|<\infty} \left|\sum_{i=1}^n (d^k_i - d_i)\left[I(Y_i \le y + h_n d_i) - I(-Y_i < y - h_n d_i)\right]\right| \\
&\qquad + \sup_{|y|<\infty} \left|\sum_{i=1}^n d^k_i \left[I(Y_i \le y + h_n d^k_i) - I(Y_i \le y + h_n d_i)\right]\right| \\
&\qquad + \sup_{|y|<\infty} \left|\sum_{i=1}^n d^k_i \left[I(-Y_i < y - h_n d^k_i) - I(-Y_i < y - h_n d_i)\right]\right| \Big)\mu_0 \\
&:= (I_1 + I_2 + I_3)\mu_0, \qquad say.
\end{aligned} \tag{3.22}$$

We have

$$\begin{aligned}
I_1 &= \sup_{|y|<\infty} \left|\sum_{i=1}^n (d^k_i - d_i)\left[I(Y_i \le y + h_n d_i) - I(-Y_i < y - h_n d_i)\right]\right| \\
&\le \| \boldsymbol{d}^k - \boldsymbol{d} \| \, n^{1/2} \\
&= O(n^{-3/2}).
\end{aligned} \tag{3.23}$$

Recall the definition of W_t from (2.3). By Corollary 2.2, assumption (B), and the fact that $|C|\, h_n = o(1)$,

$$\begin{aligned}
I_2 &= \sup_{|y|<\infty} \left|\sum_{i=1}^n d^k_i \left[I(Y_i \le y + h_n d^k_i) - I(Y_i \le y + h_n d_i)\right]\right| \\
&\le \sup_{|y|<\infty} \Big\{ \left|W_{\mathrm{d}^k}(y, h_n\boldsymbol{d}^k) - W_{\mathrm{d}^k}(y, \mathbf{0})\right| + \left|W_{\mathrm{d}^k}(y, h_n\boldsymbol{d}) - W_{\mathrm{d}^k}(y, \mathbf{0})\right| \\
&\qquad + \left|\sum_{i=1}^n d^k_i \left(F(y + h_n d^k_i) - F(y + h_n d_i)\right)\right| \Big\} \\
&= O(|\boldsymbol{C}|^{1/2} (\ln n)^{3/4}) \\
&= o((\ln n)^{1/2}).
\end{aligned} \tag{3.24}$$

Similarly,

$$I_3 = o((\ln n)^{1/2}). \tag{3.25}$$

Combining (3.22) – (3.25),

$$\sup_{\mathbf{d}\in\mathcal{D}} \left| \int \left[U_d^*(y, h_n \boldsymbol{d}) - U_{d^k}^*(y, h_n \boldsymbol{d}^k) \right] d\mu(y) \right| = o((\ln n)^{1/2}). \tag{3.26}$$

Finally, by (3.21) and (3.26),

$$\begin{aligned}
&\sup_{\mathbf{d}\in D} - \int U_d^*(y, h_n \boldsymbol{d})\, d\mu(y) \\
&\quad = \sup_{\mathbf{d}\in D} \Big(- \int U_{d^k}^*(y, h_n \boldsymbol{d}^k)\, d\mu(y) + \int \left[U_{d^k}^*(y, h_n \boldsymbol{d}^k) - U_d^*(y, h_n \boldsymbol{d}) \right] d\mu(y) \Big) \\
&\quad \le \max_{1 \le k \le m_n} \Big\{ - \int U_{d^k}^*(y, h_n \boldsymbol{d}^k)\, d\mu(y) \Big\} + o((\ln n)^{1/2}) \\
&\quad < -(K_0 \ln n)^{1/2}, \qquad\qquad wpln,
\end{aligned}$$

thereby proving (3.15) and also the lemma. □

Now, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. By the definition of $\boldsymbol{\Delta}^+$,

$$T^+(\boldsymbol{\Delta}^+) \le T^+(\mathbf{0}).$$

On the other hand, by Lemmas 3.2 and 3.4,

$$\inf_{\|\boldsymbol{b}\| \ge (K_1 \ln n)^{1/2}} T^+(\boldsymbol{b}) > K_0 \ln n > T^+(\mathbf{0}), \qquad wpln.$$

Therefore,

$$\| \boldsymbol{\Delta}^+ \| < (K_1 \ln n)^{1/2}, \qquad wpln$$

This completes the proof of Theorem 3.1. □

Next, to prove Theorem 3.2, we prove the following three lemmas.

Lemma 3.5 *Let*

$$T_{\mathbf{t}}(\mathbf{0}) = \int \Big\{ \sum_{i=1}^n t_i \Big(I(Y_i \le y) - F(y) \Big) \Big\}^2 dH(y).$$

Then, under the assumption (I), for every $\mathbf{t} \in R^n$,

$$E T_{\mathbf{t}}^2(\mathbf{0}) \le 6 \parallel \boldsymbol{t} \parallel^4 \iint_{x \le y} F(x)(1 - F(y))\, dH(x)\, dH(y) < \infty. \tag{3.27}$$

Proof. Observe that by the *Fubini* Theorem,

$$\begin{aligned}
E T_{\mathbf{t}}^2(\mathbf{0}) &= E\Big[\int \Big\{\sum_{i=1}^n t_i\big(I(Y_i \le y) - F(y)\big)\Big\}^2 dH(y)\Big]^2 \\
&= E\Big\{\Big[\int \Big\{\sum_{i=1}^n t_i\big(I(Y_i \le x) - F(x)\big)\Big\}^2 dH(x)\Big] \cdot \\
&\qquad \Big[\int \Big\{\sum_{j=1}^n t_j\big(I(Y_j \le y) - F(y)\big)\Big\}^2 dH(y)\Big]\Big\} \\
&\le 2\iint_{x \le y} \Big\{\sum_{i=1}^n t_i^4 E\Big[\big(I(Y_i \le y) - F(y)\big)^2\big(I(Y_i \le x) - F(x)\big)^2\Big] \\
&\quad + \sum\sum\nolimits_{i \ne j} t_i^2 t_j^2 E\Big[\big(I(Y_i \le y) - F(y)\big)^2\big(I(Y_j \le x) - F(x)\big)^2\Big] \\
&\quad + 2\sum\sum\nolimits_{i \ne j} t_i^2 t_j^2 E\Big[\big(I(Y_i \le x) - F(x)\big)\big(I(Y_i \le y) - F(y)\big) \\
&\qquad \big(I(Y_j \le y) - F(y)\big)\big(I(Y_j \le x) - F(x)\big)\Big]\Big\}\, dH(x)\, dH(y) \\
&= 2\iint_{x \le y} \Big\{\sum_{i=1}^n t_i^4 A + \sum\sum\nolimits_{i \ne j} t_i^2 t_j^2 B \\
&\qquad + 2\sum\sum\nolimits_{i \ne j} t_i^2 t_j^2 D\Big\}\, dH(x)\, dH(y).
\end{aligned} \tag{3.28}$$

where

$$\begin{aligned}
A &= E\Big[\big(I(Y_i \le y) - F(y)\big)^2\big(I(Y_i \le x) - F(x)\big)^2\Big], \\
B &= E\big(I(Y_i \le y) - F(y)\big)^2 E\big(I(Y_j \le x) - F(x)\big)^2,
\end{aligned}$$

and

$$\begin{aligned}
D = E\Big[&\big(I(Y_i \le x) - F(x)\big)\big(I(Y_i \le y) - F(y)\big) \\
&\big(I(Y_j \le y) - F(y)\big)\big(I(Y_j \le x) - F(x)\big)\Big].
\end{aligned}$$

Further, when $x \le y$,

$$\begin{aligned}
A &\le E\Big|\big(I(Y_i \le x) - F(x)\big)\big(I(Y_i \le y) - F(y)\big)\Big| \\
&= F(x)\big(1 - F(y)\big)\Big[1 + 2\big(F(y) - F(x)\big)\Big]
\end{aligned}$$

$$
\begin{aligned}
&\leq \; 3F(x)(1-F(y)) && (3.29) \\
B \; &= \; F(x)(1-F(x))F(y)(1-F(y)) && \\
&\leq \; F(x)(1-F(y)) && (3.30) \\
D \; &= \; F^2(x)(1-F(y))^2 && \\
&\leq \; F(x)(1-F(y)) && (3.31)
\end{aligned}
$$

Then (3.27) follows from (3.29) – (3.31). □

Lemma 3.6 . *Assume (H) and (I) hold. Then*

$$T(\mathbf{0}) = \int \mathbf{U}^2(y,\mathbf{0})\, dH(y) < |C|^{-\alpha} \qquad wpln \tag{3.32}$$

Proof. Apply Lemma 3.5 p times, the jth time to $\boldsymbol{t} = \boldsymbol{c}_{\cdot j}$, $1 \leq j \leq p$ and use the fact that $\| \boldsymbol{c}_{\cdot j} \| = 1$ together with the *Cauchy-Schwatz* inequality to obtain

$$
\begin{aligned}
ET^2(\mathbf{0}) \; &= \; E\Big[\int \| \boldsymbol{U}(y,\mathbf{0}) \|^2 \; dH(y)\Big]^2 \\
&= \; E\Big[\sum_{j=1}^{p} \int \Big\{\sum_{i=1}^{n} c_{ij}\big(I(Y_i \leq y) - F(y)\big)\Big\}^2 dH(y)\Big]^2 \\
&\leq \; \Big(\sum_{j=1}^{p}\Big\{E\Big[\int \Big\{\sum_{i=1}^{n} c_{ij}\big(I(Y_i \leq y) - F(y)\big)\Big\}^2 dH(y)\Big]^2\Big\}^{1/2}\Big)^2 \\
&\leq \; 6p^2 \iint_{x \leq y} F(x)(1-F(y))dH(x)dH(y) \\
&< \; \infty.
\end{aligned}
$$

Now, (3.32) follows from the *Markov* inequality, (H), and *Borel-Cantelli* Lemma. □

Lemma 3.7 . *Assume that (B), (C), and (F) hold for* $b_n = |C|^{-\alpha/2}$. *Then there exists a constant* $0 < K < \infty$, *such that*

$$\inf_{\|\boldsymbol{b}\| \geq K|C|^{-\alpha/2}} T(\boldsymbol{b}) > |C|^{-\alpha} \qquad wpln. \tag{3.33}$$

Proof. This proof is similar to that of Lemma 3.4. □

Proof of Theorem 3.2. Similar to that of Theorem 3.1. □

Chapter 4

Bahadur Expansion

4.1 Main result and proof

In this chapter, we further give a *Bahadur* type expansion for $\boldsymbol{\Delta}^+$ so that a similar expansion can also be obtained for β^+.

We need further assume that

($\boldsymbol{J}$) F has density function f whose derivative satisfies

$$\limsup_{\delta \to 0} \int (f'(y+\delta))^2 \, dH(y) < \infty.$$

To describe the theorem, define

$$\boldsymbol{B}^+ = -\int \boldsymbol{U}^+(y, \boldsymbol{0}) f(y) \, dH(y),$$

$$\hat{T}^+(\boldsymbol{b}) = \int \| \boldsymbol{U}^+(y, \boldsymbol{0}) + 2\boldsymbol{b} f(y) \|^2 \; dH(y),$$

and

$$\hat{\boldsymbol{\Delta}}^+ = \text{argmin}_b \, \hat{T}^+(\boldsymbol{b}).$$

Remark 4.1 . *Observe that* $2 \| f \|_H^2 \, \hat{\boldsymbol{\Delta}}^+ = \boldsymbol{B}^+$ *if* $0 < \| f \|_H < \infty$. *Because of this fact and (3.9),*

$$\| \hat{\boldsymbol{\Delta}}^+ \| < (K_2 \ln n)^{1/2}, \qquad\qquad wp1n,$$

where $K_2 = \frac{1}{4} K_0 \| f \|_H^{-2}$. □

Theorem 4.1 *In addition to the linear model (3.2) with true $\boldsymbol{\Delta} = \mathbf{0}$ and the symmetry of F around 0, assume (B) - (G) and (J) hold. Then*

$$\| f \|_H^2 \, (\boldsymbol{\Delta}^+)' = -\frac{1}{2}\int \sum_{i=1}^{n} \Big[\boldsymbol{c}_{i\cdot}\{I(Y_i \le y) - I(-Y_i < y)\}\Big] f(y) dH(y) + R_n, \qquad (4.1)$$

where

$$R_n = O(|\boldsymbol{C}|^{1/4} (\ln n)^{5/8}).$$

Proof. By the facts $\| \boldsymbol{a} \|^2 - \| \boldsymbol{b} \|^2 = (\boldsymbol{a} - \boldsymbol{b})'(\boldsymbol{a} + \boldsymbol{b})$, and

$$2 \, \| f \|_H^2 \, \hat{\boldsymbol{\Delta}}^+ = -\int \boldsymbol{U}^+(y, \mathbf{0}) f(y) \, dH(y),$$

we obtain

$$\begin{aligned}
&\hat{T}^+(\boldsymbol{\Delta}^+) - \hat{T}^+(\hat{\boldsymbol{\Delta}}^+) \\
&= \int \Big(\| \boldsymbol{U}^+(y,\mathbf{0}) + 2f(y)\boldsymbol{\Delta}^+ \|^2 - \| \boldsymbol{U}^+(y,\mathbf{0}) + 2f(y)\hat{\boldsymbol{\Delta}}^+ \|^2\Big) \, dH(y) \\
&= 2\int (\boldsymbol{\Delta}^+ - \hat{\boldsymbol{\Delta}}^+)'\Big(2\boldsymbol{U}^+(y,\mathbf{0}) + 2f(y)(\boldsymbol{\Delta}^+ + \hat{\boldsymbol{\Delta}}^+)\Big) f(y) \, dH(y) \\
&= 4 \, \| f \|_H^2 \, \| \boldsymbol{\Delta}^+ - \hat{\boldsymbol{\Delta}}^+ \|^2 . \qquad (4.2)
\end{aligned}$$

Observe that the first term of the RHS of (4.1) is $\| f \|_H^2 \, \hat{\boldsymbol{\Delta}}^+$. To prove (4.1), it thus suffices to estimate the convergence rate of the LHS of (4.2).

From Theorem 3.1 and Remark 4.1, we have

$$\| \boldsymbol{\Delta}^+ \| < h_n \quad \text{and} \quad \| \hat{\boldsymbol{\Delta}}^+ \| < h_n, \qquad wpln,$$

where $h_n = (K \ln n)^{1/2}$ with K being the maximum of the K_1 from Theorem 3.1 and K_2 from Remark 4.1. Hence,

$$\begin{aligned}
\Big|\hat{T}^+(\boldsymbol{\Delta}^+) - \hat{T}^+(\hat{\boldsymbol{\Delta}}^+)\Big| &\le \Big|\hat{T}^+(\boldsymbol{\Delta}^+) - T^+(\boldsymbol{\Delta}^+)\Big| + \Big|T^+(\hat{\boldsymbol{\Delta}}^+) - \hat{T}^+(\hat{\boldsymbol{\Delta}}^+)\Big| \\
&\le 2 \sup_{\|\mathbf{b}\| \le h_n} \Big|T^+(\boldsymbol{b}) - \hat{T}^+(\boldsymbol{b})\Big|, \qquad wpln. \qquad (4.3)
\end{aligned}$$

For a fixed $\boldsymbol{b} \in R^p$, we have

$$\begin{aligned}
\Big|T^+(\boldsymbol{b}) - \hat{T}^+(\boldsymbol{b})\Big| &= \Big|\int \Big(\| \boldsymbol{U}^+(y, \boldsymbol{C}\,\boldsymbol{b}) \|^2 - \| \boldsymbol{U}^+(y,\mathbf{0}) + 2\,f(y)\,\boldsymbol{b} \|^2\Big) \, dH(y)\Big| \\
&\le \int \| I_1 \|^2 \, dH + \int \| I_3 \|^2 \, dH \\
&\quad +2\int |I_2'(I_1 + I_3)| \, dH + 2\int |I_1' I_3| \, dH \qquad (4.4)
\end{aligned}$$

where

$$
\begin{aligned}
I_1 &= \boldsymbol{W}^+(y, \boldsymbol{C}\,\boldsymbol{b}) - \boldsymbol{W}^+(y, \boldsymbol{0}), \\
I_2 &= \boldsymbol{U}^+(y, \boldsymbol{0}) + 2f(y)\,\boldsymbol{b}, \\
I_3 &= E\boldsymbol{U}^+(y, \boldsymbol{C}\,\boldsymbol{b}) - 2f(y)\boldsymbol{b},
\end{aligned}
$$

with $\boldsymbol{W}^+(y,x) = \boldsymbol{U}^+(y,x) - E\boldsymbol{U}^+(y,x)$.

By (J), with some $|\theta_i| \le |\boldsymbol{c}_{i\cdot}\boldsymbol{b}|$ and $|\zeta_i| \le |\boldsymbol{c}_{i\cdot}\boldsymbol{b}|$,

$$
\begin{aligned}
\| I_3 \|_H &\le \Big[\sum_{j=1}^{p} \int \Big\{\sum_{i=1}^{n} c_{ij}\Big(F(y + \boldsymbol{c}_{i\cdot}\boldsymbol{b}) - F(y) - \boldsymbol{c}_{i\cdot}\boldsymbol{b}\, f(y)\Big)\Big\}^2 dH(y)\Big]^{1/2} \\
&\quad + \Big[\sum_{j=1}^{p} \int \Big\{\sum_{i=1}^{n} c_{ij}\Big(F(y - \boldsymbol{c}_{i\cdot}\boldsymbol{b}) - F(y) - \boldsymbol{c}_{i\cdot}\boldsymbol{b}\, f(y)\Big)\Big\}^2 dH(y)\Big]^{1/2} \\
&= \Big[\sum_{j=1}^{p} \int \Big\{\sum_{i=1}^{n} c_{ij}(\boldsymbol{c}_{i\cdot}\boldsymbol{b})^2 f'(y + \theta_i)\Big\}^2 dH(y)\Big]^{1/2} \\
&\qquad + \Big[\sum_{j=1}^{p} \int \Big\{\sum_{i=1}^{n} c_{ij}(\boldsymbol{c}_{i\cdot}\boldsymbol{b})^2 f'(y + \zeta_i)\Big\}^2 dH(y)\Big]^{1/2} \\
&= O\Big(\Big[\sum_{j=1}^{p}\Big(\sum_{i=1}^{n} |c_{ij}| \cdot (\boldsymbol{c}_{i\cdot}\boldsymbol{b})^2\Big)^2\Big]^{1/2}\Big) \\
&= O\Big(\Big[\sum_{j=1}^{p}\Big(\sum_{i=1}^{n} \| \boldsymbol{c}_{i\cdot} \|^2 \, |\boldsymbol{C}| \, \| \boldsymbol{b} \|^2\Big)^2\Big]^{1/2}\Big) \\
&= O(|\boldsymbol{C}| \, \| \boldsymbol{b} \|^2).
\end{aligned}
$$

Therefore

$$
\sup_{\|\boldsymbol{b}\| \le h_n} \| I_3 \|_H = O(|\boldsymbol{C}|\, h_n^2) = O(|\boldsymbol{C}| \ln n). \tag{4.5}
$$

By Lemma 3.2,

$$
\| \boldsymbol{U}^+(\cdot, \boldsymbol{0}) \|_H = O((\ln n)^{1/2}), \qquad wp1n.
$$

Hence

$$
\sup_{\|\boldsymbol{b}\| \le h_n} \| I_2 \|_H = O((\ln n)^{1/2}), \qquad wp1n \tag{4.6}
$$

Moreover,

$$
\| I_1 \|_H = \Big\{\sum_{j=1}^{p} \int \Big(W_j^+(y, \boldsymbol{C}\,\boldsymbol{b}) - W_j^+(y, \boldsymbol{0})\Big)^2 dH(y)\Big\}^{1/2}
$$

$$\leq \Big\{\sum_{j=1}^{p}\int\Big(W_j(y,\boldsymbol{C}\,\boldsymbol{b})-W_j(y,\boldsymbol{0})\Big)^2 dH(y)\Big\}^{1/2}$$
$$+\Big\{\sum_{j=1}^{p}\int\Big(W_j(-y,\boldsymbol{C}\,\boldsymbol{b})-W_j(-y,\boldsymbol{0})\Big)^2 dH(y)\Big\}^{1/2},$$

with W_j^+ and W_j being the jth component of $\boldsymbol{W}^+$ and $\boldsymbol{W}$, respectively, and $\boldsymbol{W}(y,\boldsymbol{Cb}) = \boldsymbol{U}(y,\boldsymbol{Cb}) - E\boldsymbol{U}(y,\boldsymbol{Cb})$ for $\boldsymbol{U}(y,\boldsymbol{Cb})$ as in (3.4). By Corollary 2.3,

$$\sup_{\|\boldsymbol{b}\|\leq h_n} \| I_1 \|_H = O(|\boldsymbol{C}|^{1/2}(\ln n)^{3/4}). \tag{4.7}$$

Combining (4.5), (4.6), and (4.7), we have

$$\begin{aligned}\sup_{\|\boldsymbol{b}\|\leq h_n} &|T^+(\boldsymbol{b})-\hat{T}^+(\boldsymbol{b})| \\ &= O\Big(|\boldsymbol{C}|^{1/2}(\ln n)^{5/4} + |\boldsymbol{C}|(\ln n)^{3/2} + (|\boldsymbol{C}|\ln n)^2\Big) \\ &= O(|\boldsymbol{C}|^{1/2}(\ln n)^{5/4}). \qquad (4.8)\end{aligned}$$

Finally, the theorem follows from (4.2), (4.3), and (4.8). □

REFERENCES

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **31**, 577–580.

Beran, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5**, 445–463.

Beran, R. J. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6**, 292–313.

Beran, R. J. (1982). Robust estimation in models for independent non-identically distributed data. *Ann. Statist.* **2**, 415–428.

Blackman, J. (1955). On the approximation of a distribution function by a empirical distribution. *Ann. Statist.* **26**, 256–267.

Bolthausen, E. (1977). Convergence in distribution of a minimum distance estimator. *Metrika* **24**, 215–227.

Bychkova, T. V. (1986). Some probabilistic inequalities for sums of random elements. *Problemy Ustoichivosti Stokhasticheskikh Modelei* **Trudy Seminara**, 22-26.

Donoho, D. L. and Liu, R. C. (1988a). The "automatic" robustness of minimum distance functionals. *Ann. Statist.* **16**, 552-586.

Donoho, D. L. and Liu, R. C. (1988b). Pathologies of some minimum distance estimators. *Ann. Statist.* **16**, 587-608.

Ghosh, J. K. (1971). A new proof of Bahadur's representation of quantiles and an application. *Ann. Math. Statist.* **42**, 1957-1961.

Ghosh, M. and Sen, P. K. (1972). On bounded length confidence interval for the regression coefficient based on a class of rank statistics. *Sankhya, Ser. A.* **34**, 33–52.

Ghosh, J. K. and Sukathme, S. (1974). Bahadur representation of quantiles in non-regular cases. *Technical Report Statistical Laboratory Iowa State University.*

Haan, L. and Taconis-Haantjes E. (1979). On Bahadur's representation of sample quantiles. *Ann. Inst. Statist. Math., Part A* **31**, 299-308.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. A. S. A.* **58**, 13-30.

Koul, H. L. (1979). Weighted empirical processes and the regression model. *J. Ind. Statist. Assoc.* **17**, 83–91.

Koul, H. L. (1985a). Minimum distance estimation in multiple linear regression. *Sankhya, Ser. A.* **47**, 57–74.

Koul, H. L. (1985b). Minimum distance estimation in linear regression with unknown error distributions. *Statist. & Prob. Letters.* **3**, 1-8.

Koul, H. L. (1992). *Weighted empiricals and linear models.* IMS Lecture Notes-Monograph Series **21**.

Koul, H. L. and DeWet, T. (1983). Minimum distance estimation in a linear regression model. *Ann. Statist.* **11**, 921–932.

Koul, H. L. and Zhu, Z. (1991) Bahadur representations for some minimum distance estimators in linear models. A preprint.

Millar, P. W. (1981). Robust estimation via minimum distance methods. *Zeit fur Wahrscheinlichkeitstheorie* **55**, 73–89.

Millar, P. W. (1982). Optimal estimation of general regression function. *Ann. Statist.* **10**, 717–740.

Millar, P. W. (1984). A general approach to the optimality of minimum distance estimators. *Trans. Amer. Math. Soc.* **286**, 377–418.

Parr, W. (1981). Minimum distance estimation: a bibliography. *comm. Statist. Theor. Meth.* **A10**, 1205–1224.

Parr, W. and Schucany, W. R. (1979). Minimum distance and robust estimation. *J.A.S.A.* **75**, 616–624.

Pollard, D. (1980). The minimum distance method of testing. *Metrika* **27**, 43–70.

Serfling, R. J. (1980). *Approximation theorems in mathematical statistics.* Wiley and Sons, New York.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* Wiley and Sons, New York.

Williamson, M. A. (1979). Weighted empirical-type estimators of regression parameter. PH.D. Thesis, MSU.

Williamson, M. A. (1982). Cramer-vov Mises estimations of regression parameter: The rank analogue. *J. Mult. Anal.* **12**, 248–255.

Wolfowitz, J. (1953). Estimation by minimum distance method. *Ann. Inst. Stat. Math.* **5**, 9–23.

Wolfowitz, J. (1954). Estimation of the components of stochastic structures. *Proc. Nat. Acad. Sci.* **40**, 602–606.

Wolfowitz, J. (1957). Minimum distance estimation method. *Ann. Math. Statist.* **28**, 75–88.