Ann. Inst. Statist. Math.
Vol. 41, No. 3, 521-540 (1989)

SEQUENTIAL ESTIMATION IN REGRESSION MODELS USING ANALOGUES OF TRIMMED MEANS*

ADAM T. MARTINSEK

Department of Statistics, University of Illinois, 101 Illini Hall, 725 South Wright Street, Champaign, IL 61820, U.S.A.

(Received July 16, 1988; revised December 19, 1988)

Abstract. A sequential procedure is proposed for constructing a fixed-size confidence region for the parameters of a linear regression model. The procedure is based on certain regression analogues of trimmed means, as formulated by Welsh (1987, *Ann. Statist.*, **15**, 20–36), rather than least squares estimates. For error distributions with continuous, symmetric density and some moment higher than fourth finite, if the design points of the model are bounded, then the procedure is both asymptotically consistent and asymptotically efficient as the size of the region approaches zero.

Key words and phrases: Fixed-size confidence region, asymptotic efficiency, asymptotic consistency, residual, order statistics.

1. Introduction and summary

Consider the general linear model

$$y_i = \mathbf{x}'_i\beta + \varepsilon_i , \tag{1.1}$$

$i = 1, 2, \ldots$, where each $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ is a known p-vector of design points, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is an unknown p-vector of parameters, $y_1, y_2, \ldots$ are the observed responses, and $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. with distribution function F and continuous density f that is positive on the support of F and symmetric about zero. Assume further that $x_{i1} = 1$ for all i, so that β_1 is an intercept.

Estimates of $\boldsymbol{\beta}$ that are designed to be regression analogues of trimmed means in location models have been proposed and analyzed by Koenker and Bassett (1978), Ruppert and Carroll (1980) and Welsh (1987). The formulations due to Koenker and Bassett and to Welsh have been shown

*Research supported in part by the National Science Foundation under Grants DMS 85-03321 and 88-02556 and by the Air Force under Grant AFOSR-87-0041.

(by Ruppert and Carroll and by Welsh, respectively) to have limiting distributional properties that are similar to those of trimmed means. This paper considers the problem of constructing fixed-size confidence ellipsoids for $\boldsymbol{\beta}$ using such estimates, specifically the version proposed by Welsh.

Let $\hat{\boldsymbol{\beta}}_n$ be a preliminary estimate of $\boldsymbol{\beta}$ based on the first n observations, and let

$$e_i(\hat{\boldsymbol{\beta}}_n) = y_i - \boldsymbol{x}'_i \hat{\boldsymbol{\beta}}_n ,$$

$i = 1,\ldots, n$, be the residuals from $\hat{\boldsymbol{\beta}}_n$. Define

$$e_{n1}(\hat{\boldsymbol{\beta}}_n) \le e_{n2}(\hat{\boldsymbol{\beta}}_n) \le \cdots \le e_{nn}(\hat{\boldsymbol{\beta}}_n)$$

to be the ordered residuals. Welsh's estimate, with trimming proportion α, $0 < \alpha < 1/2$, is defined essentially as follows. Put

$$\begin{aligned} \xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) &= e_{n[\alpha n]}(\hat{\boldsymbol{\beta}}_n) , \\ \xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n) &= e_{n,n-[\alpha n]}(\hat{\boldsymbol{\beta}}_n) , \end{aligned}$$

and

$$\begin{aligned} J_i &= I\{e_i(\hat{\boldsymbol{\beta}}_n) < \xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n)\} , \\ K_i &= I\{\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) \le e_i(\hat{\boldsymbol{\beta}}_n) \le \xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n)\} , \\ L_i &= I\{e_i(\hat{\boldsymbol{\beta}}_n) > \xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n)\} , \end{aligned} \tag{1.2}$$

$i = 1,\ldots, n$, where I denotes the indicator function. The estimate is then

$$\begin{aligned} \hat{\boldsymbol{\beta}}_n(\alpha) = \boldsymbol{A}_n^- \sum_{i=1}^{n} \boldsymbol{x}_i[\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n)(J_i - \alpha) \\ + y_i K_i + \xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n)(L_i - \alpha)] , \end{aligned} \tag{1.3}$$

where $\boldsymbol{A}_n^-$ is any generalized inverse of

$$\boldsymbol{A}_n = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}'_i K_i .$$

Welsh's proposal differs slightly from this, in that the endpoints of the ordered residuals at which Winsorizing takes place are defined somewhat differently. However, the estimate defined in (1.3) has the same asymptotic properties as Welsh's estimate and is a little easier to work with in the present setting.

Welsh (1987) shows that under the assumptions above, if

$$n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \tag{1.4}$$

is bounded in probability and

$$n^{-1}(\boldsymbol{X}_n'\boldsymbol{X}_n) = n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' \to \boldsymbol{\Gamma} , \tag{1.5}$$

as $n \to \infty$, where $\boldsymbol{X}_n$ is the $n \times p$ matrix with (i,j) entry x_{ij} and $\boldsymbol{\Gamma}$ is positive definite, then

$$n^{1/2}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}) \underset{d}{\to} N(0, \sigma^2(\alpha)\boldsymbol{\Gamma}^{-1}) , \tag{1.6}$$

as $n \to \infty$, where

$$\sigma^2(\alpha) = (1-2\alpha)^{-2}\left[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x^2 f(x)dx + 2\alpha(F^{-1}(\alpha))^2\right] . \tag{1.7}$$

Note that the condition

$$\sum_{i=1}^{n} x_{ij} = 0 ,$$

$j = 2,\ldots,p$ is not needed here, because f is symmetric about zero. In the rest of the paper $\hat{\boldsymbol{\beta}}_n$ will be the least squares estimate, assumed to be unique.

Based on (1.6), if $\sigma^2(\alpha)$ is known, for large n and $0 < \gamma < 1$, the ellipsoid

$$\{\boldsymbol{\Psi}: (\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\Psi})'\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\Psi}) \le \sigma^2(\alpha)\chi^2_\gamma(p)n^{-1}\} \tag{1.8}$$

will be an approximate $100(1-\gamma)\%$ confidence region for $\boldsymbol{\beta}$, where $\chi^2_\gamma(p)$ is the upper γ point of a chi-square distribution with p degrees of freedom. When $\sigma^2(\alpha)$ is unknown, as is typically the case, the results of Welsh (1987) show that it can be estimated consistently by

$$\begin{aligned} R_n^2(\alpha) = (1-2\alpha)^{-2}\Bigg[& (n-p)^{-1} \sum_{i=1}^{n} (e_i(\hat{\boldsymbol{\beta}}_n) - \bar{e}_K)^2 K_i \\ & + \alpha(\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) - \bar{e}_K)^2 + \alpha(\xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n) - \bar{e}_K)^2 \Bigg] , \end{aligned} \tag{1.9}$$

where

$$\bar{e}_K = (n - 2[\alpha n])^{-1} \sum_{i=1}^{n} e_i(\hat{\boldsymbol{\beta}}_n)K_i .$$

Hence approximate confidence ellipsoids for $\boldsymbol{\beta}$ may be obtained from (1.8) by plugging in $R_n^2(\alpha)$ for $\sigma^2(\alpha)$.

Suppose now that one wishes to construct a confidence ellipsoid for $\boldsymbol{\beta}$ that is of form (1.8), i.e., its shape is specified by $\boldsymbol{\Gamma}$, but of fixed size. In other words, one would like to use an ellipsoid of the form

$$\{\boldsymbol{\Psi}: (\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\Psi})'\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\Psi}) \leq d\}, \tag{1.10}$$

where $d > 0$ is given. If (1.10) is to be an approximate $100(1-\gamma)\%$ confidence region for $\boldsymbol{\beta}$, the smallest fixed sample size n that will work (asymptotically, as $d \to 0$) is

$$n_0 = n_0(d) \approx d^{-1}\sigma^2(\alpha)\chi_\gamma^2(p). \tag{1.11}$$

If $\sigma^2(\alpha)$ is unknown, then the sample size n_0 cannot be used. In this case, (1.11) suggests using the stopping rule

$$T = T_d = \inf\{n \geq 2: n \geq d^{-1}\chi_\gamma^2(p)(R_n^2(\alpha) + n^{-1})\} \tag{1.12}$$

to determine the sample size, and forming the confidence ellipsoid

$$\{\boldsymbol{\Psi}: (\hat{\boldsymbol{\beta}}_T(\alpha) - \boldsymbol{\Psi})'\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}_T(\alpha) - \boldsymbol{\Psi}) \leq d\}, \tag{1.13}$$

once sampling is terminated. Previous work on fixed-size confidence regions in regression models, under the assumption of normally distributed errors and using least squares estimates, has been done by Gleser (1965), Albert (1966), Srivastava (1967, 1971) and Finster (1985). For pioneering work on fixed-size confidence intervals, see Chow and Robbins (1965).

The asymptotic performance of the sequential procedure with stopping rule T is summarized in the following theorem.

THEOREM 1.1. *In addition to the assumptions in the first paragraph, assume that f has a density f' that is continuous a.e. on*

$$[F^{-1}(\alpha - \varepsilon_0), F^{-1}(1 - \alpha + \varepsilon_0)] \quad \textit{for some} \quad \varepsilon_0 > 0, \tag{1.14}$$

that

$$|x_{ij}| \leq M, \tag{1.15}$$

for some $M < \infty$ and all i and j, and that

$$E[|\varepsilon_1|^{4+\rho}] < \infty \quad \textit{for some} \quad \rho > 0. \tag{1.16}$$

Then as $d \to 0$,

(1.17) $\qquad T_d/n_0(d) \to 1 \quad a.s.,$

(1.18) $\qquad E(T_d)/n_0(d) \to 1 \quad$ *(asymptotic efficiency)*,

and

(1.19) $\qquad P\{(\hat{\boldsymbol{\beta}}_T(\alpha) - \boldsymbol{\beta})'\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}_T(\alpha) - \boldsymbol{\beta}) \leq d\} \to 1 - \gamma$

(asymptotic consistency).

The proof of the theorem requires several lemmas. Two of these are given in Section 2, where it is shown among other things that under mild conditions $R_n^2(\alpha)$ is in fact a strongly consistent estimate of $\sigma^2(\alpha)$. These two lemmas together suffice to prove (1.17) and (1.18). The proof of (1.19) requires uniform continuity in probability of the coordinates of the sequence

$$n^{1/2}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}).$$

The analysis is rather delicate and appears in Section 3, along with the proof of Theorem 1.1.

The assumption that $|x_{ij}| \leq M$ for some $M < \infty$ and all i and j is not very restrictive, as this will be satisfied in almost all practical settings. The assumption that some moment of the error distribution higher than the fourth is finite is more bothersome, as one of the reasons for trimming in both location and regression models is the possibility of heavy-tailed error distributions. However, even when *all* the moments of the error distribution are finite, $\sigma^2(\alpha)$ for some choices of $\alpha \neq 0$ may be much smaller than the variance σ^2 of the ε_i's. This is true of contaminated normal distributions, the double exponential distribution, and various t distributions, for example. Since

$$n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow[d]{} N(0, \sigma^2\boldsymbol{\Gamma}^{-1})$$

as $n \to \infty$ for the least squares estimate $\hat{\boldsymbol{\beta}}_n$, the sequential procedures defined above will be more efficient than those based on least squares in such cases.

It would be of interest to extend these results to cover a wider class of preliminary estimators. It would also be nice to have similar results for estimators other than the Welsh estimates. Use of a more robust preliminary estimator might eliminate the need for the moment assumption (1.16), while use of an alternative to the Welsh estimate could achieve, for example, a high breakdown point in addition to asymptotic consistency and efficiency. The main obstacle to such extensions is verification of uniform continuity in probability, which is typically much more delicate than proving convergence in distribution. The least squares estimate and Welsh's estimate can both be represented exactly as weighted sums of random variables to which Kolmogorov's inequality can be applied, either conditionally or unconditionally. Other robust estimators, especially those with high breakdown point such as the repeated median and least median of squares, do not admit such a representation or even (in some cases) any closed-form representation at all. The technical difficulties associated with proving uniform continuity in probability for such estimators are formidable and appear at present to be intractable.

Finster (1985) discusses fixed-size confidence regions of general shape, not necessarily ellipsoidal, for the case of normally distributed errors and least squares estimates. The approach in the present paper can be used in this more general setting. Specifically, a suitable sequential procedure based on $\hat{\boldsymbol{\beta}}_n(\alpha)$ can be defined and shown to be asymptotically efficient and asymptotically consistent as the "precision" (as defined by Finster) goes to infinity. The details are similar to those below and are omitted.

2. Preliminary lemmas

Throughout this section and the rest of the paper, for any p-vector $\boldsymbol{v}$, $\|\boldsymbol{v}\|$ will denote

$$|v_1| + \cdots + |v_p| .$$

The first lemma addresses the question of almost sure convergence of $R_n^2(\alpha)$ to $\sigma^2(\alpha)$.

LEMMA 2.1. *Assume* (1.14) *and* (1.15) *hold. Assume also that*

$$E[|\varepsilon_1|^{2q}] < \infty \quad \textit{for some} \quad q > 1 .$$

Then for any $\delta > 0$, *as* $n \to \infty$,

$$n^{1/2-\delta}(R_n^2(\alpha) - \sigma^2(\alpha)) \to 0 \quad a.s. \tag{2.1}$$

PROOF. Let ε_{nl} denote the l-th order statistic among $\varepsilon_1, \dots, \varepsilon_n$. Because there are exactly l ε_i's less than or equal to ε_{nl}, and each $|x_{ij}| \le M$, there must be at least l $e_i(\hat{\boldsymbol{\beta}}_n)$'s less than or equal to $\varepsilon_{nl} + M\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$. Hence $e_{nl}(\hat{\boldsymbol{\beta}}_n) \le \varepsilon_{nl} + M\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$. Similarly, $\varepsilon_{nl} \le e_{nl}(\hat{\boldsymbol{\beta}}_n) + M\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$, so that

$$|e_{nl}(\hat{\boldsymbol{\beta}}_n) - \varepsilon_{nl}| \le M\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| \qquad \text{for} \quad 1 \le l \le n . \tag{2.2}$$

By Theorem 1 of Lai and Wei (1982), in view of (1.5),

$$(2.3) \qquad \|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\| = O((n/\log(n))^{-1/2}) \quad \text{a.s.}$$

Because $\bar{e}_K$ is the average of the $n - 2[\alpha n]$ inner ordered residuals, from (2.2) and (2.3),

$$(2.4) \qquad |\bar{e}_K - m_n(\alpha)| = O((n/\log(n))^{-1/2}) \quad \text{a.s.},$$

where $m_n(\alpha)$ is the α-trimmed mean of the ε_i's,

$$m_n(\alpha) = (n - 2[\alpha n])^{-1} \sum_{l=[\alpha n]+1}^{n-[\alpha n]} \varepsilon_{nl}.$$

By similar reasoning, using the facts that

$$\varepsilon_{n[\alpha n]} \to F^{-1}(\alpha) \qquad \text{a.s.},$$

$$\varepsilon_{n,n-[\alpha n]} \to F^{-1}(1-\alpha) \qquad \text{a.s.},$$

and

$$m_n(\alpha) \to 0 \qquad \text{a.s.}$$

(see (2.8) of Martinsek (1984)),

$$\begin{aligned}(2.5) \qquad R_n^2(\alpha) &= (1-2\alpha)^{-2}\Big[(n-p)^{-1}\sum_{i=1}^{n}(e_i(\hat{\boldsymbol{\beta}}_n) - \bar{e}_K)^2 K_i \\ &\qquad + \alpha(\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) - \bar{e}_K)^2 + \alpha(\xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n) - \bar{e}_K)^2\Big] \\ &= (1-2\alpha)^{-2}\Big[(n-p)^{-1}\sum_{l=[\alpha n]+1}^{n-[\alpha n]}(\varepsilon_{nl} - m_n(\alpha))^2 \\ &\qquad + \alpha(\varepsilon_{n[\alpha n]} - m_n(\alpha))^2 + \alpha(\varepsilon_{n,n-[\alpha n]} - m_n(\alpha))^2\Big] \\ &\quad + O((n/\log(n))^{-1/2}) \quad \text{a.s.}\end{aligned}$$

By Lemma 2 of Martinsek (1984), for any $\delta > 0$, the difference between the right-hand side of (2.5) and $\sigma^2(\alpha)$ is $O(n^{-1/2+\delta})$ a.s., which finishes the proof.

The next lemma deals with q-th moment convergence of $R_n^2(\alpha)$ to $\sigma^2(\alpha)$.

LEMMA 2.2. *Assume that* (1.14) *and* (1.15) *hold. Assume also that*

$$E[|\varepsilon_1|^{2q}] < \infty,$$

where $q \geq 1$. *Then as* $n \to \infty$,

$$E\{|R_n^2(\alpha) - \sigma^2(\alpha)|^q\} = O(n^{-q/2}) . \tag{2.6}$$

PROOF. By a result of Brown (1971) (see also Chow and Teicher (1978), p. 398), since all $|x_{ij}| \leq M$ and (1.5) holds with Γ positive definite,

$$E\{\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|^{2q}\} = O(n^{-q}) . \tag{2.7}$$

By Jensen's inequality, (2.2) and (2.7),

$$\begin{aligned} E\{|\bar{e}_K - m_n(\alpha)|^{2q}\} &\leq E\left\{ (n - 2[\alpha n])^{-1} \sum_{l=[\alpha n]+1}^{n-[\alpha n]} |e_{nl}(\hat{\boldsymbol{\beta}}_n) - \varepsilon_{nl}|^{2q} \right\} \\ &= O(n^{-q}) . \end{aligned} \tag{2.8}$$

Similarly, by (2.2), (2.7), (2.8) and (3.14), (3.24) of Martinsek (1984), for any $l = [\alpha n] + 1, \ldots, n - [\alpha n]$,

$$\begin{aligned} &E\{|(e_{nl}(\hat{\boldsymbol{\beta}}_n) - \bar{e}_K)^2 - (\varepsilon_{nl} - m_n(\alpha))^2|^q\} \\ &\quad \leq E^{1/2}\{|e_{nl}(\hat{\boldsymbol{\beta}}_n) - \varepsilon_{nl} + m_n(\alpha) - \bar{e}_K|^{2q}\} \\ &\qquad \times E^{1/2}\{|e_{nl}(\hat{\boldsymbol{\beta}}_n) + \varepsilon_{nl} - m_n(\alpha) - \bar{e}_K|^{2q}\} \\ &\quad = O(n^{-q/2}) . \end{aligned} \tag{2.9}$$

Combining (2.9) with another application of Jensen's inequality and (3.29) of Martinsek (1984) yields the desired result.

COROLLARY 2.1. *Assume that* (1.14) *and* (1.15) *hold. If* $E[|\varepsilon_1|^{4+\rho}] < \infty$ *for some* $\rho > 0$, *then*

$$\{dT_d\colon d \leq 1\} \quad \textit{is uniformly integrable} . \tag{2.10}$$

PROOF. For $K > \chi_\gamma^2(p)(2\sigma^2(\alpha) + 1) + 1$, by Lemma 2.2,

$$\begin{aligned} \sup_{d\leq 1} P[dT_d > K] &\leq \sup_{d\leq 1} P[R^2_{[Kd^{-1}]}(\alpha) > [Kd^{-1}]d/\chi_\gamma^2(p) - 1] \\ &\leq \sup_{d\leq 1} P[R^2_{[Kd^{-1}]}(\alpha) > 2\sigma^2(\alpha)] \\ &\leq \sup_{d\leq 1} P[|R^2_{[Kd^{-1}]}(\alpha) - \sigma^2(\alpha)| > \sigma^2(\alpha)] \\ &\leq O(1) \sup_{d\leq 1} E\{|R^2_{[Kd^{-1}]}(\alpha) - \sigma^2(\alpha)|^{(2+\rho/2)}\} \\ &= O(1) \sup_{d\leq 1} [Kd^{-1}]^{-(1+\rho/4)} \end{aligned}$$

$$= O(K^{-1-p/4}),$$

from which the corollary follows.

3. Uniform continuity in probability and the proof of the theorem

A sequence $\{z_n: n \geq 1\}$ of random variables is said to be uniformly continuous in probability (u.c.i.p.) if for every $\varepsilon > 0$, if $\delta > 0$ is sufficiently small and n is sufficiently large,

$$P\left\{ \max_{0 \leq k \leq \delta_n} |z_{n+k} - z_n| \geq \varepsilon \right\} < \varepsilon \tag{3.1}$$

(see Woodroofe (1982), p. 10). The following proposition deals with uniform continuity in probability of the coordinates of $n^{1/2}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta})$.

PROPOSITION 3.1. *Assume that* (1.14) *and* (1.15) *hold. Assume in addition that*

$$E[|\varepsilon_1|^{2q}] < \infty \quad \textit{for some} \quad q > 1.$$

Then for each $j = 1, \ldots, p$,

$$\{n^{1/2}((\hat{\boldsymbol{\beta}}_n(\alpha))_j - \boldsymbol{\beta}_j): n \geq 1\} \quad \textit{is} \quad \textit{u.c.i.p.} \tag{3.2}$$

The proof of Proposition 3.1 depends on a series of lemmas, the first of which is in effect an almost sure version of part of Lemma A.3 of Ruppert and Carroll (1980). For any p-vector Δ, define

$$M_n^{\alpha}(\Delta) = n^{-1/2} \sum_{i=1}^{n} [\alpha - I\{\varepsilon_i \leq \boldsymbol{x}_i' \Delta n^{-1/2} + F^{-1}(\alpha)\}] \tag{3.3}$$

and

$$M_n^{1-\alpha}(\Delta) = n^{-1/2} \sum_{i=1}^{n} [(1-\alpha) - I\{\varepsilon_i \leq \boldsymbol{x}_i' \Delta n^{-1/2} + F^{-1}(1-\alpha)\}]. \tag{3.4}$$

Also, let $\boldsymbol{\Gamma}_1$ be the first row of $\boldsymbol{\Gamma}$.

LEMMA 3.1. *Assume that* (1.14) *and* (1.15) *hold. For any* $L > 0$, *as* $n \to \infty$,

$$\sup_{0 \leq \|\Delta\| \leq L} |M_n^{\alpha}(\Delta) - M_n^{\alpha}(0) + f(F^{-1}(\alpha)) \boldsymbol{\Gamma}_1 \Delta| \to 0 \qquad \textit{a.s.} \tag{3.5}$$

and

$$\sup_{0 \le \|\Delta\| \le L} |M_n^{1-\alpha}(\Delta) - M_n^{1-\alpha}(0) + f(F^{-1}(1-\alpha))\boldsymbol{\Gamma}_1\Delta| \to 0 \quad a.s. \tag{3.6}$$

PROOF. Let

$$S = \{\Delta : \Delta = (k_1(L+1)/n, \ldots, k_p(L+1)/n), \quad k_1, \ldots, k_p = 0, \pm 1, \ldots, \pm n\}. \tag{3.7}$$

For any Δ with $0 \le \|\Delta\| \le L$, if $n \ge L+1$, there exists $\tilde{\Delta} \in S$ with $|\Delta_j - \tilde{\Delta}_j| \le (L+1)/n, j = 1, \ldots, p$. Because each $|x_{ij}| \le M$ and all $x_{i1} = 1$, we have

$$\begin{aligned} M_n^\alpha(\Delta) &\ge n^{-1/2} \sum_{i=1}^n [\alpha - I\{\varepsilon_i \le \boldsymbol{x}_i'\tilde{\Delta} n^{-1/2} + (L+1)Mpn^{-3/2} + F^{-1}(\alpha)\}] \\ &= M_n^\alpha(\tilde{\Delta} + \boldsymbol{e}Mp(L+1)/n), \end{aligned} \tag{3.8}$$

where $\boldsymbol{e} = (1, 0, \ldots, 0)'$. Similarly,

$$M_n^\alpha(\Delta) \le M_n^\alpha(\tilde{\Delta} - \boldsymbol{e}Mp(L+1)/n). \tag{3.9}$$

It follows that

$$\begin{aligned} & M_n^\alpha(\tilde{\Delta} + \boldsymbol{e}Mp(L+1)/n) - M_n^\alpha(0) + f(F^{-1}(\alpha))\boldsymbol{\Gamma}_1\tilde{\Delta} \\ &\quad \le M_n^\alpha(\Delta) - M_n^\alpha(0) + f(F^{-1}(\alpha))\boldsymbol{\Gamma}_1\Delta + O(n^{-1}) \\ &\quad \le M_n^\alpha(\tilde{\Delta} - \boldsymbol{e}Mp(L+1)/n) - M_n^\alpha(0) + f(F^{-1}(\alpha))\boldsymbol{\Gamma}_1\tilde{\Delta} \\ &\qquad + O(n^{-1}). \end{aligned} \tag{3.10}$$

Hence, to prove (3.5) it suffices to show

$$\sup_{\Delta \in S} \|M_n^\alpha(\Delta + \boldsymbol{e}Mp(L+1)/n) - M_n^\alpha(0) + f(F^{-1}(\alpha))\boldsymbol{\Gamma}_1\Delta\| \to 0 \quad \text{a.s.} \tag{3.11}$$

and

$$\sup_{\Delta \in S} \|M_n^\alpha(\Delta - \boldsymbol{e}Mp(L+1)/n) - M_n^\alpha(0) + f(F^{-1}(\alpha))\boldsymbol{\Gamma}_1\Delta\| \to 0 \quad \text{a.s.} \tag{3.12}$$

Fix $\Delta \in S$. In view of (1.5), as $n \to \infty$,

$$(3.13)\qquad \begin{aligned} &|M_n^\alpha(\Delta + \boldsymbol{e}Mp(L+1)/n) - M_n^\alpha(0) + f(F^{-1}(\alpha))\Gamma_1\Delta| \\ &\quad \le \left| n^{-1/2} \sum_{i=1}^{n} \eta_i(\Delta) \right| + o(1)\,, \end{aligned}$$

uniformly in Δ, where

$$(3.14)\qquad \begin{aligned} \eta_i(\Delta) = {}& I\{\varepsilon_i \le F^{-1}(\alpha)\} \\ &- I\{\varepsilon_i \le \boldsymbol{x}_i'\Delta n^{-1/2} + Mp(L+1)n^{-3/2} + F^{-1}(\alpha)\} \\ &- [\alpha - F(\boldsymbol{x}_i'\Delta n^{-1/2} + Mp(L+1)n^{-3/2} + F^{-1}(\alpha))]\,. \end{aligned}$$

From (3.13), for n sufficiently large and $\varepsilon > 0$,

$$(3.15)\qquad \begin{aligned} &P[|M_n^\alpha(\Delta + \boldsymbol{e}Mp(L+1)/n) - M_n^\alpha(0) + f(F^{-1}(\alpha))\Gamma_1\Delta| > \varepsilon] \\ &\quad \le P\left[\left| \sum_{i=1}^{n} \eta_i(\Delta) \right| > \varepsilon n^{1/2}/2 \right]. \end{aligned}$$

Simple computations show that

$$(3.16)\qquad E(\eta_i(\Delta)) = 0$$

and

$$(3.17)\qquad E(\eta_i^2(\Delta)) = O(n^{-1/2})$$

as $n \to \infty$, uniformly in Δ, so for some $K > 0$ and all $\Delta \in S$, $n \ge 1$,

$$(3.18)\qquad \sum_{i=1}^{n} E(\eta_i^2(\Delta)) \le K n^{1/2}\,.$$

Using the Kolmogorov exponential bounds (see, e.g., Chow and Teicher (1978), Lemma 1, p. 338), we obtain from (3.16) and (3.18) that for some $K' > 0$ and all $\Delta \in S$, $n \ge 1$,

$$(3.19)\qquad P\left[\left| \sum_{i=1}^{n} \eta_i(\Delta) \right| > \varepsilon n^{1/2}/2 \right] \le K' \exp\left[-\varepsilon n^{1/4}/(2K^{1/2}) \right].$$

It follows from (3.19) that

$$(3.20)\qquad \begin{aligned} \sum_{n=1}^{\infty} P\Big[&\sup_{\Delta \in S} |M_n^\alpha(\Delta + \boldsymbol{e}Mp(L+1)/n) - M_n^\alpha(0) \\ &+ f(F^{-1}(\alpha))\Gamma_1\Delta| > \varepsilon \Big] \end{aligned}$$

$$\leq \sum_{n=1}^{\infty} \sum_{\Delta \in S} P\left[\left| \sum_{n=1}^{n} \eta_i(\Delta) \right| > \varepsilon n^{1/2}/2 \right]$$

$$\leq K' \sum_{n=1}^{\infty} \#(S) \exp\left[-\varepsilon n^{1/4}/(2K^{1/2}) \right]$$

$$= K' \sum_{n=1}^{\infty} (2n+1)^p \exp\left[-\varepsilon n^{1/4}/(2K^{1/2}) \right]$$

$$< \infty .$$

Because (3.20) holds for every $\varepsilon > 0$, by the Borel-Cantelli Lemma, (3.11) is established. Similar reasoning proves (3.12) and completes the proof of (3.5).

The argument for (3.6) is exactly analogous.

LEMMA 3.2. *Assume that* (1.14) *and* (1.15) *hold. Assume also that*

$$E[|\varepsilon_1|^{2q}] < \infty \quad \textit{for some} \quad q > 1 .$$

Then

$$\{n^{1/2}(\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) - F^{-1}(\alpha)) : n \geq 1\} \quad \textit{is} \quad \textit{u.c.i.p.} \tag{3.21}$$

and

$$\{n^{1/2}(\xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n) - F^{-1}(1-\alpha)) : n \geq 1\} \quad \textit{is} \quad \textit{u.c.i.p.} \tag{3.22}$$

PROOF. For any $L > 0$,

$$\begin{aligned} P\Big[\max_{0 \leq k \leq \delta n} &|(n+k)^{1/2}(\xi_{(n+k)\alpha}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(\alpha)) \\ &\qquad - n^{1/2}(\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) - F^{-1}(\alpha))| > \varepsilon \Big] \\ \leq P\Big[&\max_{0 \leq k \leq \delta n} (n+k)^{1/2} \| \hat{\boldsymbol{\beta}}_{n+k} - \boldsymbol{\beta} \\ &\quad + \boldsymbol{e}(\xi_{(n+k)\alpha}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(\alpha)) \| > L \Big] \\ + P\Big[&\max_{0 \leq k \leq \delta n} |(n+k)^{1/2}(\xi_{(n+k)\alpha}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(\alpha)) \\ &\quad - n^{1/2}(\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) - F^{-1}(\alpha))| > \varepsilon , \\ &\max_{0 \leq k \leq \delta n} (n+\mathrm{k})^{1/2} \| \hat{\boldsymbol{\beta}}_{n+k} - \boldsymbol{\beta} \\ &\quad + \boldsymbol{e}(\xi_{(n+k)\alpha}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(\alpha)) \| \leq L \Big] . \end{aligned} \tag{3.23}$$

From (4.1.8) of Shu (1987), the proof of Lemma 2 in Martinsek (1987), and convergence in distribution, if L and n are sufficiently large and δ is sufficiently small, the first probability on the right-hand side of (3.23) is less than $\varepsilon/2$. As for the second probability on the right-hand side of (3.23), it is smaller than

$$\begin{aligned}(3.24)\qquad & P\Big[\max_{0\le k\le\delta n} |(n+k)^{1/2}(\xi_{(n+k)\alpha}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(\alpha)) - \zeta_{n+k}| > \varepsilon/3, \\ & \max_{0\le k\le\delta n} (n+k)^{1/2}\|\hat{\boldsymbol{\beta}}_{n+k} - \boldsymbol{\beta} \\ & + \boldsymbol{e}(\xi_{(n+k)\alpha}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(\alpha))\| \le L\Big] \\ & + P\Big[\max_{0\le k\le\delta n} |\zeta_{n+k} - \zeta_n| > \varepsilon/3\Big],\end{aligned}$$

where

$$\begin{aligned}\zeta_n = \{ & -M_n^\alpha(n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + \boldsymbol{e}n^{1/2}(\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) - F^{-1}(\alpha))) \\ & + M_n^\alpha(0)\}/f(F^{-1}(\alpha)) - \boldsymbol{\Gamma}_1 n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) .\end{aligned}$$

By Lemma 3.1, if δ is sufficiently small and n is sufficiently large, the first probability on the right-hand side of (3.24) is less than $\varepsilon/4$. Also, by the definition of $\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n)$,

$$\begin{aligned}(3.25)\qquad & M_n^\alpha(n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + \boldsymbol{e}n^{1/2}(\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) - F^{-1}(\alpha))) \\ & = n^{-1/2}\sum_{i=1}^{n} [\alpha - I\{\varepsilon_i \le \boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + \xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n)\}] \\ & = n^{-1/2}\sum_{i=1}^{n} [\alpha - I\{Y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n \le \xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n)\}] \\ & = O(n^{-1/2}) .\end{aligned}$$

Moreover,

$$(3.26)\qquad \{\boldsymbol{\Gamma}_1 n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \colon n \ge 1\}$$

is u.c.i.p., as in the proof of Lemma 2 of Martinsek (1987), and

$$(3.27)\qquad \{M_n^\alpha(0)\} = \left\{ n^{-1/2}\sum_{i=1}^{n} [\alpha - I\{\varepsilon_i \le F^{-1}(\alpha)\}] \right\}$$

is u.c.i.p., because the $\alpha - I\{\varepsilon_i \le F^{-1}(\alpha)\}$ are i.i.d. and have mean zero. Combining (3.25)-(3.27),

$$\{\zeta_n : n \geq 1\}$$

is u.c.i.p., and hence if δ is sufficiently small and n is sufficiently large, the second probability on the right-hand side of (3.24) is less than $\varepsilon/4$. This completes the proof of (3.21). The proof of (3.22) is similar.

PROOF OF PROPOSITION 3.1. We will show first that each coordinate of $n^{-1}\boldsymbol{A}_n$ is u.c.i.p. By (A.10) of Ruppert and Carroll (1980),

$$n^{-1}\boldsymbol{A}_n \underset{P}{\to} (1-2\alpha)\boldsymbol{\Gamma} \tag{3.28}$$

as $n \to \infty$. Fix $\varepsilon > 0$. For any j, $l = 1,\ldots,p$, δ sufficiently small and n sufficiently large,

$$\begin{aligned}
(3.29)\quad & P\Big[\max_{0\le k\le \delta n} |n^{-1}(\boldsymbol{A}_n)_{jl} - (n+k)^{-1}(\boldsymbol{A}_{n+k})_{jl}| > \varepsilon\Big] \\
&= P\Big[\max_{0\le k\le \delta n} \Big| n^{-1}\sum_{i=1}^{n} x_{ij}x_{il}K_i^n - (n+k)^{-1}\sum_{i=1}^{n+k} x_{ij}x_{il}K_i^{n+k}\Big| > \varepsilon\Big] \\
&\le P\Big[(\delta/(n+\delta n))\Big|\sum_{i=1}^{n} x_{ij}x_{il}K_i^n\Big| > \varepsilon n/2\Big] \\
&\quad + P\Big[\max_{0\le k\le \delta n}\Big|\sum_{i=1}^{n} x_{ij}x_{il}K_i^n - \sum_{i=1}^{n+k} x_{ij}x_{il}K_i^{n+k}\Big| > \varepsilon n/2\Big],
\end{aligned}$$

where K_i^n is K_i defined in (1.2) and K_i^{n+k} is K_i with n replaced by $n+k$. By (3.28), if δ is sufficiently small and n is sufficiently large, the first probability on the right-hand side of (3.29) is less than $\varepsilon/2$. The second probability on the right-hand side of (3.29) is smaller than

$$\begin{aligned}
(3.30)\quad & P\Big[\max_{0\le k\le \delta n}\Big|\sum_{i=1}^{n} x_{ij}x_{il}(K_i^n - K_i^{n+k})\Big| > \varepsilon n/4\Big] \\
&\quad + P\Big[\max_{0\le k\le \delta n}\Big|\sum_{i=n+1}^{n+k} x_{ij}x_{il}K_i^{n+k}\Big| > \varepsilon n/4\Big].
\end{aligned}$$

For δ less than $\varepsilon/4M^2$, the second probability in (3.30) vanishes. The remaining term in (3.30) is smaller than

$$\begin{aligned}
(3.31)\quad & P\Big[M\max_{0\le k\le \delta n}\|\hat{\boldsymbol{\beta}}_n - \hat{\boldsymbol{\beta}}_{n+k}\| \\
&\quad + \max_{0\le k\le \delta n}|\zeta_{(n+k)\alpha}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(\alpha)| \\
&\quad + \max_{0\le k\le \delta n}|\zeta_{(n+k)(1-\alpha)}(\hat{\boldsymbol{\beta}}_{n+k}) - F^{-1}(1-\alpha)| > n^{-1/4}\Big] \\
&\quad + P\Big[\max_{0\le k\le \delta n}\Big|\sum_{i=1}^{n} x_{ij}x_{il}(K_i^n - K_i^{n+k})\Big| > \varepsilon n/4,
\end{aligned}$$

$$\max_{0\le k\le \delta n} M\|\hat{\boldsymbol\beta}_n - \hat{\boldsymbol\beta}_{n+k}\| + \max_{0\le k\le \delta n} |\xi_{(n+k)\alpha}(\hat{\boldsymbol\beta}_{n+k}) - F^{-1}(\alpha)|$$
$$+ \max_{0\le k\le \delta n} |\xi_{(n+k)(1-\alpha)}(\hat{\boldsymbol\beta}_{n+k}) - F^{-1}(1-\alpha)| \le n^{-1/4}\Big].$$

By Lemma 3.2 in this paper and the proof of Lemma 2 in Martinsek (1987), the first probability in (3.31) can be made less than $\varepsilon/4$ if δ is sufficiently small and n is sufficiently large. Because each $|x_{ij}| \le M$ and $|x_{il}| \le M$, since $K_i^n \ne K_i^{n+k}$ implies

$$|\varepsilon_i - F^{-1}(\alpha)| \le 2n^{-1/4}$$

or else

$$|\varepsilon_i - F^{-1}(1-\alpha)| \le 2n^{-1/4},$$

the second probability in (3.31) is smaller than

$$\begin{aligned} &P\Big[\sum_{i=1}^{n} I\{|\varepsilon_i - F^{-1}(\alpha)| \le 2n^{-1/4} \\ &\qquad \text{or } |\varepsilon_i - F^{-1}(1-\alpha)| \le 2n^{-1/4}\} > \varepsilon n/4M^2\Big] \\ &\qquad \le (4M^2/\varepsilon n)E\Big[\sum_{i=1}^{n} I\{|\varepsilon_i - F^{-1}(\alpha)| \le 2n^{-1/4} \\ &\qquad\qquad\qquad \text{or } |\varepsilon_i - F^{-1}(1-\alpha)| \le 2n^{-1/4}\}\Big] \\ &\qquad = O(n^{-1/4}) \to 0, \end{aligned}$$

as $n \to \infty$. In particular, the second probability in (3.31) can be made smaller than $\varepsilon/4$, and u.c.i.p. of each coordinate of $n^{-1}A_n$ follows. By (3.28),

$$P[n^{-1}A_n \text{ invertible}] \to 1$$

as $n \to \infty$, and hence by u.c.i.p. of the coordinates of $n^{-1}A_n$,

$$P[(n+k)^{-1}A_{n+k} \text{ invertible, for all } 0 \le k \le \delta n] \to 1 \tag{3.32}$$

also. U.c.i.p. of the coordinates of nA_n^- now follows from (3.32), stochastic boundedness of the coordinates of $n^{-1}A_n$ (see (3.28)), and Lemma 1.4 of Woodroofe (1982). Define

$$r_{i,n} = e_i(\hat{\boldsymbol\beta}_n) = Y_i - \boldsymbol{x}_i'\hat{\boldsymbol\beta}_n$$

and

$$
\begin{aligned}
W_{n,\alpha}(x) = {} & \xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n)(I\{x < \xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n)\} - \alpha) \\
& + xI\{\xi_{n\alpha}(\hat{\boldsymbol{\beta}}_n) \le x \le \xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n)\} \\
& + \xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n)(I\{x > \xi_{n(1-\alpha)}(\hat{\boldsymbol{\beta}}_n)\} - \alpha) .
\end{aligned}
$$

By the proof of Lemma 2 of Martinsek (1987), for $j = 1,\dots,p$,

$$\{n^{1/2}((\hat{\boldsymbol{\beta}}_n)_j - \boldsymbol{\beta}_j)\colon n \ge 1\}$$

is u.c.i.p. It therefore suffices to show

$$\{n^{1/2}((\hat{\boldsymbol{\beta}}_n(\alpha))_j - (\hat{\boldsymbol{\beta}}_n)_j)\colon n \ge 1\}$$

is u.c.i.p. Furthermore, in view of (3.32), it is enough to establish (3.1) with $z_n = n^{1/2}((\hat{\boldsymbol{\beta}}_n(\alpha))_j - (\hat{\boldsymbol{\beta}}_n)_j)$ under the assumption that $\boldsymbol{A}_n,\dots,\boldsymbol{A}_{n+[\delta n]}$ are invertible. Then for each $k = 0,\dots,[\delta n]$,

$$\hat{\boldsymbol{\beta}}_{n+k}(\alpha) - \hat{\boldsymbol{\beta}}_{n+k} = \boldsymbol{A}_{n+k}^{-1} \sum_{i=1}^{n+k} \boldsymbol{x}_i W_{n+k,\alpha}(r_{i,n+k}) . \tag{3.33}$$

By (3.33) and Lemma 1.4 of Woodroofe (1982), together with stochastic boundedness of the coordinates of $n\boldsymbol{A}_n^{-1}$ and the coordinates of

$$\sum_{i=1}^{n} \boldsymbol{x}_i W_{n,\alpha}(r_{i,n}) ,$$

it suffices to show that the sequence

$$n^{-1/2} \sum_{i=1}^{n} x_{ij} W_{n,\alpha}(r_{i,n})$$

is u.c.i.p. for $j = 1,\dots, p$. Now

$$
\begin{aligned}
P\Big[\max_{0\le k\le \delta n} \Big| & (n+k)^{-1/2} \sum_{i=1}^{n+k} x_{ij} W_{n+k,\alpha}(r_{i,n+k}) \\
& - n^{-1/2} \sum_{i=1}^{n} x_{ij} W_{n,\alpha}(r_{i,n}) \Big| > \varepsilon \Big] \\
\le P\Big[& \max_{0\le k\le \delta n} \Big| \sum_{i=1}^{n+k} x_{ij} W_{n+k,\alpha}(r_{i,n+k}) - \sum_{i=1}^{n} x_{ij} W_{n,\alpha}(r_{i,n}) \Big| > \varepsilon n^{1/2}/2 \Big] \\
& + P\Big[[(1+\delta)^{1/2} - 1] \Big| \sum_{i=1}^{n} x_{ij} W_{n,\alpha}(r_{i,n}) \Big| > \varepsilon n^{1/2}/2 \Big] .
\end{aligned} \tag{3.34}
$$

The second probability on the right-hand side of (3.34) can be made less than $\varepsilon/2$ if δ is sufficiently small and n is sufficiently large. It follows from

Lemma 3.2 that for every $\varepsilon_1 > 0$, if $\delta_1 > 0$ is sufficiently small,

$$P\Big[\max_{0\le k\le\delta_1 n} |n^{1/2}(\xi_{(n+k)\alpha}(\hat{\boldsymbol\beta}_{n+k}) - \xi_{n\alpha}(\hat{\boldsymbol\beta}_n))| + \max_{0\le k\le\delta_1 n} |n^{1/2}(\xi_{(n+k)(1-\alpha)}(\hat{\boldsymbol\beta}_{n+k}) - \xi_{n(1-\alpha)}(\hat{\boldsymbol\beta}_n))| > \varepsilon_1\Big] < \varepsilon_1 ,$$

and hence to make the first probability on the right-hand side of (3.34) less than $\varepsilon/2$ it will be enough to make

$$P\Big[\max_{0\le k\le\delta n}\Big|\sum_{i=1}^{n+k} x_{ij} W_{n,\alpha}(r_{i,n+k}) - \sum_{i=1}^{n} x_{ij} W_{n,\alpha}(r_{i,n})\Big| > \varepsilon n^{1/2}/4\Big] < \varepsilon/4 .$$

But

$$n^{1/2}|r_{i,n+k} - r_{i,n}| \le Mn^{1/2}\|\hat{\boldsymbol\beta}_n - \hat{\boldsymbol\beta}_{n+k}\| ,$$

and by u.c.i.p. of the coordinates of $n^{1/2}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta)$ (see Lemma 2 of Martinsek (1987)) with the fact that

$$|W_{n,\alpha}(x) - W_{n,\alpha}(y)| \le |x - y|$$

for every x and y, it suffices to show

$$P\Big[\max_{0\le k\le\delta n}\Big|\sum_{i=1}^{n+k} x_{ij} W_{n,\alpha}(r_{i,n}) - \sum_{i=1}^{n} x_{ij} W_{n,\alpha}(r_{i,n})\Big| = \max_{0\le k\le\delta n}\Big|\sum_{i=n+1}^{n+k} x_{ij} W_{n,\alpha}(r_{i,n})\Big| > \varepsilon n^{1/2}/8\Big] < \varepsilon/8 .$$

Given $\varepsilon_1,\ldots,\varepsilon_n$, the $W_{n,\alpha}(r_{i,n})$, $i = n+1,\ldots,n+[\delta n]$, are independent. By Kolmogorov's inequality and a result of Chow and Studden (1969),

$$\begin{aligned}
&P\Big[\max_{0\le k\le\delta n}\Big|\sum_{i=n+1}^{n+k} x_{ij}[W_{n,\alpha}(r_{i,n}) - E(W_{n,\alpha}(r_{i,n})|\varepsilon_1,\ldots,\varepsilon_n)]\Big| > \varepsilon n^{1/2}/8 \,|\, \varepsilon_1,\ldots,\varepsilon_n\Big] \\
&\le 64/(n\varepsilon^2)\sum_{i=n+1}^{n+[\delta n]} x_{ij}^2 \operatorname{Var}[W_{n,\alpha}(r_{i,n})|\varepsilon_1,\ldots,\varepsilon_n] \\
&= 64/(n\varepsilon^2)\sum_{i=n+1}^{n+[\delta n]} x_{ij}^2 \operatorname{Var}[W_{n,\alpha}(r_{i,n}) + \alpha\xi_{n\alpha}(\hat{\boldsymbol\beta}_n) + \alpha\xi_{n(1-\alpha)}(\hat{\boldsymbol\beta}_n)|\varepsilon_1,\ldots,\varepsilon_n] \\
&\le 64/(n\varepsilon^2)\sum_{i=n+1}^{n+[\delta n]} x_{ij}^2 \operatorname{Var}[\varepsilon_i + \mathbf{x}_i'(\boldsymbol\beta - \hat{\boldsymbol\beta}_n)|\varepsilon_1,\ldots,\varepsilon_n]
\end{aligned}$$

$$= 64/(n\varepsilon^2) \sum_{i=n+1}^{n+[\delta n]} x_{ij}^2 \operatorname{Var}(\varepsilon_i)$$

$$\leq 64M^2\delta n\sigma^2/(n\varepsilon^2) < \varepsilon/16 \quad \text{a.s.},$$

if δ is sufficiently small. Hence

$$P\left[\max_{0\leq k\leq \delta n}\left|\sum_{i=n+1}^{n+k} x_{ij}[W_{n,\alpha}(r_{i,n}) - E(W_{n,\alpha}(r_{i,n})|\varepsilon_1,\ldots,\varepsilon_n)]\right| > \varepsilon n^{1/2}/8\right] < \varepsilon/16 .$$

It remains to show that if δ is sufficiently small and n is sufficiently large,

$$P\left[\max_{0\leq k\leq \delta n}\left|\sum_{i=n+1}^{n+k} x_{ij}E(W_{n,\alpha}(r_{i,n})|\varepsilon_1,\ldots,\varepsilon_n)\right| > \varepsilon n^{1/2}/8\right] < \varepsilon/16 . \tag{3.35}$$

We have that for $i = n+1,\ldots,n+[\delta n]$,

$$\begin{aligned} E[W_{n,\alpha}(r_{i,n})|\varepsilon_1,\ldots,\varepsilon_n] &= \xi_{n\alpha}(\hat{\boldsymbol\beta}_n)[F(\mathbf{x}_i'(\hat{\boldsymbol\beta}_n - \boldsymbol\beta) + \xi_{n\alpha}(\hat{\boldsymbol\beta}_n)) - \alpha] \\ &+ \int_{\mathbf{x}_i'(\hat{\boldsymbol\beta}_n-\boldsymbol\beta)+\xi_{n\alpha}(\hat{\boldsymbol\beta}_n)}^{\mathbf{x}_i'(\hat{\boldsymbol\beta}_n-\boldsymbol\beta)+\xi_{n(1-\alpha)}(\hat{\boldsymbol\beta}_n)} [x - \mathbf{x}_i'(\hat{\boldsymbol\beta}_n - \boldsymbol\beta)]f(x)dx \\ &+ \xi_{n(1-\alpha)}(\hat{\boldsymbol\beta}_n)[1 - F(\mathbf{x}_i'(\hat{\boldsymbol\beta}_n - \boldsymbol\beta) + \xi_{n(1-\alpha)}(\hat{\boldsymbol\beta}_n)) - \alpha] . \end{aligned} \tag{3.36}$$

From (2.2),

$$\xi_{n\alpha}(\hat{\boldsymbol\beta}_n) = F^{-1}(\alpha) + O_p(n^{-1/2})$$

and

$$\xi_{n(1-\alpha)}(\hat{\boldsymbol\beta}_n) = F^{-1}(1-\alpha) + O_p(n^{-1/2}) .$$

It follows from this and (3.36) that

$$E[W_{n,\alpha}(r_{i,n})|\varepsilon_1,\ldots,\varepsilon_n] = O_p(n^{-1/2}) ,$$

which proves (3.35) and hence Proposition 3.1.

PROOF OF THEOREM 1.1. First note that since

$$T_d \geq (\chi_\gamma^2(p)d^{-1})^{1/2} ,$$

$T_d \to \infty$ a.s. as $d \to 0$. From Lemma 2.1,

$$R_n^2(\alpha) \to \sigma^2(\alpha) \qquad \text{a.s.}$$

as $n \to \infty$, so we have

$$R_{T_d}^2(\alpha) \to \sigma^2(\alpha) \qquad \text{a.s.} \tag{3.37}$$

and

$$R_{T_d-1}^2(\alpha) \to \sigma^2(\alpha) \qquad \text{a.s.} \tag{3.38}$$

as $d \to 0$. (1.17) now follows from (3.37), (3.38) and the defining relation for T_d. (1.18) is immediate from (1.17) and Corollary 2.1. Finally, by (1.6), Proposition 3.1, (1.17) and Anscombe's (1952) theorem, as $d \to 0$,

$$T_d^{1/2}(\hat{\boldsymbol{\beta}}_{T_d}(\alpha) - \boldsymbol{\beta}) \underset{d}{\to} N(0, \sigma^2(\alpha)\boldsymbol{\Gamma}^{-1}),$$

and (1.19) follows from this and (1.17).

References

Albert, A. (1966). Fixed size confidence ellipsoids for linear regression parameters, *Ann. Math. Statist.*, **37**, 1602–1630.

Anscombe, F. (1952). Large sample theory of sequential estimation, *Proc. Cambr. Philos. Soc.*, **48**, 600–607.

Brown, B. M. (1971). A note on convergence of moments, *Ann. Math. Statist.*, **42**, 777–779.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width confidence intervals for the mean, *Ann. Math. Statist.*, **36**, 457–462.

Chow, Y. S. and Studden, W. J. (1969). Monotonicity of the variance under truncation and variations of Jensen's inequality, *Ann. Math. Statist.*, **40**, 1106–1108.

Chow, Y. S. and Teicher, H. (1978). *Probability Theory*, Springer, New York.

Finster, M. (1985). Estimation in the general linear model when the accuracy is specified before data collection, *Ann. Statist.*, **13**, 663–675.

Gleser, L. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters, *Ann. Math. Statist.*, **36**, 463–467.

Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Ann. Statist.*, **10**, 154–166.

Martinsek, A. T. (1984). Sequential determination of estimator as well as sample size, *Ann. Statist.*, **12**, 533–550.

Martinsek, A. T. (1987). Sequential point estimation in regression models: The non-parametric case, Tech. Report, University of Illinois.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model, *J. Amer. Statist. Assoc.*, **75**, 828–838.

Shu, W. Y. (1987). Limit theorems for processes and stopping rules in adaptive sequential estimation, Ph.D. Dissertation, University of Illinois.

Srivastava, M. S. (1967). On fixed-width confidence bounds for regression parameters and the mean vector, *J. Roy. Statist. Soc. Ser. B*, **29**, 132–140.

Srivastava, M. S. (1971). On fixed-width confidence bounds for regression parameters, *Ann. Math. Statist.*, **42**, 1403–1411.

Welsh, A. H. (1987). The trimmed mean in the linear model, *Ann. Statist.*, **15**, 20–36.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.