Econometric Theory, **16**, 2000, 301–323. Printed in the United States of America.

LINEARIZATION OF RANDOMLY WEIGHTED EMPIRICALS UNDER LONG RANGE DEPENDENCE WITH APPLICATIONS TO NONLINEAR REGRESSION QUANTILES

KANCHAN MUKHERJEE
National University of Singapore

This paper discusses some asymptotic uniform linearity results of randomly weighted empirical processes based on long range dependent random variables. These results are subsequently used to linearize nonlinear regression quantiles in a nonlinear regression model with long range dependent errors, where the design variables can be either random or nonrandom. These, in turn, yield the limiting behavior of the nonlinear regression quantiles. As a corollary, we obtain the limiting behavior of the least absolute deviation estimator and the trimmed mean estimator of the parameters of the nonlinear regression model. Some of the limiting properties are in striking contrast with the corresponding properties of a nonlinear regression model under independent and identically distributed error random variables. The paper also discusses an extension of rank score statistic in a nonlinear regression model.

1. INTRODUCTION

The least absolute deviation method of estimation appears to be older than the least squares method of estimation in linear regression models, with origins dating back to the middle of the eighteenth century. Despite this, the asymptotic theory of the least absolute deviation estimator (LAD) (see (2.2), which follows, for the definition) has been developed only recently. Koenker and Bassett (1978) considered a linear regression model with nonrandom regressors and errors that are independent and identically distributed (i.i.d.) and defined an αth regression quantile (RQ) (see (2.2) with $h(u, X_{ni}) = u'X_{ni}$ for the definition), which is a multidimensional analogue of the $([n\alpha] + 1)$th-order statistic. They derived central limit theorems (CLT's) for the RQ's, which in particular

I thank the referees for their useful comments. I am grateful to Professor Donald W.K. Andrews for a very careful reading of the paper and for many suggestions leading to its improvement. I thank my colleague Professor A.J. Berrick for his help in polishing up the English of the paper. The revision of the paper was partially supported by Academic Research Grant RP3982708 from the National University of Singapore. Address correspondence to: Kanchan Mukherjee, Department of Statistics and Applied Probability, National University of Singapore, Republic of Singapore, 119260; e-mail: stamk@leonis.nus.edu.sg.

© 2000 Cambridge University Press 0266-4666/00 $9.50

established a CLT for the LAD. It turned out that the LAD is asymptotically more efficient than the least squares estimator (LSE) whenever the median is asymptotically more efficient than the mean as an estimator of the location in the one-sample location model.

Several ramifications of Koenker and Bassett's work were developed by a number of researchers during the past 20 years, either by assuming various kinds of dependence structure on the error random variables or by imposing different types of regularity conditions on the regressor variables. Bloomfield and Steiger (1983) and Pollard (1991) considered random regressors and the errors generated by stationary, ergodic, martingale differences to establish the CLT for the LAD. Portnoy (1991) and Koul and Mukherjee (1994) obtained asymptotic representations of RQ's when the errors are m-dependent and long range dependent (L.R.D.) (see (3.1), which follows, for the definition), respectively, in linear models with nonrandom regressors.

Another line of development, resulting from the attempt to extend the concept of the LAD and RQ's to nonlinear regression models, was carried out by Oberhofer (1982), Richardson and Bhattachariya (1987), Weiss (1991), and Jurečková and Procházka (1994), among others. They obtained the asymptotic normality of the LAD and nonlinear regression quantiles (see (2.2)) under various sets of regularity conditions on the design points with i.i.d. errors.

There is an increasing interest in stationary time series that exhibit long range dependence (characterized by hyperbolic decay of correlations). More specifically, L.R.D. models with a regression trend to arise in many fundamental applications of econometrics, business, and environmental studies. Asymptotic theory for the classical best linear unbiased estimator and the LSE in linear regression models with nonrandom regressor and L.R.D. errors was developed by Yajima (1991), and the corresponding theory for robust estimation procedures was developed by Beran (1991), Koul and Mukherjee (1993, 1994), and Mukherjee (1994, 1999b), among others. In view of the importance of random regressors (say, a linear state-space model with long range dependence as in Robinson, 1992) and the nonlinearity of the regression function, in this paper we discuss the asymptotic behavior of nonlinear regression quantiles and rank scores (see definitions (2.2) and (2.5)) in nonlinear regression models with random regressors, when the errors are L.R.D. Nonlinear quantiles and rank scores are useful in both estimation and testing problems, and so their asymptotics discussed in this paper pave the way for the statistical inference in these nonlinear models. Also, an interesting order statistics property of nonlinear quantiles is proved that shows that they are right analogues of the order statistics in nonlinear regression models.

Some very intriguing phenomena regarding the asymptotic behavior of nonlinear regression quantiles are observed under the L.R.D. errors setup. They are fundamentally different from what are generally observed under the i.i.d. errors setup and the L.R.D. errors setup with nonrandom regressors. For example, the rates of convergence of the different estimators are influenced by the nature of the randomness of the regressors and are not always $n^{1/2}$. Also, depending on the type of randomness, the rates of convergence for intercept and slope parameters may turn out to be different. Moreover, unlike the i.i.d. errors case, the limiting distributions, if they exist, are not always normal. For details, see Remarks 4.3–4.5.

There are several approaches to deriving the asymptotic distribution of the RQ in linear regression models. For example, in the case of the LAD, Amemiya (1982) and Bloomfield and Steiger (1983, Theorem 2, p. 44) used a smooth approximation of a function $\rho_{1/2}$ (defined following (2.2)), whereas Pollard (1991) used some convexity properties. Because of the nonlinearity of the regression function and the dependence among errors, these approaches may not work in our case. However, we use a linearization result (Theorem A.1) of a randomly weighted empirical process V_h (see the Appendix for definition) to obtain a Taylor-type expansion of the statistic $T(u,\alpha)$ (see Section 4 for definition) in Theorem 4.1(i). This expansion, coupled with a consistency assumption (4.1), gives us the desired result. This technique was used previously by Jurečková (1984), Portnoy (1991), and Koul and Mukherjee (1994), among others, in similar but simpler situations.

The remainder of the paper is organized as follows. The quantiles and rank scores are defined in Section 2. The error distributional assumptions are stated in Section 3. Section 4 contains the main results of the paper. They are followed by several remarks on the implications of these results. Section 5 contains some examples of the L.R.D. models with linear and nonlinear trends where the results of Section 4 are applied. Finally, the Appendix contains the proofs of the linearization results of randomly weighted empirical processes and some auxiliary results on the properties of nonlinear regression quantiles that are analogous to those of order statistics.

2. NONLINEAR REGRESSION QUANTILES, *L*-ESTIMATORS AND RANK SCORES

Consider the nonlinear regression model where one observes an array of random vectors $\{(Y_{n,i}, X_{n,i})'; 1 \le i \le n\}$ satisfying

$$Y_{ni} = h(\beta, X_{ni}) + \epsilon_i, \qquad 1 \le i \le n. \tag{2.1}$$

Here $\beta' := (\beta_0, \tilde{\beta}') \in \Omega$ is the unknown parameter where Ω is an open subset of $\mathbb{R}^{1+d}$ of the form $\Omega = \mathbb{R}^1 \times \tilde{\Omega}$; $h : \Omega \times \mathbb{R}^p \to \mathbb{R}^1$ is a known function of the form $h\{(u_0, u_1, \dots, u_d), x\} = u_0 + \tilde{h}\{(u_1, \dots, u_d), x\}$ such that for each $x \in \mathbb{R}^p$, the function $h(\cdot, x)$ is differentiable with vector of partial derivatives at u represented by $\dot{h}(u, x)$; $\{X_{ni}; 1 \le i \le n\}$ is an array of p-dimensional random vectors representing regressor variables; and $\{\epsilon_i; 1 \le i \le n\}$ are the unobservable error random variables.

Following Jurečková and Procházka (1994), an αth nonlinear regression quantile (NLRQ) $\hat{\beta}_n(\alpha)$ in the model (2.1) is defined as

$$\hat{\beta}_n(\alpha) := \operatorname{argmin}\left\{\sum_{i=1}^{n} \rho_\alpha(Y_{ni} - h(u, X_{ni}))\right\}, \tag{2.2}$$

where $\rho_\alpha(y) := -(1-\alpha)yI(y \le 0) + \alpha y I(y > 0)$. When $\alpha = \frac{1}{2}$, $\hat{\beta}_n(\frac{1}{2})$ is called the LAD estimator and is denoted by $\hat{\beta}_{LAD}$. As is common to this kind of implicit definition, in general there may not be any unique minimizer or any minimizer at all in (2.2), and thus the definition suffers from some kind of ambiguity. However, if the function h is sufficiently smooth, Koul (1996b) has shown that the LSE is consistent for β, and thus one can choose a minimizer in (2.2), which is closest to the LSE by, say, $1/n$, as estimator. Koenker and Park (1996) provided some sufficient conditions on the function h for the existence of minimizers. They suggested an "interior point algorithm" that can be used for computation.

Note that if $Y_{ni} - h(u, X_{ni}) \neq 0$ for all $1 \le i \le n$, then u can not minimize $\sum_{i=1}^n \rho_\alpha(Y_{ni} - h(u, X_{ni}))$. Therefore, a minimizer $\hat{\beta}_n(\alpha)$ must satisfy $Y_{ni} - h(\hat{\beta}_n(\alpha), X_{ni}) = 0$ for some i. Define

$$B_n(\alpha) := \{i; Y_{ni} = h(\hat{\beta}_n(\alpha), X_{ni})\}.$$

Let N, P, and Z denote the numbers of elements of the sets $\{i;\ Y_{ni} < h(\hat{\beta}_n(\alpha), X_{ni})\}$, $\{i; Y_{ni} > h(\hat{\beta}_n(\alpha), X_{ni})\}$, and $B_n(\alpha)$, respectively. Then, Assumption (h.0) of Section 4 and the continuity of the underlying random variables (Assumption (M.3) of Section 3) imply that $Z = 1 + d$ with probability one and

$$N \le n\alpha, \qquad P \le n(1-\alpha). \tag{2.3}$$

For a proof, see Lemma A.4(i). Recall that if $Y_{([n\alpha]+1)}$ denotes the $([n\alpha]+1)$th-order statistic of $\{Y_{n1}, \ldots, Y_{nn}\}$ and N_0, P_0, and Z_0 denote the numbers of elements of the sets $\{i;\ Y_{ni} < Y_{([n\alpha]+1)}\}$, $\{i; Y_{ni} > Y_{([n\alpha]+1)}\}$, and $\{i; Y_{ni} = Y_{([n\alpha]+1)}\}$, respectively, then $N_0 \le n\alpha$, $P_0 \le n(1-\alpha)$, and $Z_0 = 1$. In this sense, $\hat{\beta}_n(\alpha)$ is a proper analogue of the $([n\alpha]+1)$th order statistic of $\{Y_{n1}, \ldots, Y_{nn}\}$.

In analogy to Koenker and Portnoy (1987), we define an L-estimator of β in the model (2.1) as

$$\hat{\beta}_\Lambda := \int_0^1 \hat{\beta}_n(\alpha)\Lambda(d\alpha), \tag{2.4}$$

where

(L.1) Λ is a finite signed measure on a compact subinterval of (0,1), with $\Lambda\{(0,1)\} = 1$.

Note that $\hat{\beta}(\alpha_1, \alpha_2) = \hat{\beta}_\Lambda$, corresponding to the probability measure $\Lambda(d\alpha) = (\alpha_2 - \alpha_1)^{-1} I(\alpha_1 \le \alpha \le \alpha_2)d\alpha$; $0 < \alpha_1 < \frac{1}{2} < \alpha_2 < 1$ is an analogue of the trimmed mean in the location model with $100\alpha_1\%$ lower trimming and $100(1-\alpha_2)\%$ upper trimming.

In the context of the linear regression model, Gutenbrunner and Jurečková (1992, display (3.11)) introduced regression rank score process (RRSP) as an extension of Hájek and Šidák's definition of rank process (1967, Sect. V.3.5) of the location model. RRSP and linear rank statistics based on RRSP are useful in various hypothesis testing problems. Following Mukherjee (1999a), here we give a definition of nonlinear rank score process (NLRSP) that is not an extension of RRSP but nevertheless has similar asymptotic properties. Moreover, for the one-sample location model, the definition of NLRSP reduces to Hájek and Šidák's definition of rank process. Toward that end, fix any $\hat{\beta}_n(\alpha)$ and the corresponding index set $B_n(\alpha)$. For $i \notin B_n(\alpha)$, define

$$\begin{aligned} \hat{a}_{ni}(\alpha) &= 1 \quad \text{if } Y_{ni} - h(\hat{\beta}_n(\alpha), X_{ni}) > 0 \\ &= 0 \quad \text{if } Y_{ni} - h(\hat{\beta}_n(\alpha), X_{ni}) < 0. \end{aligned}$$

Also for $i \in B_n(\alpha), \hat{a}_{ni}(\alpha)$ is defined as a solution of

$$\sum_{i=1}^{n} h(\hat{\beta}_n(\alpha), X_{ni})\hat{a}_{ni}(\alpha) = (1-\alpha)\sum_{i=1}^{n} h(\hat{\beta}_n(\alpha), X_{ni}), \tag{2.5}$$

with $\hat{a}_{ni}(\alpha) \in [0,1]$.

Let $\{w_{ni}; 1 \le i \le n\}$ be an array of $\mathbb{R}^r$ valued random vectors. Then the sequence of NLRSP $\hat{W}_n^w$ on (0,1) is defined by

$$\hat{W}_n^w(\alpha) := \sum_{i=1}^{n} w_{ni}\,\hat{a}_{ni}(\alpha), \qquad 0 < \alpha < 1.$$

Also, let $W_n^{\dot{h}}$ stand for W_n^w with $w_{ni} = \dot{h}_{ni}(\beta)$ and $r = 1 + d$.

Next, for a function $b:(0,1) \to \mathbb{R}^1$ that is of bounded variation and constant outside a compact subinterval of (0,1), the ith score is defined as

$$\hat{b}_{ni} := \int_0^1 \hat{a}_{ni}(\alpha)b(d\alpha), \qquad 1 \le i \le n.$$

Gutenbrunner and Jurečková (1992) showed that in the location model, $\{\hat{b}_{ni}\}$'s reduce to familiar rank scores as discussed in Hájek and Šidák (1967). Based on the NLRSP, a nonlinear rank statistic (NLRS) $\hat{U}_n^w$ is defined as

$$\hat{U}_n^w := \sum_{i=1}^{n} w_{ni}\,\hat{b}_{ni} = \int_0^1 \hat{W}_n^w(\alpha)b(d\alpha).$$

3. LONG RANGE DEPENDENCE, ERROR DISTRIBUTIONAL ASSUMPTIONS, AND HERMITE RANK

Let $\{\eta_i; i \ge 1\}$ be a sequence of stationary random variables with a standard normal marginal distribution and with correlation at "lag" k,

$$\rho(k) := \text{correlation}(\eta_1, \eta_{1+k}) = L(k)/k^{\theta}, \qquad k \ge 1. \tag{3.1}$$

Here $0 < \theta < 1$ and L is a positive (eventually) and slowly varying function at infinity, i.e., $\lim_{t\to\infty} L(tx)/L(t) = 1$, for all $x > 0$. The long range dependence among the error random variables $\{\epsilon_i; 1 \le i \le n\}$ is modeled by assuming the following:

(M.1) $\epsilon_i = G(\eta_i)$, $1 \le i \le n$, where $G : \mathbb{R}^1 \to \mathbb{R}^1$ is an unknown function.

L.R.D. data are encountered in the fields of hydrology, economics, time series analysis, and other sciences. Existing statistical methodologies may exhibit very different characteristics when applied to L.R.D. data. These facts are reflected in the review paper by Beran (1992) and the references therein. These are more popularly known as long memory data among the economists.

Let $\mathcal{F} := \{x; 0 < F(x) < 1\}$ be an open interval denoting the support of the error random variables $\{G(\eta_i)\}$ with distribution function F. Let m be the Hermite rank of the class of functions $\{I(G(\eta_1) \le x), x \in \mathcal{F}\}$ (for the definition, see Dehling and Taqqu, 1989). For $x \in \mathcal{F}$, define $J_m(x) := E[I(G(\eta_1) \le x)H_m(\eta_1)]$ (which is nonzero for some $x \in \mathcal{F}$) and $J_m^+(x) := E[I(G(\eta_1) \le x)|H_m(\eta_1)|]$. We further assume the following:

(M.2) For each $n \ge 1$, the sigma fields generated by $\{X_{ni}; 1 \le i \le n\}$ and $\{\epsilon_i; 1 \le i \le n\}$ are independent.
(M.3) The error distribution function F has a continuous density f that is positive on $\mathcal{F}$.
(M.4) The functions J_m and J_m^+ are continuously differentiable on $\mathcal{F}$.

For L and θ as in (3.1), assume that $m < \theta^{-1}$ and define

$$\tau_n := \{2m!(1 - m\theta)^{-1}(2 - m\theta)^{-1} n^{1-m\theta} L^m(n)\}^{1/2}. \tag{3.2}$$

4. ASYMPTOTIC DISTRIBUTIONS OF NONLINEAR REGRESSION QUANTILES AND RANK SCORE PROCESSES

In the following, let $h(u, X_{ni})$ and $\dot{h}(u, X_{ni})$ be denoted by $h_{ni}(u)$ and $\dot{h}_{ni}(u)$, respectively. Define $T(u,\alpha)$, the almost everywhere derivative of $\sum_{i=1}^{n} \rho_\alpha(Y_{ni} - h_{ni}(u))$, as

$$\begin{aligned} T(u,\alpha) &:= \sum_{i=1}^{n} \dot{h}_{ni}(u)\{I(Y_{ni} - h_{ni}(u) \le 0) - \alpha\} \\ &= \sum_{i=1}^{n} \dot{h}_{ni}(u)\{I(\epsilon_i \le h_{ni}(u) - h_{ni}(\beta)) - \alpha\}, \qquad u \in \Omega, \qquad \alpha \in (0,1). \end{aligned}$$

To find out the asymptotic distribution of $\hat{\beta}_n(\alpha)$, define the centering constant $\beta(\alpha) := \beta + F^{-1}(\alpha)e_1$, where $F^{-1}(\alpha) := \inf\{x \in \mathcal{F}; \alpha \le F(x)\}$ and $e_j := [0,\ldots,1,\ldots,0]'$, a vector with one at the jth position and zeros at all other positions, $1 \le j \le 1 + d$. To proceed further, note that our aim is to approximate $A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\}$ by $B_n^{-1} T(\beta(\alpha),\alpha)$, for appropriate matrices A_n and B_n

(see (h.0) and (4.2), which follow). Therefore, we must choose B_n in such a way that $B_n^{-1}T(\beta(\alpha),\alpha) = O_p(1)$. In the i.i.d. errors case, $T(\beta(\alpha),\alpha)$ is a sum of i.i.d. centered random variables, and hence the obvious choice for B_n is $n^{1/2}$. By contrast, in the L.R.D. case, the order of B_n depends on both $\{\dot{h}(\beta, X_{ni})\}$ and $\{I(\epsilon_i \le F^{-1}(\alpha)) - \alpha\}$. We shall discuss different choices of B_n in Remark 4.3 under different conditions on the underlying design and error random variables. Accordingly, we make the following assumptions on h.

(h.0) For any set of $1 + d$ equations, $Y_{ni_j} = h(u, X_{ni_j})$, $1 \le j \le 1 + d$, there is a unique solution in u. Let $\dot{\mathcal{H}}_n$ denote an $n \times (1 + d)$ matrix with ith row $\dot{h}_{ni}(\beta)$. We assume that the matrix $\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n$ is positive definite with probability one. Also, let D_n be a matrix such that $B_n^{-1}T(\beta(\alpha),\alpha) = O_p(1)$, where $B_n := \tau_n D_n$. Defining $A_n := B_n^{-1}\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n$, we further assume that $\|A_n^{-1}\| = o_p(1)$.

(h.1) $n \max\{E\|D_n^{-1}\dot{h}_{ni}(\beta)\|^2; 1 \le i \le n\} = O(1)$

For $b > 0$, let $\mathcal{N}_b := \{t \in \mathbb{R}^{1+d}; \|t\| \le b\}$. The following assumptions hold for each $t \in \mathbb{R}^{1+d}$ and $b > 0$.

(h.2t) $n \max\{E\|D_n^{-1}(\dot{h}_{ni}(\beta + A_n^{-1}t) - \dot{h}_{ni}(\beta))\|^2; 1 \le i \le n\} = o(1)$.

Moreover, for all $1 \le j \le 1 + d$,

$$0 < k_j(t) < e_j' E\left[D_n^{-1} \sum_{i=1}^n \dot{h}_{ni}(\beta + A_n^{-1}t)\dot{h}_{ni}'(\beta + A_n^{-1}t)D_n^{-1}\right]e_j,$$

for some constants $k_j(t)$.

(h.1b) For every $\alpha > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left[\sup\left\{\tau_n^{-1}\sum_{i=1}^n \|D_n^{-1}(\dot{h}_{ni}(\beta + A_n^{-1}s) - \dot{h}_{ni}(\beta + A_n^{-1}t))\|;\right.\right.$$
$$\left.\left. s, t \in \mathcal{N}_b, \|s - t\| \le \delta\right\} < \alpha\right] \ge 1 - \alpha.$$

(h.2b) For every $\alpha > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left[\sup\left\{\tau_n^{-1}\sum_{i=1}^n \|D_n^{-1}\dot{h}_{ni}(\beta + A_n^{-1}s)(h_{ni}(\beta + A_n^{-1}s) - h_{ni}(\beta + A_n^{-1}t))\|;\right.\right.$$
$$\left.\left. s, t \in \mathcal{N}_b, \|s - t\| \le \delta\right\} < \alpha\right] \ge 1 - \alpha.$$

Remark 4.1. Condition (h.0) was assumed by Jurečková and Procházka (1994). When specialized to linear models, it reduces to the condition that the rank of the design matrix is $1 + d$. Variants of (h.0) and (h.1) are standard assumptions for the linear regression model with random design points as in Pollard (1991, Theorem 2). When $X_{ni} = X_i$, say, (free from n), and X_i's are stationary random variables, conditions (h.1) and (h.2t) reduce to the following two mild second moment conditions on the stationary distribution:

$$nE\|D_n^{-1}\dot{h}(\beta, X_1)\|^2 = O(1),$$

and

$$nE\|D_n^{-1}\{\dot{h}(\beta + A_n^{-1}t, X_1) - \dot{h}(\beta, X_1)\}\|^2 = o(1).$$

Also, (h.0), (h.1), and (h.2t) imply that for each t,

$$\max\{|h_{ni}(\beta + A_n^{-1}t) - h_{ni}(\beta)|; 1 \leq i \leq n\} = o_p(1)$$

and

$$\tau_n^{-1}\sum_{i=1}^{n} E|e_j' D_n^{-1}\dot{h}_{ni}(\beta + A_n^{-1}t)\{h_{ni}(\beta + A_n^{-1}t) - h_{ni}(\beta)\}| = O(1).$$

When $\sup\{\|\dot{h}_{ni}(u)\|; 1 \leq i \leq n, \|u - \beta\| \leq \tilde{\delta}\} = O_p(1)$ for some $\tilde{\delta} > 0$, simpler sufficient conditions for (h.1b) and (h.2b) are the following.

(h.1*b) For every $\alpha > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\Bigg[\tau_n^{-1}\sum_{i=1}^{n}\sup\{\|D_n^{-1}(\dot{h}_{ni}(\beta + A_n^{-1}s) - \dot{h}_{ni}(\beta + A_n^{-1}t))\|; s, t \in \mathcal{N}_b, \|s - t\| \leq \delta\} < \alpha\Bigg] \geq 1 - \alpha.$$

(h.2*b) For every $\alpha > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\Bigg[\tau_n^{-1}\|D_n^{-1}\|\sum_{i=1}^{n}\sup\{|h_{ni}(\beta + A_n^{-1}s) - h_{ni}(\beta + A_n^{-1}t)|; s, t \in \mathcal{N}_b, \|s - t\| \leq \delta\} < \alpha\Bigg] \geq 1 - \alpha.$$

THEOREM 4.1. *Assume that in the model* (2.1), (M.1)–(M.3), (h.0), (h.1), (h.2t), (h.1b), *and* (h.2b) *hold. Let* $q(\alpha) = f(F^{-1}(\alpha))$.

(i) *Then for each* $\alpha \in (0,1)$ *and* $b > 0$,

$$\sup\{\|B_n^{-1}[T(\beta(\alpha) + A_n^{-1}t, \alpha) - T(\beta(\alpha), \alpha)] - tq(\alpha)\|; t \in \mathcal{N}_b\} = o_p(1).$$

(ii) *In addition, suppose there exists a sequence of minimizers* $\hat{\beta}_n(\alpha)$ *in* (2.2) *such that*

$$A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\} = O_p(1). \tag{4.1}$$

Then,

$$A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\} = [-q(\alpha)B_n]^{-1}T(\beta(\alpha),\alpha) + o_p(1) \quad \mathbf{(4.2)}$$

$$= [-q(\alpha)B_n m!]^{-1}J_m(F^{-1}(\alpha)) \times \sum_{i=1}^{n} \dot{h}_{ni}(\beta)H_m(\eta_i) + o_p(1). \quad \mathbf{(4.3)}$$

(iii) *In particular, when* $\alpha = \frac{1}{2}$ *and* F *has the median at* 0,

$$A_n(\hat{\beta}_{LAD} - \beta) = [-f(0)B_n m!]^{-1} \sum_{i=1}^{n} \dot{h}_{ni}(\beta)\left\{I(\epsilon_i \le 0) - \frac{1}{2}\right\} + o_p(1)$$

$$= [-f(0)B_n m!]^{-1}J_m(0) \sum_{i=1}^{n} \dot{h}_{ni}(\beta)H_m(\eta_i) + o_p(1).$$

Proof. It is easy to see that $h_{ni}(\beta(\alpha) + A_n^{-1}t) = F^{-1}(\alpha) + h_{ni}(\beta + A_n^{-1}t)$ and $\dot{h}_{ni}(\beta(\alpha) + A_n^{-1}t) = \dot{h}_{ni}(\beta + A_n^{-1}t)$. Also, the continuity of F implies that $F(F^{-1}(\alpha)) = \alpha$. Using these

$$\begin{aligned} &B_n^{-1}[T(\beta(\alpha) + A_n^{-1}t,\alpha) - T(\beta(\alpha),\alpha)] - tq(\alpha) \\ &\quad = V_h(F^{-1}(\alpha),t) - V_h(F^{-1}(\alpha),0) - tf(F^{-1}(\alpha)) \\ &\qquad - B_n^{-1}\sum_{i=1}^{n}\{\dot{h}_{ni}(\beta + A_n^{-1}t) - \dot{h}_{ni}(\beta)\}\alpha, \end{aligned}$$

and hence (i) follows from Theorem A.1(i), which appears in the Appendix.

Now note that by (4.1) and (i),

$$\begin{aligned} &B_n^{-1}[T(\beta(\alpha) + A_n^{-1}A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\},\alpha) - T(\beta(\alpha),\alpha)] \\ &\quad - A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\}q(\alpha) = o_p(1). \end{aligned}$$

Also, by Lemma A.4(ii), found in the Appendix, $B_n^{-1}T(\hat{\beta}_n(\alpha),\alpha) = o_p(1)$. Hence

$$A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\}q(\alpha) = -B_n^{-1}T(\beta(\alpha),\alpha) + o_p(1), \quad \mathbf{(4.4)}$$

and (4.2) follows. Also, (4.3) follows from (4.2) by applying Lemma A.2(ii) (in the Appendix) with γ_{ni} equal to the kth coordinate of $D_n^{-1}\dot{h}_{ni}(\beta)$, $1 \le k \le 1 + d$. ■

Remark 4.2. Following Koul (1996a, Corollary 1.1), a sufficient condition for the existence of $\hat{\beta}_n(\alpha)$ satisfying (4.1) is as follows.

For every $\epsilon, M > 0$, there exist n_0 and $b > 0$ such that for all $n \ge n_0$,

$$P[\inf\{\|B_n^{-1}T(\beta(\alpha) + A_n^{-1}t,\alpha)\|;\ \|t\| \ge b\} \ge M] \ge 1 - \epsilon.$$

This condition, in turn, is implied by the following condition.

For every $e \in \mathbb{R}^{1+d}$ with $\|e\| = 1$ and for all large n, $e' B_n^{-1} T(\beta(\alpha) + A_n^{-1} re, \alpha)$ is monotone in $r \in (0,\infty)$, almost surely.

The preceding condition is satisfied, for example, whenever h is linear in parameters.

Remark 4.3. Now we discuss the choice of B_n. First consider the case when $\{X_{ni}\}$'s are nonrandom; $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \dot{h}_{ni}(\beta)\dot{h}'_{ni}(\beta)$ exists and is positive definite. Then one can choose $B_n = \tau_n [\sum_{i=1}^n \dot{h}_{ni}(\beta)\dot{h}'_{ni}(\beta)]^{1/2}$. Here the rate of convergence of $\hat{\beta}_n(\alpha)$ to $\beta(\alpha)$ is $\tau_n^{-1} n^{1/2}$, which is slower than the typical $n^{1/2}$ rate of convergence.

Now consider the case when $X_{ni} = X_i$ and $\{X_i; i \geq 1\}$'s are i.i.d. Denoting $\dot{h}(\beta, X_i)$ by $[1, \xi_i']'$, we also assume that $E\xi_1 = 0$ and $E\xi_1\xi_1'$ is positive definite. Because $T(\beta(\alpha),\alpha) = [\sum_{i=1}^n \{I(\epsilon_i \leq F^{-1}(\alpha)) - \alpha\}, \sum_{i=1}^n \xi_i'\{I(\epsilon_i \leq F^{-1}(\alpha)) - \alpha\}]'$, we can calculate its L^2 norm to see that

$$\left[(n^{1/2}\tau_n)^{-1} \sum_{i=1}^n \{I(\epsilon_i \leq F^{-1}(\alpha)) - \alpha\}, n^{-1/2} \sum_{i=1}^n \xi_i'\{I(\epsilon_i \leq F^{-1}(\alpha)) - \alpha\}\right]' = O_p(1).$$

Hence, we may choose

$$B_n = \begin{pmatrix} n^{1/2}\tau_n & 0 \\ 0 & n^{1/2} I_{d\times d} \end{pmatrix},$$

and in this case, the intercept parameter has $\tau_n^{-1} n^{1/2}$ rate of convergence, whereas the corresponding rate for each of the slope parameters is $n^{1/2}$. Therefore, i.i.d. regressors overcome the effect of the dependence among errors by retaining the traditional rate of convergence.

Finally, consider a case when $X_{ni} = X_i$ and $\dot{h}(\beta, X_{ni})$ is of the form $[1, G^*(\eta_{i1}^*), \ldots, G^*(\eta_{id}^*)]'$, where $G^*: \mathbb{R}^1 \to \mathbb{R}^1$ is a measurable function (possibly dependent on β) and $\{\eta_i^* = [\eta_{i1}^*, \ldots, \eta_{id}^*]'; i \geq 1\}$ is a sequence of L.R.D. normal vectors in the sense of (3.1) in Arcones (1994). Also $\{\eta_i^*; i \geq 1\}$ and $\{\eta_i; i \geq 1\}$ are independent. If the Hermite rank (Arcones, 1994, Definition 2.2) of the function $Z_k(\eta^*, \eta) := G^*(\eta_k^*)\{I(G(\eta) \leq F^{-1}(\alpha)) - \alpha\}$ is m^* for all $\alpha \in (0,1)$ with $m^* \geq m$, then by Theorem 6 of Arcones (1994) we can choose

$$B_n = \begin{pmatrix} n^{1/2}\tau_n & 0 \\ 0 & n^{1/2}\tau_n^* I_{d\times d} \end{pmatrix},$$

where τ_n^* is τ_n in (3.2) with $m = m^*$.

Remark 4.4. Note that under (h.1), $E\|B_n^{-1} \sum_{i=1}^n \dot{h}_{ni}(\beta) H_m(\eta_i)\|^2 = O(1)$, and hence the sequence $[-q(\alpha) B_n m!]^{-1} J_m(F^{-1}(\alpha)) \sum_{i=1}^n \dot{h}_{ni}(\beta) H_m(\eta_i)$ of random vectors in the right hand side of (4.3) is tight. But unlike the i.i.d. errors case, this tight sequence of random vectors need not converge in distribution, and even if it converges, no closed form expression for the limiting distribution is available. However, if either G is strictly monotone and continuous function, or G is an odd function with the property that $\{x \in \mathbb{R}^1; G(x) \leq 0\}$ equals either $(-\infty,0]$ or $[0,\infty)$, then by Koul and Mukherjee (1993), $m = 1$ and $H_m(\eta_i) = \eta_i$. In this case, if the conditional dispersion matrix of $S_n := B_n^{-1} \sum_{i=1}^n \dot{h}_{ni}(\beta)\eta_i$ given $\{X_{ni}; 1 \leq i \leq n\}$ converges in probability to a positive definite matrix Γ, say, then using the convergence theorem of characteristic functions, it can be shown that S_n converges in distribution to the normal random vector $N_{1+d}[0,\Gamma]$.

In the sequel, for a vector-valued stochastic process $\{M_n(\alpha); \alpha \in (0,1)\}$, $\|M_n\|_a$ denotes $\sup\{\|M_n(\alpha)\|; a \leq \alpha \leq 1-a\}$, and we say that $M_n(\alpha) = O_p^*(1)(o_p^*(1))$ if for every $a \in (0,\frac{1}{2}]$, $\|M_n\|_a = O_p(1)(o_p(1))$. The following theorem gives the uniform asymptotic representation of the nonlinear regression quantile process on compact subintervals of (0,1). The uniform representation is then used to obtain the asymptotic representations of L-estimators.

THEOREM 4.2. *Assume that in the model* (2.1), (M.1)–(M.4), (h.0), (h.1), (h.2t), (h.1b), *and* (h.2b) *hold.*

(i) *Then for every* $a \in (0,\frac{1}{2}]$,

$$\sup\{\|B_n^{-1}[T(\beta(\alpha) + A_n^{-1}t,\alpha) - T(\beta(\alpha),\alpha)] - tq(\alpha)\|;\\ (\alpha,t) \in [a,1-a] \times \mathcal{N}_b\} = o_p(1).$$

(ii) *In addition, suppose there exists a sequence of minimizers* $\hat{\beta}_n(\alpha)$ *of* (2.2) *such that*

$$A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\} = O_p^*(1). \tag{4.5}$$

Then

$$A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\} = [-q(\alpha)B_n]^{-1}T(\beta(\alpha),\alpha) + o_p^*(1)$$
$$= [-q(\alpha)B_n m!]^{-1}J_m(F^{-1}(\alpha)) \sum_{i=1}^n \dot{h}_{ni}(\beta)H_m(\eta_i) + o_p^*(1).$$

(iii) *Consequently, if* Λ *satisfies* (L.1), *then*

$$A_n\left\{\hat{\beta}_\Lambda - \beta - e_1\int_0^1 F^{-1}(\alpha)\Lambda(d\alpha)\right\}$$
$$= B_n^{-1}\sum_{i=1}^n \dot{h}_{ni}(\beta) \int_0^1 -q^{-1}(\alpha)\{I(\epsilon_i \leq F^{-1}(\alpha)) - \alpha\}\Lambda(d\alpha) + o_p(1) \tag{4.6}$$
$$= B_n^{-1}\sum_{i=1}^n \dot{h}_{ni}(\beta)H_m(\eta_i)\int_0^1 \{-q(\alpha)m!\}^{-1}J_m(F^{-1}(\alpha))\Lambda(d\alpha) + o_p(1). \tag{4.7}$$

(iv) *In particular*, *with* $\Lambda(d\alpha) := (\alpha_2 - \alpha_1)^{-1} I(\alpha_1 \le \alpha \le \alpha_2)\, d\alpha$,

$$A_n \left\{ \hat{\beta}(\alpha_1, \alpha_2) - \beta - e_1 (\alpha_2 - \alpha_1)^{-1} \int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} x F(dx) \right\}$$

$$= B_n^{-1} \sum_{i=1}^{n} \dot{h}_{ni}(\beta) H_m(\eta_i) \{-m!(\alpha_2 - \alpha_1)\}^{-1} \int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} J_m(x)\, dx + o_p(1).$$

Proof. Relation (i) follows from Theorem A.1(ii) in the same fashion as Theorem 4.1(i) follows from Theorem A.1(i). The proofs of other assertions follow similarly. ■

Remark 4.5 Comparison with Other Estimators. Using the linearity results of the Appendix, we can obtain the following asymptotic representations of some robust estimators of β in the model (2.1). With $\bar{\dot{h}}_n(\beta) := n^{-1} \sum_{i=1}^{n} \dot{h}_{ni}(\beta)$ and $D_{nc} := [\sum_{i=1}^{n} \{\dot{h}_{ni}(\beta) - \bar{\dot{h}}_n(\beta)\}\{\dot{h}_{ni}(\beta) - \bar{\dot{h}}_n(\beta)\}']^{1/2}$,

$$A_n(\hat{\beta}_M - \beta) = B_n^{-1} \sum_{i=1}^{n} \dot{h}_{ni}(\beta) \left(- \int f\, d\psi \right)^{-1} \psi(\epsilon_i) + o_p(1), \qquad \textbf{(4.8)}$$

and

$$\tau_n^{-1} D_{nc}(\hat{\beta}_R - \beta)$$

$$= \tau_n^{-1} D_{nc}^{-1} \sum_{i=1}^{n} \{\dot{h}_{ni}(\beta) - \bar{\dot{h}}_n(\beta)\} \left(\int f\, d\phi(F) \right)^{-1} \phi(F(\epsilon_i)) + o_p(1), \qquad \textbf{(4.9)}$$

where $\hat{\beta}_M$ is an M-estimator based on a nondecreasing function ψ with $E(\psi(\epsilon_1)) = 0$ and $\hat{\beta}_R$ is a rank estimator based on a nondecreasing function ϕ. On the other hand, recall that

$$A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\} = B_n^{-1} \sum_{i=1}^{n} \dot{h}_{ni}(\beta) \times \{-q^{-1}(\alpha)\}\{I(\epsilon_i \le F^{-1}(\alpha)) - \alpha\}$$
$$+ o_p(1)$$

and

$$A_n \left\{ \hat{\beta}_\Lambda - \beta - e_1 \int_0^1 F^{-1}(\alpha) \Lambda(d\alpha) \right\}$$

$$= B_n^{-1} \sum_{i=1}^{n} \dot{h}_{ni}(\beta) \int_0^1 - q^{-1}(\alpha) \{I(\epsilon_i \le F^{-1}(\alpha)) - \alpha\} \Lambda(d\alpha) + o_p(1).$$

These representations enable us to study and compare the asymptotic relationships between the classes of M-, R-, and L-estimators. For example, the LAD is asymptotically more efficient than the LSE ($\hat{\beta}_M$ with $\psi(x) \equiv x$), whenever the same is true in the model (2.1) with i.i.d. errors. In particular, $\hat{\beta}_{LAD}$ and $\hat{\beta}(\alpha_1,\alpha_2)$ are asymptotically more robust (in the sense of asymptotic efficiency) than the LSE for heavy tailed F.

Also, corresponding to an M-estimator with score function ψ that is constant outside a compact interval of the real line, there is an asymptotically equivalent L-estimator given by

$$\Lambda(du) = \left\{\int_0^1 \psi'(F^{-1}(\alpha))\,d\alpha\right\}^{-1} \psi'(F^{-1}(u))\,du, \qquad u \in (0,1).$$

Similarly, corresponding to an R-estimator with a differentiable score function ϕ that is constant outside a closed interval of (0,1) and with $\bar{\dot{h}}_n(\beta) = 0$, an asymptotically equivalent L-estimator is given by

$$\Lambda(du) = \left\{\int \phi'(F(x)) f^2(x)\,dx\right\}^{-1} \phi'(u) f(F^{-1}(u))\,du, \qquad u \in (0,1).$$

Recall that (4.3) and (4.7) are typical Hermite expansions of (4.2) and (4.6), respectively. Under some regularity conditions, one can get Hermite expansions of the random vectors in (4.8) and (4.9) in a similar fashion. From these expansions, it is easy to see that when $G(\eta_i) = \eta_i$, i.e., when the errors are L.R.D. normal, the asymptotic distributions of $\hat{\beta}_n(\alpha)$, $\hat{\beta}_\Lambda$, $\hat{\beta}_M$, and $\hat{\beta}_R$ are same irrespective of α, Λ, ψ, and ϕ. *This is in complete contrast with the i.i.d. errors case.*

To derive asymptotic representations of $\hat{W}_n^w(\cdot)$ and $\hat{U}_n^w$, we assume that the coefficients $\{w_{ni};1 \le i \le n\}$ satisfy the following conditions.

(W.0) For each $n \ge 1$, the sigma fields generated by $\{w_{ni};\, 1 \le i \le n\}$ and $\{\epsilon_i;\, 1 \le i \le n\}$ are independent. Also, letting $\mathcal{W}_n$ = an $n \times r$ matrix with ith row w_{ni}', we assume that the matrix $\mathcal{W}_n'\mathcal{W}_n$ is positive definite. Moreover, define $A_w := \tau_n^{-1} D_w$ and $B_w := \tau_n D_w$, where $D_w := [\mathcal{W}_n'\mathcal{W}_n]^{1/2}$.

(W.1) $n \max\{E\|D_w^{-1} w_{ni}\|^2;1 \le i \le n\} = O(1)$.

Define an approximating sequence of processes W_n^w by

$$W_n^w(\alpha) := \sum_{i=1}^n w_{ni} I(\epsilon_i > F^{-1}(\alpha)), \qquad 0 < \alpha < 1.$$

Let $v_{ni} := D_w^{-1} w_{ni} - \mathcal{M}_n(\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n)^{-1/2}\dot{h}_{ni}(\beta)$, where $\mathcal{M}_n := D_w^{-1}\mathcal{W}_n'\dot{\mathcal{H}}_n \times (\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n)^{-1/2}$. The following result is useful for testing problems concerning β based on rank statistic.

THEOREM 4.3. *In addition to the assumptions of Theorem* 4.2(ii), *assume that* (W.0) *and* (W.1) *hold. Then* $\hat{W}_n^w$ *and* $\hat{U}_n^w$ *admit the following asymptotic representations*:

(i) $B_w^{-1}\left\{\hat{W}_n^w(\alpha) - \sum_{i=1}^n w_{ni}(1-\alpha)\right\}$

$$= \tau_n^{-1}\sum_{i=1}^n v_{ni}\{I(\epsilon_i > F^{-1}(\alpha)) - (1-\alpha)\} + o_p^*(1)$$

$$= (-\tau_n m!)^{-1} J_m(F^{-1}(\alpha)) \sum_{i=1}^n v_{ni} H_m(\eta_i) + o_p^*(1).$$

Consequently,

(ii) $B_w^{-1}\left\{\hat{U}_n^w - \sum_{i=1}^n w_{ni}\int_0^1 (1-\alpha)b(d\alpha)\right\}$

$$= (-\tau_n m!)^{-1}\sum_{i=1}^n v_{ni} H_m(\eta_i)\int_0^1 J_m(F^{-1}(\alpha))b(d\alpha) + o_p(1).$$

Proof. Write $Z_n(\alpha) := A_n(\hat{\beta}_n(\alpha) - \beta(\alpha))$. Note that for $i \notin B_n(\alpha)$,

$$\hat{a}_{ni}(\alpha) = I\{\epsilon_i > h_{ni}(\beta(\alpha) + A_n^{-1}Z_n(\alpha)) - h_{ni}(\beta)\}$$
$$= I\{\epsilon_i > F^{-1}(\alpha) + h_{ni}(\beta + A_n^{-1}Z_n(\alpha)) - h_{ni}(\beta)\}.$$

Also, for $i \in B_n(\alpha)$, $I\{\epsilon_i > h_{ni}(\hat{\beta}_n(\alpha)) - h_{ni}(\beta)\} = 0$. Hence,

$$B_w^{-1}\hat{W}_n^w(\alpha) = B_w^{-1}\sum_{i=1}^n w_{ni} I\{\epsilon_i > F^{-1}(\alpha) + h_{ni}(\beta + A_n^{-1}Z_n(\alpha)) - h_{ni}(\beta)\}$$
$$+ B_w^{-1}\sum_{i\in B_n(\alpha)} w_{ni}\hat{a}_{ni}(\alpha). \quad \textbf{(4.10)}$$

By (W.1), the last term in (4.10) is $o_p(1)$. Next, applying Lemma A.2(v) r number of times on the first term with γ_{ni} equals the kth $(1 \le k \le r)$ entry of $D_w^{-1}w_{ni}$, $\xi_{ni} = h_{ni}(\beta + A_n^{-1}t) - h_{ni}(\beta)$, and $y = F^{-1}(\alpha)$ and then letting $t = Z_n(\alpha)$, we get

$$B_w^{-1}\hat{W}_n^w(\alpha) = B_w^{-1}W_n^w(\alpha)$$
$$- B_w^{-1}\sum_{i=1}^n w_{ni}[F\{F^{-1}(\alpha) + h_{ni}(\beta + A_n^{-1}Z_n(\alpha)) - h_{ni}(\beta)\}$$
$$- F(F^{-1}(\alpha))] + o_p^*(1)$$
$$= B_w^{-1}W_n^w(\alpha) - \mathcal{M}_n(\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n)^{-1/2}D_n Z_n(\alpha)q(\alpha) + o_p^*(1).$$

But, from Theorem 4.2(ii),

$$Z_n(\alpha)q(\alpha) = B_n^{-1} W_n^{\dot{h}}(\alpha) - B_n^{-1} \sum_{i=1}^{n} \dot{h}_{ni}(\beta)(1-\alpha) + o_p^*(1),$$

and therefore (i) follows after centering by $B_w^{-1} \sum_{i=1}^n w_{ni}(1-\alpha)$. Moreover, (ii) follows from (i) by integration with respect to the measure generated by $b(\cdot)$. ■

5. EXAMPLES

Example 1.

A Linear Model with L.R.D. Errors. Consider the model (2.1) with $h(\beta, x_i) = x_i'\beta$ where x_i's are assumed to be nonrandom. Let X be the design matrix with ith row x_i', $1 \le i \le n$. In addition to (M.1), (M.3), and (M.4), suppose that the following conditions hold:

(E.1) $(X'X)^{-1}$ exists.
(E.2) $n \max\{x_{ni}'(X'X)^{-1}x_{ni}; 1 \le i \le n\} = O(1)$.

Then by Lemma A.2(iii), $\tau_n^{-1}(X'X)^{-1/2}T(\beta(\alpha),\alpha) = O_p(1)$, and hence we can choose $D_n = (X'X)^{1/2}$, $B_n = \tau_n(X'X)^{1/2}$, and $A_n = \tau_n^{-1}(X'X)^{1/2}$. Also, $\dot{h}_{ni}(u) = x_i$ for all u and i. With these choices, it is easy to verify that (h.0), (h.1), (h.2t), (h.1b), and (h.2b) are satisfied. Finally, to verify (4.5), fix $a \in (0, \frac{1}{2})$. Note that for $M, \eta > 0$

$$\begin{aligned}
P&\Big[\sup_{\alpha\in[a,1-a]} \|A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\}\| \ge M\Big] \\
&\le P\Big[\sup_{\alpha\in[a,1-a]} \|A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\}\| \ge M, \\
&\qquad \sup_{\alpha\in[a,1-a]} \|B_n^{-1}T(\beta(\alpha) + A_n^{-1}A_n\{\hat{\beta}_n(\alpha) - \beta(\alpha)\}, \alpha)\| \le \eta\Big] \\
&\quad + P\Big[\sup_{\alpha\in[a,1-a]} \|B_n^{-1}T(\hat{\beta}_n(\alpha),\alpha)\| \ge \eta\Big] \\
&\le P\Big[\inf_{\|u\|\ge M} \sup_{\alpha\in[a,1-a]} \|B_n^{-1}T(\beta(\alpha) + A_n^{-1}u, \alpha)\| \le \eta\Big] \\
&\quad + P\Big[\sup_{\alpha\in[a,1-a]} \|B_n^{-1}T(\hat{\beta}_n(\alpha),\alpha)\| \ge \eta\Big] \\
&= T_1 + T_2,
\end{aligned}$$

say. Using the polar representation of u and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&\inf_{\|u\|\geq M} \sup_{\alpha\in[a,1-a]} \|B_n^{-1}T(\beta(\alpha)+A_n^{-1}u,\alpha)\| \\
&\quad = \inf_{\|b\|=1,\, r\geq M} \sup_{\alpha\in[a,1-a]} \|B_n^{-1}T(\beta(\alpha)+A_n^{-1}rb,\alpha)\| \\
&\quad \geq \inf_{\|b\|=1,\, r\geq M} \sup_{\alpha\in[a,1-a]} b'B_n^{-1}T(\beta(\alpha)+A_n^{-1}rb,\alpha) \\
&\quad = \inf_{\|b\|=1} \sup_{\alpha\in[a,1-a]} b'B_n^{-1}T(\beta(\alpha)+A_n^{-1}Mb,\alpha) \\
&\quad \geq \inf_{\|b\|=1} \Bigg\{ \sup_{\alpha\in[a,1-a]} Mq(\alpha) - \sup_{\alpha\in[a,1-a]} |b'B_n^{-1}T(\beta(\alpha),\alpha)| \\
&\qquad\qquad - \sup_{\alpha\in[a,1-a]} |b'[B_n^{-1}T(\beta(\alpha)+A_n^{-1}Mb,\alpha) \\
&\qquad\qquad\qquad - B_n^{-1}T(\beta(\alpha),\alpha) - Mbq(\alpha)]| \Bigg\},
\end{aligned}$$

where the second equality follows from the fact that $b'B_n^{-1}T(\beta(\alpha)+A_n^{-1}rb,\alpha)$ is monotonically nondecreasing in r. The last term in the last inequality is $o_p(1)$ by Theorem 4.2(i), and the second term is $O_p(1)$ by the choice of B_n. Therefore T_1 can be made arbitrarily small by choosing sufficiently large M. Also, T_2 can be made arbitrarily small by Lemma A.4(ii), and thus (4.5) is verified.

Example 2.

A Long Memory Time Series Model with Nonlinear Trend Function. Consider the model (2.1) with $d = 1 = p$ and $h(\beta_0,\beta_1,x) = \beta_0 + (\beta_1 + x_i)\log(\beta_1 + x)$ where the ith design point $x_i = i$, $1 \leq i \leq n$. In addition to (M.1), (M.3), and (M.4), suppose that $\beta_1 > -1$. Note that $\dot{h}_{ni}(u) = [1, 1 + \log(u_1 + x)]'$ for $u = [u_0,u_1]'$. Let $\dot{h}_{ni}(\beta) = [1,a_i]'$ where $a_i = 1 + \log(\beta_1 + i)$. Recall that $\dot{\mathcal{H}}_n$ is the $n \times 2$ matrix of rank 2 with ith row $\dot{h}_{ni}(\beta)$, and we choose $D_n = (\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n)^{1/2}$. Once we verify (h.1) with this choice of D_n, we can choose $B_n = \tau_n(\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n)^{1/2}$ and $A_n = \tau_n^{-1}(\dot{\mathcal{H}}_n'\dot{\mathcal{H}}_n)^{1/2}$ because Lemma (A.2)(iii) implies that $B_n^{-1}T(\beta(\alpha),\alpha) = O_p(1)$.

Verification of (h.0) is trivial. Assumptions (h.1) and (h.2t) can be verified by calculating the underlying quantities directly. The calculations involve the inversion of a square matrix of order 2 and showing that for $j = 1,2$,

$$e_j'D_n^{-1}\sum_{i=1}^{n}\{\dot{h}_{ni}(\beta + A_n^{-1}t)\dot{h}_{ni}'(\beta + A_n^{-1}t) - \dot{h}_{ni}(\beta)\dot{h}_{ni}'(\beta)\}D_n^{-1}e_j = o(1).$$

Similar direct calculations are used to verify (h.1b) and (h.2b). Consequently, under (4.1), the sequence of nonlinear quantiles has representations (4.2) and (4.3).

REFERENCES

Amemiya, T. (1982) Two stage least absolute deviations estimators. *Econometrica* 50, 698–711.

Arcones, M. (1994) Limit theorems for nonlinear functionals of stationary Gaussian sequence of vectors. *Annals of Probability* 22, 2242–2274.

Beran, J. (1991) *M*-estimators of location for data with slowly decaying serial correlations. *Journal of the American Statistical Association* 86, 704–708.

Beran, J. (1992) Statistical methods for data with long-range dependence. *Statistical Science* 7, 404–427.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Bloomfield, P. & W. Steiger (1983) *Least Absolute Deviations: Theory, Applications, and Algorithms.* Boston: Birkhauser.

Dehling, H. & M.S. Taqqu (1989) The empirical process of some long-range dependent sequences with an application to *U*-statistic. *Annals of Statistics* 17, 1767–1783.

Gutenbrunner, C. & J. Jurečková (1992) Regression rank scores and regression quantiles. *Annals of Statistics* 20, 305–329.

Hájek, J. & Z. Šidák (1967) *Theory of Rank Tests.* New York: Academic Press.

Jurečková, J. (1984) Regression quantiles and trimmed least squares estimator under a general design. *Kybernetika* 20, 345–357.

Jurečková, J. & B. Procházka (1994) Regression quantiles and trimmed least squares estimators in nonlinear regression models. *Journal of Nonparametric Statistics* 3, 201–222.

Jurečková, J. & P.K. Sen (1996) *Robust Statistical Procedures.* New York: Wiley.

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

Koenker, R. & B. Park (1996) An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics* 71, 265–283.

Koenker, R. & S. Portnoy (1987) *L*-estimation for linear models. *Journal of the American Statistical Association* 82, 851–857.

Koul, H.L. (1992) *Weighted Empiricals and Linear Models.* IMS Lecture Notes–Monograph Series, vol. 21. Hayward, California.

Koul, H.L. (1996a) Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Annals of Statistics* 24, 380–404.

Koul, H.L. (1996b) Asymptotics of *M*-estimators in non-linear regression with long range dependent errors. In P.M. Robinson & M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series*, vol. 2, Springer Verlag Lecture Notes Series 115. New York: Springer-Verlag.

Koul, H.L. & K. Mukherjee (1993) Asymptotics of *R*-, *MD*-, and *LAD* estimators in linear regression models with long range dependent errors. *Probability Theory and Related Fields* 95, 535–553.

Koul, H.L. & K. Mukherjee (1994) Regression quantiles and related processes under long range dependent errors. *Journal of Multivariate Analysis* 51, 318–337.

Mukherjee, K. (1994) Minimum distance estimation in linear models with long range dependent errors. *Statistics and Probability Letters* 51, 318–337.

Mukherjee, K. (1999a) Asymptotics of quantiles and rank scores in nonlinear time series. *Journal of Time Series Analysis* 20, 173–192.

Mukherjee, K. (1999b) The asymptotic distribution of a class of L-estimators under long range dependence. *Canadian Journal of Statistics* 27, 345–360.

Oberhofer, W. (1982) The consistency of nonlinear regression minimizing the l_1-norms. *Annals of Statistics* 10, 316–319.

Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.

Portnoy, S. (1991) Asymptotic behavior of regression quantiles in non-stationary dependent cases. *Journal of Multivariate Analysis* 38, 100–113.

Richardson, G.D. & B.B. Bhattachariya (1987) Consistent L_1-estimators in nonlinear regression for noncompact parameter space. *Sankhya A* 49, 46–68.

Robinson, P.M. (1992) Semiparametric methods for time series. In D. Brillinger et al. (eds.) *New Directions in Time Series Analysis*, part I, 315–326. New York: Springer-Verlag.

Weiss, A.A. (1991) Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* 7, 46–68.

Yajima, Y. (1991) Asymptotic properties of the LSE in a linear regression model with long memory stationary errors. *Annals of Statistics* 19, 158–177.

APPENDIX: LINEARIZATION RESULTS OF RANDOMLY WEIGHTED EMPIRICALS

Let $\{(\gamma_{ni},\xi_{ni});1 \le i \le n\}$ be an array of random variables and consider the following processes for $x \in \mathcal{F}$.

$$V_n(x) := \tau_n^{-1} \sum_{i=1}^{n} \gamma_{ni} I(\epsilon_i \le x + \xi_{ni}), \qquad \mu_n(x) := \tau_n^{-1} \sum_{i=1}^{n} \gamma_{ni} F(x + \xi_{ni}),$$

$$V_n^*(x) := \tau_n^{-1} \sum_{i=1}^{n} \gamma_{ni} I(\epsilon_i \le x) \quad \text{and} \quad \mu_n^*(x) := \tau_n^{-1} \sum_{i=1}^{n} \gamma_{ni} F(x).$$

The proof of Theorem A.1 is based on Lemma A.2(v), which can be viewed as the uniform closeness of the randomly weighted perturbed empirical process $V_n(\cdot) - \mu_n(\cdot)$ to the unperturbed empirical process $V_n^*(\cdot) - \mu_n^*(\cdot)$ on compact subintervals $\mathcal{F}_k$ where for $k > 0$, $\mathcal{F}_k := \mathcal{F} \cap [-k,k]$.

When $\{\epsilon_i\}$'s are weakly dependent, and $\{(\gamma_{ni},\xi_{ni});1 \le i \le n\}$ are nonrandom, one can first derive the pointwise convergence in probability to zero of $V_n(\cdot) - \mu_n(\cdot) - V_n^*(\cdot) + \mu_n^*(\cdot)$ and then demonstrate the tightness of the process $V_n(\cdot) - \mu_n(\cdot) - V_n^*(\cdot) + \mu_n^*(\cdot)$ to conclude its uniform convergence. The tightness of the process generally follows from the bound on (usually) the second or fourth moments of the product of two increment processes over disjoint intervals by some power (more than unity) of the difference of some monotonically increasing continuous function (Billingsley, 1968, Theorem 15.6). This technique was applied in Billingsley (1968, p. 141) to exhibit the tightness of ordinary empirical processes and was adopted successfully by many other researchers in suitable contexts. See Koul (1992, Lemma 2.2a.1) and Jurečková and Sen (1996, Sects. 5.2, 5.3), among others, for some examples of this technique. Because Hermite expansion is an L^2 expansion, in the L.R.D. context, we can compute bound on the second moment of the increment process (see Lemma A.1), which is the difference of monotonically increasing continuous function with power unity only, and hence Billingsley's technique can not be used to prove Lemma A.2(v). To circumvent this, Dehling and Taqqu (1989) came up with an ingenious chaining argument to obtain the uniform convergence of the ordinary empirical process ($V_n^*(\cdot) - \mu_n^*(\cdot)$ with $\gamma_{ni} = n^{-1/2}$) to the process $J_m(\cdot)Z_m$, where Z_m is some random variable. Koul and Mukherjee (1993, Theorem 1.1) used a similar chaining argument to obtain a variant of Lemma A.2 when the weights $\{(\gamma_{ni},\xi_{ni});1 \le i \le n\}$ are nonrandom. Because the proofs of Lemma A.1 and Lemma A.2 of this paper are similar to the those of Lemma 2.1 and Theorem 1.1 of Koul and Mukherjee, we present only brief outlines of the proofs here. Note that the conclusion of Lemma A.2(v) (uniform convergence over $\mathcal{F}_k$) is weaker than the conclusion of Theorem 1.1 of Koul and Mukherjee (uniform convergence over $\mathcal{F}$) because our (M.3) and (M.4) are also weaker than the corresponding conditions (A.5) and (A.6) of Koul and Mukherjee.

Accordingly, for $x \in \mathcal{F}$, define

$$\mathcal{S}_n(x) := \tau_n^{-1} \sum_{i=1}^{n} \gamma_{ni}\{I(\epsilon_i \le x + \xi_{ni}) - F(x + \xi_{ni}) - (m!)^{-1} J_m(x + \xi_{ni}) H_m(\eta_i)\},$$

$$\mathcal{S}_n^*(x) := \tau_n^{-1} \sum_{i=1}^{n} \gamma_{ni}\{I(\epsilon_i \le x) - F(x) - (m!)^{-1} J_m(x) H_m(\eta_i)\}.$$

LEMMA A.1. *Suppose that* (M.1) *holds. Let* $\mathcal{A}_n$ *and* $\mathcal{B}_n$ *denote the sigma fields generated by* $\{(\gamma_{ni},\xi_{ni});1 \le i \le n\}$ *and* $\{\epsilon_i;1 \le i \le n\}$, *respectively*, *satisfying the following*.

(A.0) *For each* $n \ge 1$, $\mathcal{A}_n$ *and* $\mathcal{B}_n$ *are independent. Then*

$$\tau_n^2 E[\mathcal{S}_n(y) - \mathcal{S}_n(x)]^2 \le \sum_{i=1}^{n}\sum_{j=1}^{n} E[|\gamma_{ni}\gamma_{nj}||\{F(y+\xi_{ni}) - F(x+\xi_{ni})\} \times \{F(y+\xi_{nj}) - F(x+\xi_{nj})\}|^{1/2}]|\rho(i-j)|^{(1+m)}.$$

Proof. Write $E[\mathcal{S}_n(y) - \mathcal{S}_n(x)]^2 = E[E\{[\mathcal{S}_n(y) - \mathcal{S}_n(x)]^2|\mathcal{A}_n\}]$. Evaluate the conditional expectation by its Hermite expansion as in Lemma 2.1 of Koul and Mukherjee to obtain the result. ■

For the next result, consider the following assumptions:

(A.1) $n \max\{E\gamma_{ni}^2;1 \le i \le n\} = O(1)$.
(A.2) $0 < k_1 < \sum_{i=1}^n E\gamma_{ni}^2$, for some constant k_1.
(A.3) $\max\{|\xi_{ni}|;1 \le i \le n\} = o_p(1)$.
(A.4) $\tau_n^{-1}\sum_{i=1}^n E|\gamma_{ni}\xi_{ni}| = O(1)$.

LEMMA A.2.

(i) *Under* (M.1), (M.3), *and* (A.0)–(A.3), *for each* $x \in \mathcal{F}_k$,

$$V_n(x) - \mu_n(x) - V_n^*(x) + \mu_n^*(x) = o_p(1).$$

(ii) *Under* (M.1), (A.0)–(A.2), $\sup\{|\mathcal{S}_n^*(x)|; x \in \mathcal{F}_k\} = o_p(1)$.
(iii) *Under* (M.1), (A.0), *and* (A.1), *for each* $x \in \mathcal{F}_k$,

$$V_n^*(x) - \mu_n^*(x) = O_p(1).$$

(iv) *Under* (M.1), (M.4), (A.0), (A.1), *and* (A.3),

$$\sup\left\{\tau_n^{-1}\left|\sum_{i=1}^n \gamma_{ni}\{J_m(x+\xi_{ni}) - J_m(x)\}(m!)^{-1}H_m(\eta_i)\right|; x \in \mathcal{F}_k\right\} = o_p(1).$$

(v) *Under* (M.1), (M.3), (M.4), *and* (A.0)–(A.4), $\sup\{|\mathcal{S}_n(x)|; x \in \mathcal{F}_k\} = o_p(1)$ *and hence*

$$\sup\{|V_n(x) - \mu_n(x) - V_n^*(x) + \mu_n^*(x)|; x \in \mathcal{F}_k\} = o_p(1).$$

Proof. Using the L^2 convergence to zero, we can prove (i). For proving (ii) and (v), construct a chain with $\kappa = [\log_2\{\lambda(d)\sum_{i=1}^n E|\gamma_{ni}|/(\delta\tau_n)\}] + 1$ as in Koul and Mukherjee (display 9) and use Lemma 2.1 to proceed analogously. Assertion (iii) follows from the fact that $E[V_n^*(x) - \mu_n^*(x)]^2 = O(1)$. For (iv), one can use an argument similar to Koul and Mukherjee (display 21). ∎

Let $\{(\gamma_{ni}(t), \xi_{ni}(t)); 1 \le i \le n, t \in \mathbb{R}^{1+d}\}$ be an array of real valued stochastic processes with $\xi_{ni}(0) = 0$, for all $1 \le i \le n$, and consider the following processes for $x \in \mathcal{F}, t \in \mathbb{R}^{1+d}$.

$$V_n(x,t) := \tau_n^{-1}\sum_{i=1}^n \gamma_{ni}(t)I(\epsilon_i \le x + \xi_{ni}(t)), \qquad \mu_n(x,t) := \tau_n^{-1}\sum_{i=1}^n \gamma_{ni}(t)F(x+\xi_{ni}(t)).$$

For the next result, suppose that the following conditions hold for each fixed $t \in \mathbb{R}^{1+d}$ and $b > 0$.

(B.0) For each $n \ge 1$, the sigma fields generated by $\{\gamma_{ni}(\cdot), \xi_{ni}(\cdot); 1 \le i \le n\}$ and $\{\epsilon_i; 1 \le i \le n\}$ are independent.
(B.1) $n \max\{E\gamma_{ni}^2(0); 1 \le i \le n\} = O(1)$.
(B.2t) $n \max\{E[\gamma_{ni}(t) - \gamma_{ni}(0)]^2; 1 \le i \le n\} = o(1)$ and $0 < k_1(t) < \sum_{i=1}^n E\gamma_{ni}^2(t)$ for some constant $k_1(t)$.
(B.3t) $\max\{|\xi_{ni}(t)|; 1 \le i \le n\} = o_p(1)$.
(B.4t) $\tau_n^{-1}\sum_{i=1}^n E|\gamma_{ni}(t)\xi_{ni}(t)| = O(1)$.
(B.1b) For every $\alpha > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left[\sup\left\{\tau_n^{-1}\sum_{i=1}^n |\gamma_{ni}(s) - \gamma_{ni}(t)|; s,t \in \mathcal{N}_b, \|s-t\| \le \delta\right\} < \alpha\right] \ge 1-\alpha.$$

(B.2b) For every $\alpha > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left[\sup\left\{\tau_n^{-1}\sum_{i=1}^n |\gamma_{ni}(s)\|\xi_{ni}(s) - \xi_{ni}(t)|; s,t \in \mathcal{N}_b, \|s-t\| \le \delta\right\} < \alpha\right] \ge 1-\alpha.$$

LEMMA A.3. *Suppose that* (M.1), (M.3), (B.0), (B.1), (B.2t), (B.3t), (B.1b), *and* (B.2b) *hold for every* $t \in \mathbb{R}^{1+d}$ *and* $b > 0$.

(i) *Then for every* $x \in \mathcal{F}_k$,

$$\sup\{|V_n(x,t) - \mu_n(x,t) - V_n(x,0) + \mu_n(x,0)|; t \in \mathcal{N}_b\} = o_p(1).$$

(ii) *If, in addition,* (M.4) *and* (B.4t) *hold, then*

$$\sup\{|V_n(x,t) - \mu_n(x,t) - V_n(x,0) + \mu_n(x,0)|;(x,t) \in \mathcal{F}_k \times \mathcal{N}_b\} = o_p(1).$$

Proof. Write $V_n(x,t) - \mu_n(x,t) - V_n(x,0) + \mu_n(x,0) = U_1(x,t) + U_2(x,t)$, where

$$U_1(x,t) := \tau_n^{-1} \sum_i \{\gamma_{ni}(t) - \gamma_{ni}(0)\}\{I(\epsilon_i \le x) - F(x)\},$$

and

$$U_2(x,t) := \tau_n^{-1} \sum_i \gamma_{ni}(t)\{I(\epsilon_i \le x + \xi_{ni}(t)) - F(x + \xi_{ni}(t)) - I(\epsilon_i \le x) + F(x)\}.$$

The pointwise convergence of U_1 follows by showing that $E[U_1(x,t)]^2 = o(1)$. Tightness of U_1 follows from (B.1b). For U_2, note that by Lemma A.2(v), $\sup\{\|U_2(x,t)\|; x \in \mathcal{F}_k\} = o_p(1)$. Now the tightness of the process $[\tau_n^{-1}\sum_i \gamma_{ni}(t)\{I(\epsilon_i \le x) - F(x)\}]$ follows from (B.1b), whereas that of $[\tau_n^{-1}\sum_i \gamma_{ni}(t)\{I(\epsilon_i \le x + \xi_{ni}(t)) - F(x + \xi_{ni}(t))\}]$ follows as in Koul (1996b, Lemma 3.2). ■

Next, fix $k, b > 0$ and consider the following processes for $(x,t) \in \mathcal{F}_k \times \mathcal{N}_b$.

$$V_n(x,t) := \tau_n^{-1} \sum_{i=1}^{n} D_n^{-1} \dot{h}_{ni}(\beta + A_n^{-1}t) I\{\epsilon_i \le x + h_{ni}(\beta + A_n^{-1}t) - h_{ni}(\beta)\},$$

and

$$\mu_h(x,t) := \tau_n^{-1} \sum_{i=1}^{n} D_n^{-1} \dot{h}_{ni}(\beta + A_n^{-1}t) F\{x + h_{ni}(\beta + A_n^{-1}t) - h_{ni}(\beta)\}.$$

THEOREM A.1. *Assume that* (M.1)–(M.3), (h.0), (h.1), (h.2t), (h.1b), *and* (h.2b) *hold. Then for each* $x \in \mathcal{F}_k$,

(i) $\sup\{\|V_h(x,t) - V_h(x,0) - \mu_h(x,t) + \mu_h(x,0)\|; t \in \mathcal{N}_b\} = o_p(1)$ *and consequently*,

$$\sup\left\{\left\| V_h(x,t) - V_h(x,0) - tf(x) - B_n^{-1} \sum_{i=1}^{n} \{\dot{h}_{ni}(\beta + A_n^{-1}t) - \dot{h}_{ni}(\beta)\}F(x)\right\|; \; t \in \mathcal{N}_b\right\} = o_p(1).$$

(ii) *If, in addition,* (M.4) *holds, then*

$$\sup\{\|V_h(x,t) - V_h(x,0) - \mu_h(x,t) + \mu_h(x,0)\|;(x,t) \in \mathcal{F}_k \times \mathcal{N}_b\} = o_p(1)$$

and consequently,

$$\sup\left\{\left\| V_h(x,t) - V_h(x,0) - tf(x) - B_n^{-1} \sum_{i=1}^{n} \{\dot{h}_{ni}(\beta + A_n^{-1}t) - \dot{h}_{ni}(\beta)\}F(x)\right\|; \; (x,t) \in \mathcal{F}_k \times \mathcal{N}_b\right\} = o_p(1).$$

Proof. Applying Lemma A.3 with $\gamma_{ni}(t)$ equal to the kth coordinate of $D_n^{-1}\dot{h}_{ni}(\beta + A_n^{-1}t)$ and $\xi_{ni}(t)$ equal to $h_{ni}(\beta + A_n^{-1}t) - h_{ni}(\beta)$, $1 \le k \le 1 + d$, and noting that $\mu_h(x,t) - \mu_h(x,0) = tf(x) + B_n^{-1}\sum_{i=1}^n \{\dot{h}_{ni}(\beta + A_n^{-1}t) - \dot{h}_{ni}(\beta)\}F(x) + o_p(1)$, we obtain the conclusions. ∎

LEMMA A.4.

(i) *Under* (h.0), (2.3) *holds*.
(ii) *Under* (h.0) *and* (h.1), $B_n^{-1}T(\hat{\beta}_n(\alpha),\alpha) = o_p(1)$.

Proof. Because ρ_α is a convex function and $h_{ni}(u)$ is differentiable in u, all directional derivatives of $\sum_{i=1}^n \rho_\alpha(Y_{ni} - h_{ni}(u))$ exist and are positive at a minimum $\hat{\beta}_n(\alpha)$. Let $\mathcal{D}_i(u,w)$ denote the w-directional derivative of $\rho_\alpha(Y_{ni} - h_{ni}(u)) = (h_{ni}(u) - Y_{ni})I(h_{ni}(u) - Y_{ni} > 0) - \alpha(h_{ni}(u) - Y_{ni})$ with respect to $u \in \Omega$, where $w \in W := \{u \in \mathbb{R}^{1+d}; \|u\| = 1\}$. Then, using the chain rule for differentiation,

$$
\begin{aligned}
\mathcal{D}_i(u,w) &= w'\dot{h}_{ni}(u)(1-\alpha), \quad \text{if } h_{ni}(u) - Y_{ni} > 0,\\
&= w'\dot{h}_{ni}(u) \times -\alpha, \quad \text{if } h_{ni}(u) - Y_{ni} < 0,\\
&= w'\dot{h}_{ni}(u)(1-\alpha), \quad \text{if } h_{ni}(u) - Y_{ni} = 0 \quad \text{and} \quad w'h_{ni}(u) > 0,\\
&= w'\dot{h}_{ni}(u) \times -\alpha, \quad \text{if } h_{ni}(u) - Y_{ni} = 0 \quad \text{and} \quad w'h_{ni}(u) < 0.
\end{aligned}
$$

Let $\dot{h}_{ni}(\hat{\beta}_n(\alpha))$ be denoted by $g_{ni} = [g_{ni,1},\ldots,g_{ni,1+d}]'$, $1 \le i \le n$. Then, for each $w \in W$, we have with probability one,

$$
\begin{aligned}
0 &< \sum_{i=1}^n \mathcal{D}_i(\hat{\beta}_n(\alpha),w)\\
&= w'\Bigg[\sum_{i \notin B_n(\alpha)} g_{ni}\{I(Y_{ni} \le h_{ni}(\hat{\beta}_n(\alpha))) - \alpha\}\\
&\qquad + \sum_{i \in B_n(a); w'g_{ni} > 0} g_{ni}(1-\alpha) + \sum_{i \in B_n(\alpha); w'g_{ni} < 0} g_{ni} \times -\alpha\Bigg].
\end{aligned} \tag{A.1}
$$

Now, choosing $w = e_k := [0,\ldots,0,1,0,\ldots,0]$, $1 \le k \le 1 + d$ and applying (A.1) we get

$$
\begin{aligned}
&\sum_{i \notin B_n(\alpha)} g_{ni,k}\{I(Y_{ni} \le h_{ni}(\hat{\beta}_n(\alpha))) - \alpha\} + \sum_{i \in B_n(\alpha); g_{ni,k} > 0} g_{ni,k}(1-\alpha)\\
&\qquad + \sum_{i \in B_n(\alpha); g_{ni,k} < 0} g_{ni,k} \times -\alpha > 0,
\end{aligned}
$$

and hence,

$$
\begin{aligned}
&\sum_{i \notin B_n(\alpha)} g_{ni,k}\{I(Y_{ni} \le h_{ni}(\hat{\beta}_n(\alpha))) - \alpha\} > (\alpha - 1)\sum_{i \in B_n(\alpha); g_{ni,k} > 0} g_{ni,k}\\
&\qquad + \alpha \sum_{i \in B_n(\alpha); g_{ni,k} < 0} g_{ni,k}.
\end{aligned} \tag{A.2}
$$

Similarly, applying (A.1) once more with $w = -e_k$, we get

$$(\alpha - 1) \sum_{i \in B_n(\alpha); g_{ni,k} < 0} g_{ni,k} + \alpha \sum_{i \in B_n(\alpha); g_{ni,k} > 0} g_{ni,k}$$
$$> \sum_{i \notin B_n(\alpha)} g_{ni,k} \{ I(Y_{ni} \leq h_{ni}(\hat{\beta}_n(\alpha))) - \alpha \}. \qquad \textbf{(A.3)}$$

Putting $k = 1$ in (A.2) and (A.3) and using $g_{ni,1} = 1$ for all $1 \leq i \leq n$, we get

$$N(1 - \alpha) + P(0 - \alpha) > (\alpha - 1)Z + \alpha \times 0$$

and

$$(\alpha - 1) \times 0 + \alpha Z > N(1 - \alpha) + P(0 - \alpha).$$

Now (2.3) follows by using $n = N + P + Z$.

For (ii), combining (A.2) and (A.3) we get that for all $1 \leq k \leq 1 + d$,

$$\sum_{i \in B_n(\alpha); g_{ni,k} < 0} g_{ni,k} = \left[(\alpha - 1) \sum_{i \in B_n(\alpha); g_{ni,k} > 0} g_{ni,k} + \alpha \sum_{i \in B_n(\alpha); g_{ni,k} < 0} g_{ni,k} \right]$$
$$+ \sum_{i \in B_n(\alpha)} g_{ni,k} \{ I(Y_{ni} \leq h_{ni}(\hat{\beta}_n(\alpha))) - \alpha \}$$
$$< \sum_{i \notin B_n(\alpha)} g_{ni,k} \{ I(Y_{ni} \leq h_{ni}(\hat{\beta}_n(\alpha))) - \alpha \} + \sum_{i \in B_n(\alpha)} g_{ni,k} \{ I(Y_{ni} \leq h_{ni}(\hat{\beta}_n(\alpha))) - \alpha \}$$
$$= T_k(\hat{\beta}_n(\alpha), \alpha)$$
$$< (\alpha - 1) \sum_{i \in B_n(\alpha); g_{ni,k} < 0} g_{ni,k} + \alpha \sum_{i \in B_n(\alpha); g_{ni,k} > 0} g_{ni,k}$$
$$+ \sum_{i \in B_n(\alpha)} g_{ni,k} \{ I(Y_{ni} \leq h_{ni}(\hat{\beta}_n(\alpha))) - \alpha \} = \sum_{i \in B_n(\alpha); g_{ni,k} > 0} g_{ni,k},$$

where $T_k(\hat{\beta}_n(\alpha), \alpha)$ is the kth $(1 \leq k \leq 1 + d)$ coordinate of $T(\hat{\beta}_n(\alpha), \alpha)$. Now the conclusion follows from (h.0) and (h.1). ∎