This is to certify that the

dissertation entitled

Weak Convergence of Weighted Empirical Processes Under Long Range Dependence with Applications to Robust Estimation in Linear Models

presented by

Kanchan Mukherjee

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Statistics

Major professor

Date July 12, 1993

WEAK CONVERGENCE OF WEIGHTED EMPIRICAL PROCESSES UNDER LONG RANGE DEPENDENCE WITH APPLICATIONS TO ROBUST ESTIMATION IN LINEAR MODELS

By

Kanchan Mukherjee

A DISSERTATION

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1993

ABSTRACT

WEAK CONVERGENCE OF WEIGHTED EMPIRICAL PROCESSES UNDER LONG RANGE DEPENDENCE WITH APPLICATIONS TO ROBUST ESTIMATION IN LINEAR MODELS

By

Kanchan Mukherjee

A discrete time stationary stochastic process is said to be long range dependent if its covariances decrease to zero like a power of lag as the lag tends to infinity but their absolute sum diverges. In this dissertation, a *uniform closeness result* of weighted residual empirical process to its natural estimator is derived under the LRD setup. These are then used to prove the *asymptotic uniform linearity* of a class of linear rank statistics and the *asymptotic uniform quadraticity* of a class of L^2- distance statistics. These results, in turn, are applied to investigate the asymptotic behavior of the above estimators in a linear regression model when the errors are function of LRD Gaussian random variables.

Some intriguing phenomena are observed in connection with the inherent nature of the limiting distributions of the above estimators. Unlike the weakly dependent case the limiting distributions may not be normal always. Moreover, when the errors are LRD Gaussian and the design matrix is centered, the asymptotic covariances of the class of rank and minimum distance estimators become those of the least square estimator- a phenomenon which is in complete contrast with the i.i.d. error case. Similar statement applies to LAD and a large class of M- estimators. These results are proved under some conditions on the design matrix that are very similar to those under the i.i.d. setup

The dissertation also considers the asymptotic behavior of regression quantiles and regression rank scores which are natural generalization of the notion of order statistics and regression rank scores processes from the one sample model to the linear model. Under the LRD error setup the aforementioned *uniform closeness result* is used to obtain the asymptotic representations of regression

quantiles and the asymptotic uniform linearity of regression rank score statistics. These results are useful for testing subhypotheses and for estimation in the presence of nuisance parameters.

Finally, as a byproduct, extension of some of the results of Dehling and Taqqu (The empirical process of some long range dependent sequences with an application to U-statistics. 1989. *Ann. Statist.*) from the ordinary empirical to the weighted empirical processes are derived.

To my parents and grandparents

ACKNOWLEDGMENTS

I would like to express my deepest regard to Professor Hira Lal Koul for his careful guidance, encouragement and affection throughout my graduate study at Michigan State University. His erudition, dedication and contribution to statistics have been my main source of inspiration to work in the area of nonparametric statistics.

I would like to thank Professors Joseph Gardiner, James Hannan and Habib Salehi for serving on my thesis committee and for their suggestions leading to the improvement of my dissertation. I owe my special thanks to Professors Dennis Gilliland, Connie Page and James Stapleton from whom I learned many aspects of good teaching and statistical consulting.

The dedication of my parents and grandparents to the pursuit of my higher education is greatly appreciated. The support and encouragement of my sisters, Baishakhi and Mousumi, my friend Dr. Suman Majumdar, and his wife Jayasri during my doctoral study were invaluable. The careful guidance provided by my teachers at *UPANISHAD* during the early stage of my education deserves special mention.

TABLE OF CONTENTS

CHAPTER 0

INTRODUCTION

A discrete time stationary stochastic process is called a *long range dependent* (LRD) or a *long memory* process if its correlations decay to zero like a power of lag as the lag tends to infinity but their absolute sum diverges. Quite often, econometric and time series data appears to be stationary and exhibit strong correlation between observations separated by large lag which does not decay to zero at a fast enough rate to be absolutely summable. Similar phenomena have been observed in hydrology in connection with the construction of the Aswan dam over the river Nile, Egypt, when hydrologist Hurst noticed that the annual volume of river-flow shows long term behavior over time (Mandelbrot and Van Ness, 1966). Mandelbrot and Van Ness proposed *fractional gaussian noise* to model observations with such strong correlation. Later, Granger and Joyeux (1980) and Hosking (1981) independently came up with *fractional* ARIMA model to include more processes with non-gaussian marginal distributions. The salient features of these processes are that their spectral density diverges at zero and their correlations are not absolutely summable and that creates considerable mathematical difficulties for its analysis.

The usefulness of LRD processes in modeling a wide variety of physical phenomena heralded an upsurge of interest among many researchers who explored different probabilistic aspects of LRD processes in the last two decades. Investigation of the behavior of different statistics and estimators based on LRD observations are also carried out by some authors. Taqqu (1975) obtained the *weak convergence results of partial sum processes* based on random variables (r.v.'s) that are a measurable function of LRD Gaussian r.v.'s.Taqqu characterized the limiting distribution of the partial sum process for r.v.'s having Hermite rank (see Remark 1.2.1) one and two. Later, Dobrushin and Major (1979) and Taqqu (1979) independently characterized the limiting process (called *Hermite process*) for r.v.'s having arbitrary Hermite ranks through a multiple Weiner-Ito integral representation. It was observed that the limiting process is non-gaussian if its Hermite rank is more than unity.

Along with these technical results, parallel research has proliferated on the estimation of some parameters describing the correlation structure. Fox and Taqqu (1986) and Yajima (1985) proposed *maximum likelihood estimation* of the index of the LRD θ (see (1.1.2)) based on LRD Gaussian sequence and Yajima (1985) considered the *least square estimation* (l.s.e.) of θ based on LRD r.v.'s.

In linear regression model with LRD errors, Yajima (1988, 1991) obtained the strong consistency and the asymptotic distribution of the l.s.e. of the regression parameters under some conditions on the cumulants of the marginal error distribution function. Koul (1992a) derived the *asymptotic uniform linearity* (AUL) of M-statistic and limiting representation of normalized M-estimators in a regression model when the errors are a function of LRD Gaussian r.v. and the score function is absolutely continuous.

Motivated by the seminal work of Koul (1992a), in this dissertation we derive the asymptotic representation of some more robust estimators of the regression parameters in a linear regression model when the errors are a function of LRD Gaussian r.v.'s. In particular, we consider the behavior of a class of *rank estimator* (R-estimator) proposed by Jureckova (1969) and Jaeckel (1972), and *minimum distance estimators* (m.d.e.) proposed by Koul and Dewet (1983) and

Koul (1985b). Finally, we also consider regression quantiles (RQ) proposed by Koenker and Basset (1978) of which the special case is the *least absolute deviation estimator* (LAD).

The investigation of the behavior of robust estimators based on dependent observations started with the work of Gastwirth and Rubin (1975). Gastwirth and Rubin studied the behavior of R-, M- and L-estimators in a location model under Δ-mixing errors. Koul (1977) generalized their results to linear regression model with strongly mixing errors which contains Δ-mixing class. In all of the above weakly dependent cases the correlations are absolutely summable and thus the effect of dependency becomes relegated, at least asymptotically. Consequently, the limiting distributions of the suitably normalized estimators is Gaussian as in the case of independent identically distributed (i.i.d.) errors. But, in the case of the LRD errors, the limiting distributions of these estimators are quite different from their weakly dependent counterparts in two fundamental ways. First of all, the normalizing factors are different and secondly the limiting distribution is not always normal. For more on these see Remark 2.2.1.

The crux of proving AUL theorems in linear models is unified in the work of Koul (1991, 1992) in the form of a *uniform closeness theorem* for a weighted residual empirical process to its natural estimate. Hence the fundamental tool for proving most of the results in this dissertation is the *uniform closeness result* of weighted residual empirical process to its natural estimate in a linear regression setting when the errors are a function of LRD Gaussian r.v.

The technical difficulties for proving the *uniform closeness result* in this setup is surmounted by using a modification of an ingenious *chaining argument* of Dehling and Taqqu (1989). Dehling and Taqqu came up with a *chaining argument* to prove the *uniform weak reduction principle* of ordinary empirical process. Theorem 3.1 of Dehling and Taqqu (1989) obtains an upper-bound for

$$P[\sup_{k \le n,\ x \in I} |\tau_n^{-1} \sum_{i=1}^{k} n^{-1/2} \{I(\epsilon_i \le x) - F(x) - J_m(x)\ (m!)^{-1} H_m(\eta_i)\}| > \delta],$$

which converges to zero $\forall \delta > 0$. Here m, J_m, τ_n and H_m are defined in Section 1.2. In Theorem 1.2.1 of this dissertation, we invoke a similar *chain* to prove that

$$P[\sup_{x \in I} |\tau_n^{-1} \sum_{i=1}^{n} \gamma_{ni} \{I(\epsilon_i \le x + \xi_{ni}) - F(x + \xi_{ni}) - J_m(x + \xi_{ni})(m!)^{-1}\} H_m(\eta_i) > \delta]$$

converges to zero. Then, this result is used to derive the *uniform closeness result.* In other words, we obtain a partial generalization of the *uniform weak reduction principle* of Dehling and Taqqu from the ordinary empirical process to a very general weighted empirical processes with nonzero ξ_{ni}, $1 \le i \le n$. This partial generalization allows us to prove the *uniform closeness result*, which along with some other results in Chapter 1, is used to obtain the asymptotic representations of the rank estimators, minimum distance estimators and regression quantiles. As a byproduct of Theorem 1.2.1, we also obtain the *weak convergence results* of weighted empiricals. In our case the novelty in proving Theorem 1.2.1 lies in choosing the chain suitable for weighted empirical process which, of course, reduces to the Dehling and Taqqu chain when the weighted empirical reduces to the ordinary empirical.

Notation. In this dissertation, I(A) denotes the indicator function of an event A. The index i in the summation varies from 1 to n unless specified otherwise. For a vector $\mathbf{u} \in R^p$, $\mathbf{u}^t$ ($\|\mathbf{u}\|$) denotes its transpose (Euclidian norm). If $\mathbf{D}$ is an $n \times p$ matrix, then $\mathbf{d}_{ni}^t$, $1 \le i \le n$, denotes its ith row and $\mathbf{d}_{.j}, 1 \le j \le p$, its jth column. For $b \in (0, \infty)$, the set $\{s \in R^p : \|s\| \le b\}$ is denoted by $N(b)$. For a sequence of numbers (r.v.'s), $O(1)$ ($O_p(1)$) denotes its boundedness (in probability) and $o(1)$ ($o_p(1)$) denotes the convergence to zero (in probability). An expression number (x.y) of the section x of chapter z is called (x.y) in chapter z, (z.x.y) in other.

CHAPTER 1

PRELIMINARIES

1.1. Introduction.

We consider the following multiple linear regression model in this dissertation. Let $\{\eta_i,\ i \geq 1\}$ be a stationary, mean zero, unit variance Gaussian process with correlation $\rho(k) := E(\eta_1 \eta_{1+k})$, $k \geq 1$. Suppose $\epsilon_i := G(\eta_i)$, $i \geq 1$, where G is a measurable function from R^1 to R^1 and let X denote the $n \times p$ design matrix of known constants whose ith row is x_{ni}^t, $1 \leq i \leq n$. Consider a linear model where one observes the response variable $\{Y_{ni}\}$, $1 \leq i \leq n$, satisfying,

$$(1.1) \qquad Y_{ni} = x_{ni}^t \beta + \epsilon_i,\ 1 \leq i \leq n, \quad \text{for some } \beta \in R^p.$$

The long range dependence of the r.v.'s $\{\eta_i\}$ is implied by assuming that for some $0 < \theta < 1$,

$$(1.2) \qquad \rho(k) = k^{-\theta} L(k), \quad k \geq 1,$$

where $L(k)$ is a slowly varying function at infinity, i.e., $L(tx)/L(t) \rightarrow 1$ as $t \rightarrow \infty$ for every $x > 0$. We assume that $L(k)$ is positive for large k . From Lemma VIII.8 of Feller (1968, vol 2), it follows that $\sum_1^\infty \rho(k) = \infty$. Examples of such functions L are positive constants or $L(k) := \log k$.

In this dissertation, we derive the asymptotic representation of some familiar estimators of the regression parameter β that are known to be robust in the linear models with independent errors. In particular, we consider a family of R-estimators, m.d.e., and regression quantiles. In this chapter, we describe the basic probabilistic results and their proofs that are needed throughout the rest of this dissertation.

The technique of obtaining asymptotic representation of suitably normalized estimators defined as a solution of a system of equations goes back to Cramer (1946). The basic idea is to derive an asymptotic *Taylor type expansion* of a suitable score function and to ensure the existence of stochastically bounded solutions. Therefore using the same technique with suitable modifications one can obtain the asymptotic representation of R-estimators and M-estimators, defined as a solution of a system of equations. In linear regression models, different authors have used different techniques to derive a *Taylor type expansion*, see, e.g., Koul (1969) and Jureckova (1971) for R-estimators, among others. Koul (1991) envisaged a unified approach to these problems as a consequence of *uniform closeness* of some weighted residual empirical processes to its natural estimate, centered at its expectation. Using *uniform closeness* and the smoothness of the error distribution function one can obtain an one step *Taylor Type Expansions* of the nonsmooth empirical processes and *asymptotic uniform linearity* is a consequence of that.

1.2. Uniform Closeness of Weighted Empiricals.

Let ϵ (η) have the same distribution as that of the marginal distribution of $\{\epsilon_i,\ i \geq 1\}$ ($\{\eta_i,\ i \geq 1\}$). Let F be the distribution function of ϵ and I := {x: 0<F(x)<1}; $\{H_q : q \geq 1\}$ denote the Hermite polynomials (see Remark 2.1 for definition). It has the following properties:

(2.1) $\quad EH_q(\eta) = 0,\ q \geq 1, \quad EH_q(\eta_j)H_q(\eta_k) = q!\rho^q(j - k)\ \forall q \geq 0,\ \forall j,\ k \geq 1$, and

$$EH_q(\eta_j)H_r(\eta_k) = 0\ \forall q \neq r,\ \forall j,\ k \geq 1.$$

A quick proof of the last fact with $j \neq k$ can be given as follows. Suppose that $q < r$ and write $\eta_j = \rho\eta_k + \eta^*$, where $\rho := \rho(j - k)$. Observe that $H_q(\rho\eta_k + \eta^*)$ is a polynomial of degree q in η_k with coefficients involving η^*. Since η_k and η^* are independent the result follows from $E\ \eta_k^c H_r(\eta_k) = 0\ \forall 0 \leq c < k$ which is a consequence of the fact that the k th degree Hermite polynomial is obtained through the Gram-Schmidt orthogonalization of the square integrable r.v.'s $\{\eta_k^c,\ 1 \leq c \leq k\}$.

Define,

$$J_q(x) := EI(G(\eta) \leq x)H_q(\eta),\ J_q^+(x) := EI(G(\eta) \leq x)|H_q(\eta)|\ x \in I,\ q \geq 1,$$

and let m be the Hermite rank of the class of functions $\{I(G(\eta) \leq x),\ x \in I\}$ introduced by Dehling and Taqqu (1989), i.e.,

(2.2) $$m := \text{minimum}\{m(x) \colon x \in I\}$$

where

$$m(x) := \text{minimum}\{q \geq 1;\ J_q(x) \neq 0\}.$$

In view of (2.2) we have,

(2.3) $$I(G(\eta_i) \leq x) - F(x) = \sum_{q \geq m} J_q(x)/q!H_q(\eta_i),\quad \forall\ x\ \in\ I,\quad \forall\ i \geq 1,$$

where the above equality is in the L^2-sense.

For L and θ in (1.2) define $\tau_n := n^{(1-m\theta)/2}L^{m/2}(n),\ n \geq 1$. **Throughout this dissertation, we assume that $0 < \theta < 1/m$.**

Let $\{\gamma_{ni}, \xi_{ni};\ 1 \leq i \leq n\}$ be arrays of real numbers and define, for $x \in I$,

$$V_n(x) := \tau_n^{-1}\sum_i \gamma_{ni} I(\epsilon_i \leq x + \xi_{ni}), \qquad V_n^*(x) := \tau_n^{-1}\sum_i \gamma_{ni} I(\epsilon_i \leq x),$$

$$\mu_n(x) := \tau_n^{-1}\sum_i \gamma_{ni} F(x + \xi_{ni}), \qquad \mu_n^*(x) := \tau_n^{-1}\sum_i \gamma_{ni} F(x),$$

$$U_n(x) := V_n(x) - \mu_n(x), \qquad U_n^*(x) := V_n^*(x) - \mu_n^*(x)$$

$$S_n(x) := \tau_n^{-1} \sum_i \gamma_{ni}\{I(\epsilon_i \le x + \xi_{ni}) - F(x + \xi_{ni}) - J_m(x + \xi_{ni})(m!)^{-1}\}H_m(\eta_i),$$

$$S_n^*(x) := \tau_n^{-1} \sum_i \gamma_{ni}\{I(\epsilon_i \le x) - F(x) - J_m(x)\,(m!)^{-1}H_m(\eta_i)\}.$$

The following theorem obtains the *uniform closeness* of the processes U_n and U_n^*. Here, $\epsilon_i := G(\eta_i)$ and (1.2) is satisfied. Also τ_n is defined as above.

Theorem 2.1. *Suppose that the weights* $\{\gamma_{ni};\ 1 \le i \le n\}$ *satisfy the following conditions:*

(A.1) $\quad n \max_{1 \le i \le n} \gamma_{ni}^2 = O(1). \qquad$ (A.2) $\quad \sum_1^n \gamma_{ni}^2 = 1, \ \forall\ n \ge 1.$

Then,

$$\sup_{x \in I} |S_n^*(x)| = o_p(1). \tag{2.4}$$

Moreover, assume that (A.3), (A.4), (A.5) *and* (A.6) *hold, where*

(A.3) $\quad \max_{1 \le i \le n} |\xi_{ni}| = o(1), \qquad$ (A.4) $\quad \tau_n^{-1} \sum_i |\gamma_{ni}\xi_{ni}| = O(1),$

(A.5) *The d.f.* F *of* ϵ *has uniformly continuous density* f *on* I, f>0 *a.e.* (*Lebesgue*) on I,

(A.6) *The functions* J_m *and* J_m^+ *are continuously differentiable with respective derivatives* J_m' *and* $J_m^{+\prime}$. *Moreover,* $J_m'(x)$ *and* $J_m^{+\prime}(x)$ *converge to zero as* x *converges to* $c := \inf I$ *and* $d := \sup I$.

Then,

(2.5) $$\sup_{x \in I} |S_n(x)| = o_p(1),$$

(2.6) $$\sup_{x \in I} |U_n(x) - U_n^*(x)| = o_p(1),$$

and

(2.7) $$\sup_{x \in I} |\tau_n^{-1} \sum_i \gamma_{ni}\{I(\epsilon_i \le x + \xi_{ni}) - I(\epsilon_i \le x)\} - \tau_n^{-1} \sum_i \gamma_{ni}\xi_{ni} f(x)| = o_p(1).$$

The proof of the above theorem uses a chaining argument similar to Dehling and Taqqu (1989) and appears in the next section. An analog of (2.6) when the errors are i.i.d. appears in Koul (1969, 1970) and was further generalized to include the case of strongly-mixing errors by Koul (1977, Proposition A1).

Using the technique similar to the proof of Theorem 2.1 one can obtain *pointwise* and *uniform convergence over compact* versions of the above theorem. This is stated below in a form that is useful in Chapter 4.

Theorem 2.2. *Suppose* $\epsilon_i := G(\eta_i)$, $i \ge 1$ *and* (1.2) *is satisfied.*

If (A.1), (A.2) *and* (A.3) *hold, and* F *is continuous at* $x \in I$, *then*

(2.8) $$|U_n(x) - U_n^*(x)| = o_p(1).$$

If, in addition, F *has a continuous density* f *at* $x \in I$ *and* (A.4) *holds, then*

(2.9) $$|\tau_n^{-1} \sum_i \gamma_{ni}\{I(\epsilon_i \le x + \xi_{ni}) - I(\epsilon_i \le x)\} - \tau_n^{-1} \sum_i \gamma_{ni}\xi_{ni} f(x)| = o_p(1).$$

In addition to (A.1) - (A.4), *suppose that the following hold:*

(A.7) f *is continuous and positive on* I

(A.8) J_m, J_m^+ *are continuously differentiable.*

Then for every $b \in (0, \infty)$,

$$(2.10) \quad \sup_{\{x \in I, |x| \le b\}} |\tau_n^{-1} \sum_i \gamma_{ni}\{I(\epsilon_i \le x + \xi_{ni}) - I(\epsilon_i \le x)\} - \tau_n^{-1} \sum_i \gamma_{ni}\xi_{ni} f(x)| = o_p(1).$$

Finally, as a byproduct of (2.4) and (2.6), we obtain a *weak convergence result* of *weighted residual empirical processes* based on LRD observations. Note that from (A.1) and (Π.2) of Section (1.3), the sequence of r.v.'s $\{ \tau_n^{-1} \sum_i \gamma_{ni} H_m(\eta_i) \}$ is bounded, uniformly in L^2 norm and hence has a subsequence converging to a r.v. $Z_m(\gamma)$, say. The determination of $Z_m(\gamma)$ for general weights $\{\gamma_{ni}, 1 \le i \le n\}$ with $m > 1$ is still an open problem. In the following corollary, the weak convergence is understood in $D[-\infty, \infty]$ equipped with the σ-field generated by the open balls of the *sup* metric.

Corollary 2.1. *Under* (A.1)-(A.6) *the process* $\{\tau_n^{-1} \sum_i \gamma_{ni}[I(\epsilon_i \le . + \xi_{ni}) - F(. + \xi_{ni})]\}$ *is tight and converges weakly to the process* $\{(m!)^{-1} J_m(.) Z_m(\gamma)\}$ *along a subsequence.*

//

Some remarks about the Hermite ranks and the assumptions of the above theorem are now in order.

Remark 2.1. Let η be a standard Gaussian r.v. and Q: $R^1 \to R^1$ be a measurable function such that $EQ^2(\eta) < \infty$. Recall from Sansone (1959) that the Hermite polynomials $\{H_k, k \ge 0\}$, defined by $\frac{d^k}{dx^k} \phi(x) = (-1)^k H_k(x)\phi(x)$ (alternatively, $e^{-(t^2-2tx)/2} = \sum_0^\infty t^k H_k(x)/k!$) have the property that $\{H_k/(k!)^{1/2}\}$ is an orthonormal basis for $L^2(R^1, B^1, d\Phi)$. The index of the *first nonzero coefficient* in the Fourier expansion of the r.v. $Q(\eta)$ with respect to this orthonormal basis is called its Hermite rank (see Taqqu, 1975). Clearly, if Q is an odd (even) function then its Hermite rank is 1 (2). Also *integration by parts* shows that if Q is monotone and right continuous such that the function $Q\phi$ vanishes at $-\infty$ and ∞ then the Hermite rank of Q is 1. //

Remark 2.2. Here we consider some examples of the Hermite ranks of a class of functions, see (2.2). If Q is strictly monotone and continuous, then the Hermite rank m in (2.2) is equal to 1. To see this, consider the case when Q is strictly increasing. Then using the fact that $\phi(x)\ H_r(x)\ dx = -\ d\{\phi(x)\ H_{r-1}(x)\}$, we obtain that for $y \in I$, $r \geq 1$, $EI(\epsilon \leq y)H_r(\eta) = EI(\eta \leq Q^{-1}(y))H_r(\eta) = -\ \phi(Q^{-1}(y))H_{r-1}(Q^{-1}(y))$, which is nonzero for $r = 1$. The same is true when Q is strictly decreasing.

Now let Q be an odd function with the additional property that $\{x \in R^1\colon Q(x) \leq 0\}$ equals either $[0, \infty)$ or $(-\infty, 0]$. Then also m is equal to one. To see this, consider the case $\{x \in R^1\colon Q(x) \leq 0\} = (-\infty, 0]$. Note that this implies that $Q(x) \geq 0$ for $x > 0$. Then for $y \leq 0$, $y \in I$, $EI(Q(\eta) \leq y)\eta = EI(Q(\eta) \leq y)\eta\ I(\eta \leq 0) < 0$, since the range of Q is I. Similarly, for $y \geq 0$, $y \in I$,

$$\begin{aligned} EI(Q(\eta) \leq y)\eta &= EI(Q(\eta) \leq 0)\eta I(\eta \leq 0) + EI(0 \leq Q(\eta) \leq y)\eta\ I(\eta>0) \\ &< -\phi(0) + \phi(0) = 0. \end{aligned}$$

The proof is similar in the case $\{x \in R^1\colon Q(x) \leq 0\} = [0, \infty)$.

An example of Q for which $m = 2$ in (2.2) and conditions (A.5) and (A.6) are satisfied is given by $Q(x) = |x|^{1/\delta}$, $\delta>1$. Dehling and Taqqu (1989) showed that for any $m \geq 1$, there is a Q for which the Hermite rank of the class of functions $\{I(Q(\eta) \leq x),\ x \in I\}$ is m. //

Remark 2.3. Note that J_m and J_m^+ are functions of bounded variation and hence are differentiable almost everywhere. Also, under (A.6), $\sup_{x \in I} |J_m'(x)|$ and $\sup_{x \in I} |J_m^{+\prime}(x)|$ are finite. //

If G is strictly monotone and continuous with d.f. F, then from Remark 2.2, m=1. The following proposition states that in this case (A.6) is satisfied if the *Fisher information* I(f) of the density f is finite.

Proposition 2.1. *Assume that* G *is strictly monotone, continuous and* (A.5) *holds. If, moreover* $G(\eta)$ *has an absolutely continuous density* f *and* $I(f) := \int [f'/f]^2 dF < \infty$, *then* (A.6) *holds.* //

Proof. Consider only the case when G is strictly increasing and continuous. Note that this entails $G = F^{-1}\Phi$. Hence from Remark 2.2, $J_1'(x) = \phi'(G^{-1}(x))/G'(G^{-1}(x)) = -f(x)\Phi^{-1}F(x)$. Note that for any $a \in (0, 1)$ there is $K_a \in (0, \infty)$ such that $|\Phi^{-1}(u)| \le K_a[u(1-u)]^{-a}$, $0 \le u \le 1$. Fix a $b \in I$ such that $F(b) > 0$. Then for $x > b$,

$$|f(x)\, \Phi^{-1}(F(x))| \le K_a\, f(x)\, [F(x)\, (1-F(x)]^{-a}$$

$$\le K_a\, [F(b)]^{-a} \int_x^d -f'(t)\, dt/[1-F(x)]^a$$

$$\le K_a\, [F(b)]^{-a} \int_x^d |-f'(t)/f(t)|/[1-F(t)]^{-a} dF(t)$$

$$\le K_a\, [F(b)]^{-a} \left\{\int_x^d | -f'(t)/f(t)|^2\, dF(t)\right\}^{1/2} \left\{\int_x^d [1-F(t)]^{-2a}\, dF(t)\right\}^{1/2}.$$

Upon choosing $2a < 1$ in this inequality, it follows that $|f(x)\Phi^{-1}(F(x))| \to 0$ as $x \to d$. Similarly one can prove that $|f(x)\Phi^{-1}(F(x))| \to 0$ as $x \to c$. //

1.3. Proofs.

In order to prove Theorem 2.1, we need some preliminary facts about Hermite expansions and ranks which are summarized below. These can be found in Taqqu (1975, 1979) and Dehling and Taqqu (1989). In what follows, L, with or without suffix, is a generic notation for slowly varying functions and c is a generic constant.

Let $\{\eta_i, i \geq 1\}$ and ρ be as in Section 1.1 and $p > 0$ be a fixed integer. Then the following facts hold:

(П.1) $\sum\sum|\rho^p(i-r)| = O(n^{2-p\theta}[L(n)]^p)$ if $p\theta < 1$,

$= O(nL_0(n))$ if $p\theta = 1$,

$= O(n)$ if $p\theta > 1$.

(П.2) For any measurable function $h \in L^2(R^1, B^1, d\Phi)$ with Hermite rank $p < 1/\theta$,

$$\text{Variance}(\sum_i h(\eta_i)) = O(\sum_i\sum_r|\rho^p(i-r)|) = O(n^{2-p\theta}[L(n)]^p).$$

(П.3) (i) Reciprocal and product of slowly varying functions are slowly varying.

(ii) For any slowly varying function V, $V(n)n^\delta \to \infty$ (0) for all $\delta > 0$ (< 0).

Now, we shall give the proof of (2.5). That of (2.4) is similar and simpler. To begin with we obtain an upper bound for the expected value of the square increment of the S_n process. Throughout the rest of this chapter, for a function $h : \mathbf{R}^1 \to \mathbf{R}^1$, $h(x, y)$ stands for $h(y) - h(x)$, $x \leq y$.

Lemma 3.1. *Suppose* $\epsilon_i := G(\eta_i)$, $i \geq 1$, *and* (1.2) *is satisfied. Then*

(3.1) $\tau_n{}^2E[S_n(x, y)]^2$

$$\leq \sum_i\sum_j|\gamma_{ni}\gamma_{nj}|\,[F(x+\xi_{ni}, y+\xi_{ni})\,F(x+\xi_{nj}, y+\xi_{nj})]^{1/2}\,|\rho(i-j)|^{m+1}.$$

Proof. The *Hermite expansion* of $\{I(\epsilon_i \leq x+\xi_{ni}) - F(x+\xi_{ni})\}$ and (2.3) yields that

$\tau_n^{\,2}E[\,S_n(x, y)]^2$

$$=\sum_i \sum_j \gamma_{ni}\gamma_{nj} \sum_{q \ge m+1} \sum_{r \ge m+1} J_q(x+\xi_{ni},\; y+\xi_{ni})/q!\; J_r(x+\xi_{nj},\; y+\xi_{nj})/r!$$

$$\times EH_q(\eta_j)H_r(\eta_k).$$

Using (2.1), the above is equal to

$$|\sum_i \sum_j \gamma_{ni}\gamma_{nj} \sum_{q \ge m+1} J_q(x+\xi_{ni},\; y+\xi_{ni})\; J_q(x+\xi_{nj}, y+\xi_{nj})/q!\; \rho^q(i\text{-}j)|$$

$$\le \sum_i \sum_j |\gamma_{ni}\gamma_{nj}| \sum_{q \ge m+1} |J_q(x+\xi_{ni},\; y+\xi_{ni})\; J_q(x+\xi_{nj}, y+\xi_{nj})/q!|\;\; |\rho^{m+1}(i\text{-}j)|$$

$$\le \sum_i \sum_j |\gamma_{ni}\gamma_{nj}|\; [\sum_{q \ge m} J_q^2(x+\xi_{ni},\; y+\xi_{ni})/q!]^{1/2}\, [\sum_{r \ge m} J_r^2(x+\xi_{nj},\; y+\xi_{nj})/r!]^{1/2}\; |\rho^{m+1}(i\text{-}j)|,$$

where the last step follows from the Cauchy-Schwarz inequality applied to the sum involving q. But the Hermite expansion of

$\{I(x+\xi_{ni} < \epsilon_i \le y+\xi_{ni}) - F(x+\xi_{ni},\; y+\xi_{ni})\}$ and (2.1) yield that

$$\sum_{q \ge m} J_q^2(x+\xi_{ni},\; y+\xi_{ni})/q! = E\,[I(x+\xi_{ni}<\epsilon_i \le y+\xi_{ni})\text{-}F(x+\xi_{ni},\; y+\xi_{ni})]^2$$

$$\le F(x+\xi_{ni},\; y+\xi_{ni}), \qquad \forall\; i \ge 1.$$

Hence the lemma is proved. //

Proof of (2.5). Without loss of generality, assume that $\gamma_{ni} \ge 0,\; 1 \le i \le n$. For the general $\{\gamma_{ni}\}$ the result follows from $\gamma_{ni} = \gamma_{ni}^+ - \gamma_{ni}^-$ and the triangle inequality. In what follows, we use a modification of the *chaining argument* of Dehling and Taqqu (1989). Let

(3.2) $$\lambda(x) := F(x) + J_m^+(x)/\,m!, \quad x \in I.$$

Note the following facts:

(3.3) λ is strictly increasing, $\lambda(d) < \infty$, and by (A.5), (A.6), λ is differentiable with uniformly continuous derivative $\lambda' = f + J_m^{+\prime}/m!$ satisfying $\lambda'(x) \to 0$ as $x \to c$ and d.

(3.4) $F(x, y) \leq \lambda(x, y), \quad x \leq y.$

(3.5) $|J_m(x, y)/m!| \leq J_m^{+}(x, y)/\, m! \leq \lambda(x, y), \quad x \leq y.$

Fix a $\delta > 0$ and an $n \geq 1$. Recall that $\tau_n := n^{(1-m\theta)/2} L^{m/2}(n)$ and let

(3.6) $$\kappa = \kappa(\delta, n) := \text{integer part of } \log_2\{\lambda(d)\sum_{i=1}^{n} \gamma_{ni}\, (\delta\tau_n)^{-1}\} + 1.$$

By (A.1), (A.2) and the inequality

$$\tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni} \geq n^{m\theta/2}\,[L(n)]^{-m/2}\sum_{i=1}^{n}\gamma_{ni}^2 \,/(n^{1/2} \max_{1 \leq i \leq n} \gamma_{ni}),$$

it follows that

(3.7) $$\tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni} \to \infty \qquad \text{and} \qquad \lambda(d)/2^{\kappa-1} \approx \delta/(\tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni}).$$

Next define **refining partitions of I** as follows. Note that λ is invertible and define

(3.8) $$\pi_{j,\,k} := \lambda^{-1}[\,\lambda(d)\, j\, 2^{-k}],\ j = 0, 1, ..., 2^k,\ k=0, 1, 2, ..., \kappa.$$

Clearly,

$$c = \pi_{0,\,k} < \pi_{1,\,k} < ... < \pi_{2^k-1,\,k} < \pi_{2^k,\,k} = d,$$

and

(3.9) $$\lambda(\pi_{j+1,\,k}) - \lambda(\pi_{j,\,k}) = \lambda(d)/2^k.$$

For an $x \in I$ and a $k \in \{0, 1, ..., \kappa\}$ define j_k^x by the relation

(3.10) $$\pi_{j_k^x,\,k} \leq x < \pi_{j_k^x+1,\,k}\ .$$

Now define a *chain linking each point* $x \in I$ to c by

(3.11) $$c = \pi_{j_0^x,0} \le \pi_{j_1^x,1} \le \dots\dots \le \pi_{j_\kappa^x,\kappa} \le x < \pi_{j_\kappa^x+1,\kappa}.$$

Then,

(3.12) $$S_n(x) = S_n(\pi_{j_0^x,0}, \pi_{j_1^x,1}) + S_n(\pi_{j_1^x,1}, \pi_{j_2^x,2}) + \dots\dots + S_n(\pi_{j_{\kappa-1}^x,\kappa-1}, \pi_{j_\kappa^x,\kappa}) + S_n(\pi_{j_\kappa^x,\kappa}, x).$$

First *consider the last term* in the *right hand side of* (3.12). By the triangle inequality, its absolute value is bounded by $\mathcal{A}_n(x) + \mathfrak{B}_n(x) + \mathfrak{C}_n(x)$, where

$$\mathcal{A}_n(x) := \tau_n^{-1}\sum_{i=1}^n \gamma_{ni}\, I(\pi_{j_\kappa^x,\kappa} + \xi_{ni} < \epsilon_i \le x + \xi_{ni}),$$

$$\mathfrak{B}_n(x) := \tau_n^{-1}\sum_{i=1}^n \gamma_{ni}\, F(\pi_{j_\kappa^x,\kappa} + \xi_{ni}, x + \xi_{ni}),$$

$$\mathfrak{C}_n(x) := \tau_n^{-1}\,\Big|\sum_{i=1}^n \gamma_{ni}\, J_m(\pi_{j_\kappa^x,\kappa} + \xi_{ni}, x + \xi_{ni})\, H_m(\eta_i)\Big|/m!, \qquad x \in I.$$

By the monotonicity of the indicator function, triangle inequality and (3.4), (3.5),

$$\mathcal{A}_n(x) \le \tau_n^{-1}\sum_{i=1}^n \gamma_{ni}\, I(\pi_{j_\kappa^x,\kappa} + \xi_{ni} < \epsilon_i \le \pi_{j_\kappa^x+1,\kappa} + \xi_{ni})$$

$$\le |S_n(\pi_{j_\kappa^x,\kappa}, \pi_{j_\kappa^x+1,\kappa})| + \tau_n^{-1}\sum_{i=1}^n \gamma_{ni}\, F(\pi_{j_\kappa^x,\kappa} + \xi_{ni}, \pi_{j_\kappa^x+1,\kappa} + \xi_{ni})$$
$$+ \tau_n^{-1}\Big|\sum_{i=1}^n \gamma_{ni}\, J_m(\pi_{j_\kappa^x,\kappa} + \xi_{ni}, \pi_{j_\kappa^x+1,\kappa} + \xi_{ni})\, H_m(\eta_i)\Big|/m!$$

$$\le |S_n(\pi_{j_\kappa^x,\kappa}, \pi_{j_\kappa^x+1,\kappa})| + \tau_n^{-1}\sum_{i=1}^n \gamma_{ni}(1+|H_m(\eta_i)|)\, \lambda(\pi_{j_\kappa^x,\kappa} + \xi_{ni}, \pi_{j_\kappa^x+1,\kappa} + \xi_{ni})$$

$$= |S_n(\pi_{j_\kappa^x,\kappa}, \pi_{j_\kappa^x+1,\kappa})| + b_n(x), \qquad \text{say.}$$

Similarly, $\mathfrak{B}_n(x) + \mathfrak{C}_n(x) \le b_n(x)$ and hence

(3.13) $$|S_n(\pi_{j_\kappa^x,\kappa}, x)| \le |S_n(\pi_{j_\kappa^x,\kappa}, \pi_{j_\kappa^x+1,\kappa})| + 2\, b_n(x).$$

We now show that

$$\sup_{x \in I} b_n(x) = o_p(1). \tag{3.14}$$

To that effect, by the *mean value theorem*,

$$b_n(x) = \tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni}\,(1 + |H_m(\eta_i)|\,)\,[\lambda\,(\pi_{j_\kappa^x,\,\kappa}\,,\,\pi_{j_\kappa^x+1\,,\,\kappa}) + \xi_{ni}\{\lambda'(u_{nix}) - \lambda'(v_{nix})\}],$$

where, u_{nix} (v_{nix}) is a number in $[\pi_{j_\kappa^x+1,\,\kappa}\,,\,\pi_{j_\kappa^x+1,\,\kappa} + \xi_{ni}]$ ($[\pi_{j_\kappa^x,\,\kappa}\,,\,\pi_{j_\kappa^x,\,\kappa} + \xi_{ni}]$). Therefore by (3.9),

$$|b_n(x)| \le \tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni}\,(1 + |H_m(\eta_i)|\,)\,\lambda\,(d)\,2^{-\kappa}$$

$$+ \sup_{i,\,x}\,|\lambda'(u_{nix}) - \lambda'(v_{nix})|\,\tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni}\,(1 + |H_m(\eta_i)|\,)\,|\xi_{ni}|.$$

But by (A.4), the stationarity and the Gaussianity of $\{\eta_i\}$,

$$\tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni}\,(1 + |H_m(\eta_i)|\,)\,|\xi_{ni}|^j = O_p(\,[\tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni}]^{1-j}), \quad j = 0, 1. \tag{3.15}$$

Hence by (3.7)

$$b_n(x) \le \{\delta + \sup_{i,\,x}\,|\lambda'(u_{nix}) - \lambda'(v_{nix})|\,\}\,O_p(1). \tag{3.16}$$

Now to show that $\sup_{i,\,x}\,|\lambda'(u_{nix}) - \lambda'(v_{nix})| = o(1)$, note that,

$$\max_{1 \le i \le n}\,|\,u_{nix} - v_{nix}\,| \le \pi_{j_\kappa^x+1\,,\,\kappa} - \pi_{j_\kappa^x\,,\,\kappa} + 2\,\max_{1 \le i \le n}\,|\xi_{ni}|.$$

The uniform continuity of λ^{-1} on compacts and the fact

$$\lambda(\pi_{j_\kappa^x+1\,,\,\kappa}) - \lambda(\pi_{j_\kappa^x\,,\,\kappa}) = \lambda(d)\,2^{-\kappa} \le \delta/(\tau_n^{-1}\sum_{i=1}^{n}\gamma_{ni}) \to 0$$

imply that for every $b \in (0, \infty)$, $\sup\{\pi_{j_\kappa^x+1\,,\,\kappa} - \pi_{j_\kappa^x\,,\,\kappa}\;;\;|x| \le b,\, x \in I\} \to 0$. Hence from (3.3) it follows that

$$\sup \{ |\lambda'(u_{nix}) - \lambda'(v_{nix})|;\ 1 \le i \le n,\ x \in I\} = o(1).$$

This together with (3.16) and the arbitrariness of δ readily yields (3.14).

Now from (3.12) and (3.13) we obtain

$$(3.17)\quad \sup_{x \in I} |S_n(x)| \le \sup_x |S_n(\pi_{j_0^x,\,0}\,,\ \pi_{j_1^x,\,1})| + \sup_x |S_n(\pi_{j_1^x,\,1}\,,\ \pi_{j_2^x,\,2})| + \ldots\ldots$$

$$+ \sup_x |S_n(\pi_{j_{\kappa-1}^x,\,\kappa-1}\,,\ \pi_{j_\kappa^x,\,\kappa})| + \sup_x S_n(\pi_{j_\kappa^x,\,\kappa}\,,\ \pi_{j_\kappa^x+1,\,\kappa})$$

$$+ 2 \sup_{x \in I} b_n(x).$$

Fix a $k \in \{0, 1, \ldots, \kappa\text{-}1\}$. Observe that

$$\sup_{x \in I} |S_n(\pi_{j_k^x,\,k}\,,\ \pi_{j_{k+1}^x,\,k+1})|$$

$$= \max_{0 \le i \le 2^{k+1}-1}\ \sup_{x \in [\pi_{i,\,k+1},\, \pi_{i+1,\,k+1})} |S_n(\pi_{j_k^x,\,k}\,,\ \pi_{j_{k+1}^x,\,k+1})|.$$

If $\pi_{i,\,k+1} = \pi_{r,\,k}$ for some $r \in \{0, 1, 2, \ldots, 2^k\}$ then for $x \in [\pi_{i,\,k+1}\,,\ \pi_{i+1,\,k+1})$, $\pi_{j_k^x,\,k} = \pi_{i,\,k+1} = \pi_{j_{k+1}^x,\,k+1}$, so that the above supremum for the i[th] interval is zero. If $\pi_{i,\,k+1} \neq \pi_{r,\,k}$ for all $r \in \{1, 2, \ \ldots, 2^k\}$ then i is odd. Therefore for $x \in [\pi_{i,\,k+1}\,,\ \pi_{i+1,\,k+1})$, $\pi_{j_k^x,\,k} = \pi_{i-1,\,k+1}$ and $\pi_{j_{k+1}^x,\,k+1} = \pi_{i,\,k+1}$ so that

$$\sup_{x \in [\pi_{i,\,k+1},\, \pi_{i+1,\,k+1})} |S_n(\pi_{j_k^x,\,k}\,,\ \pi_{j_{k+1}^x,\,k+1})| = |S_n(\pi_{i-1,\,k+1}\,,\ \pi_{i,\,k+1})|.$$

Hence by Chebychev's inequality and Lemma 3.1, $\forall\ \delta>0$,

$$P[\sup_{x \in I} | S_n(\pi_{j_k^x, k}, \pi_{j_k^x+1, k}) | > \delta/2]$$

$$\le \sum_{j=0}^{2^{k+1}-1} P[| S_n(\pi_{j, k+1}, \pi_{j+1, k+1}) | > \delta/2]$$

$$\le 4\, \delta^{-2}\, \tau_n^{-2} \sum_i \sum_r \gamma_{ni}\gamma_{nr} |\rho(i-r)|^{m+1} \times$$

$$\sum_{j=0}^{2^{k+1}-1} [F(\pi_{j, k+1} + \xi_{ni}, \pi_{j+1, k+1} + \xi_{ni})\ F(\pi_{j, k+1} + \xi_{nr}, \pi_{j+1, k+1} + \xi_{nr})]^{1/2}$$

$$\le 4\, \delta^{-2}\, \tau_n^{-2} \sum_i \sum_r \gamma_{ni}\gamma_{nr} |\rho(i-r)|^{m+1},$$

where the last inequality follows from the Cauchy-Schwarz inequality applied to the factor involving F and the fact that

$$\sum_{j=0}^{2^{k+1}-1} [F(\pi_{j, k+1} + a, \pi_{j+1, k+1} + a) \le 1, \quad \forall\, a \in R^1.$$

Similarly, we also obtain

$$P[\sup_{x \in I} | S_n(\pi_{j_\kappa^x, \kappa}, \pi_{j_\kappa^x+1, \kappa}) | > \delta/2] \le 4\, \delta^{-2}\, \tau_n^{-2} \sum_i \sum_r \gamma_{ni}\gamma_{nr} |\rho(i-r)|^{m+1}.$$

Hence from (3.17) we obtain that

(3.18) $$P[\sup_{x \in I} |S_n(x)| > \delta]$$

$$\le 4\, (\kappa + 1)\, \delta^{-2}\, \tau_n^{-2} \sum_i \sum_r \gamma_{ni}\gamma_{nr} |\rho(i-r)|^{m+1} + P[\sup_{x \in I} b_n(x) > \delta/4].$$

We now analyze the first term in this upper-bound. Since for large n, $\tau_n > 1$, and by the Cauchy-Schwarz inequality, $\sum_i \gamma_{ni} \le n^{1/2}$, (3.6) yields that

$$\kappa + 1 < 2 + \log_2 [\lambda(d)/\delta] + c\, \ln n,$$

where c is a constant and ln denotes natural logarithm. We shall next show that,

(3.19) $$a_n := (ln\ n)\ \tau_n^{-2} \sum_i \sum_r \gamma_{ni}\gamma_{nr}\ |\rho(i-r)|^{m+1} = o(1), \ \ \theta < 1/m.$$

Consider first the case when $(m + 1)\theta < 1$.

$$a_n \leq (\ ln\ n)\ (n \max_{1 \leq i \leq n} \gamma_{ni}^2\ )\ n^{-1}\ \tau_n^{-2} \sum_i \sum_r |\rho(i-r)|^{m+1}$$

$$\approx L(n)\ [n^{2-(m+1)\theta}] / [n^{2-m\theta}] .$$

Here, the last asymptotic equivalence follows from (A.1), (Π.3)(i) and (Π.1). Assertion (3.19) for this case, now follows from (Π.3)(ii). Assertion (3.19) for other cases follows from (Π.1) in a similar fashion.

The proof of (2.5) now follows from (3.14) (3.18) and (3.19). //

Remark 3.1. The main difference between the above chain and the Dehling-Taqqu chain is in the definition of κ in (3.6). Note that when $\gamma_{ni} = n^{-1/2}$, the above chain reduces to that of the Dehling-Taqqu. Also, (2.5) generalizes some of the results of Dehling and Taqqu (1989) from the ordinary empiricals to the weighted empiricals. //

Proof of (2.6). Define,

$$\nu_n(x) := [U_n(x) - S_n(x)] - [\ U_n^*(x) - S_n^*(x)]$$

$$= \tau_n^{-1} \sum_i \gamma_{ni}[J_m(x+\xi_{ni}) - J_m(x)]\ (m!)^{-1} H_m(\eta_i), \qquad x \in \mathbf{I}.$$

Since by (2.5), $\sup_{x \in \mathbf{I}} |S_n(x)| = o_p(1)$, and also $\sup_{x \in \mathbf{I}} |S_n^*(x)| = o_p(1)$ when (2.5) is applied with $\xi_{ni} = 0\ \forall\ i,\ 1 \leq i \leq n$, it is enough to show that

(3.20) $$\sup_{x \in \mathbf{I}} |\nu_n(x)| = o_p(1).$$

Note that

$$\nu_n(x) := \tau_n^{-1} \sum_i \gamma_{ni}\xi_{ni} J'_m(u_{nix})\ H_m(\eta_i)/m!,$$

where u_{nix} is some number between $x+\xi_{ni}$ and x, $1 \le i \le n$.
Let $w_n := \max_{1\le i\le n}\{\,|\xi_{ni}|\}$. Then for any $k > 0$,

$$\sup\{\,|\nu_n(x)| : |x| > k,\ x \in I\}$$

$$\le \sup\{\,|J'_m(x)| : |x| > k - w_n,\ x \in I\}\ \tau_n^{-1} \sum_i |\gamma_{ni}\xi_{ni}|\,|H_m(\eta_i)|/m!.$$

By (A.6), $\sup\{\,|J'_m(x)| : |x| > k,\ x \in I\} \to 0$ as $k \to \infty$. Hence, in view of (3.15) with $j = 1$, to prove (3.20) it suffices to show that $\forall\ 0 < k < \infty$,

$$\sup\{\,|\nu_n(x)| : |x| \le k,\ x \in I\} = o_p(1). \tag{3.21}$$

Fix a k. Viewing ν_n as a process on $[-k, k]\cap I$, it is enough to show that

(3.22) $\qquad$ (a) $\nu_n(x) = o_p(1)\ \forall\ x \in I$,

and

(b) $E[\nu_n(x) - \nu_n(y)]^2 \le c\,(x - y)^2$, $\quad$ for some $c \in (0, \infty)$

For (a), note that

$$E[\nu_n(x)]^2 = \tau_n^{-2}\sum_i\sum_j \gamma_{ni}\gamma_{nj}\ J_m(x, x+\xi_{ni})\ J_m(x, x+\xi_{nj})\ \rho^m(i-j)/\ m!$$

$$\le \max_{1\le i\le n} [J_m(x, x+\xi_{ni})]^2\ (n \max_{1\le i\le n} \gamma_{ni}^2)\ n^{-1}\ \tau_n^{-2}\sum_i\sum_r |\rho(i-r)|^m \to 0,$$

since J_m is continuous in a neighborhood of x and by (A.1), (Π.1) the rest of the term is bounded. For (b), note that the *mean value theorem* entails that for $x \le y$, $x, y \in [-k, k]\cap I$,

$$\nu_n(y) - \nu_n(x) = \tau_n^{-1}\sum_i \gamma_{ni}\{[J_m(y+\xi_{ni}) - J_m(x+\xi_{ni})] - [J_m(y) - J_m(x)]\}\ (m!)^{-1} H_m(\eta_i)$$

$$= \tau_n^{-1} \sum_i \gamma_{ni} \, (y-x)[J_m'(u_{nixy}) - J_m'(u_{xy})]H_m(\eta_i)(m!)^{-1},$$

where $u_{nixy} \in [x+\xi_{ni}, y+\xi_{ni}]$, $u_{xy} \in [x, y]$. Therefore,

$$E[\nu_n(x) - \nu_n(y)]^2 = (y-x)^2 \tau_n^{-2} \sum_i \sum_j \gamma_{ni}\gamma_{nj} \, \Delta_{nixy} \, \Delta_{njxy} \, \rho^m(i-j)/\,(m!)^2 = C_n(y-x)^2,$$

where, $\Delta_{nixy} := J_m'(u_{nixy}) - J_m'(u_{xy})$. Note that, $|\Delta_{nixy}| \le 2 \sup_{x \in I} |J_m^{+\prime}(x)| < \infty \; \forall$ n, i, x, y. Hence using (A.1) and (F.1) once more as in (a) we obtain that $C_n = O(1)$. Now (3.22) follows from Billingsley (1968; Theorem 12.3). This completes the proof of (2.6) also. //

Proof of (2.7) . This follows from (2.6) in conjunction with the following:

$$\sup_{x \in I} |\tau_n^{-1} \sum_i \gamma_{ni}\{F(x + \xi_{ni})\} - F(x)\} - \tau_n^{-1} \sum_i \gamma_{ni}\xi_{ni} f(x)| = o(1),$$

which can be proved by applying the *mean value theorem* on $\{F(x+\xi_{ni})\}-F(x)\}$ and using the uniform continuity of f and (A.4). //

Proof of Theorem 2.2. Proof of (2.8) follows directly from the Hermite expansion at each fixed x and the continuity of J_m, which in turn follows from the assumption of the continuity of the d.f. F. Proof of (2.9) follows from (2.8) in a simple manner. Finally, assertion (2.10) follows along the same line as in the proof of (2.5) and (2.6) with suitable modifications. //

CHAPTER 2

RANK ESTIMATION

2.1 Introduction.

The idea of estimating location parameter based on rank statistics finds its root in the seminal work of Hodges and Lehmann (1963). Since then, a major branch of nonparametric statistics deals with the rank-estimation (R-estimation) of parameters by *minimizing certain dispersions* based on the ranks of observation. Generally, these dispersions are expressed in terms of linear rank statistics. Thus, the widespread applicability of *linear rank statistics* for a variety of testing problems leads to its use in estimation in a natural way. The key tool for studying the R-estimators is the AUL of linear rank statistics.

To explain R-estimation in a regression set up, let ψ be a nondecreasing real-valued function on (0, 1) and $\{R_{i\Delta},\ 1 \le i \le n,\ \Delta \in R^p\}$ denote the *residual ranks*, i.e., $R_{i\Delta}$ is the rank of $Y_{ni} - x_{ni}^t\Delta$ among $Y_{nj} - x_{nj}^t\Delta$, $1 \le j \le n$. Define

(1.1)
$$S(\Delta) := \sum_i (x_{ni} - \bar{x})\, \psi(R_{i\Delta}/(n+1)) = [\, S_1(\Delta), \dots S_p(\Delta)]^t;$$
$$\mathfrak{J}(\Delta) := \sum_i \psi(R_{i\Delta}/(n+1)\, (Y_{ni} - x_{ni}^t\Delta).$$

Note that S is a linear rank statistic. Jureckova (1971) defined a class of linear rank estimators of β, the regression parameter, by

(1.2)
$$\hat{\beta}_{JU} := \operatorname{argmin}_{\Delta \in R^p}\{ \sum_{j=1}^{n} |S_j(\Delta)| \}.$$

For square integrable score function ψ, Jureckova (1971) obtained the AUL of the linear rank statistics $S(\Delta)$ and the asymptotic normality of the standardized $\hat{\beta}_{JU}$ by exploiting the *contiguity* of a sequence of densities of the i.i.d. errors. Koul (1969, 1970) obtained the AUL of S for right continuous and bounded ψ under weaker assumptions on the design matrix X and the error density. He used *weak convergence* techniques and the inherent monotonicity of the w.e.p. to come up with a better result for the bounded score functions. Koul (1992, Theorem 3.2.4) also obtained the AUL of rank statistics under more general heteroscedastic errors that allow the study of the robustness of linear rank estimators under independent errors.

Jaeckel (1972) defined another class of R-estimators by

$$(1.3) \qquad \hat{\beta}_J := \mathrm{argmin}_{\Delta \in R^p}\, \mathcal{J}(\Delta),$$

where

$$\mathcal{J}(\Delta) := \sum_i \psi(R_{i\Delta} /(n+1)\, (Y_{ni} - x^t_{ni}\Delta).$$

Jaeckel showed that if the score function ψ satisfies

$$(L.1) \qquad \sum_{i=1}^{n} \psi(i/(n+1)) = 0,$$

then ψ is nonnegative and convex on R^p. Moreover, using the AUL results of Jureckova and the observation that almost everywhere differential of $\mathcal{J}$ is -S, Jaeckel proved the asymptotic equivalence of $\hat{\beta}_{JU}$ and $\hat{\beta}_J$.

Our aim in this chapter is to investigate the asymptotic behavior of Jureckova-Jaeckel type estimators under the model (1.1.1) and (1.1.2). We consider nondecreasing, right continuous bounded score function ψ and follow the weak convergence approach of Koul. The question of the AUL of rank statistics is first reduced to that of the asymptotic uniform continuity of the w.e.p. of the residuals. The results of Chapter 1 are then applied to yield the asymptotic representations of the above estimators in the LRD set up.

2.2. AUL of linear rank statistics.

To proceed further, we introduce more notation. Let $\bar{X}$ denote the $n \times p$ matrix whose jth column consists of $n \times 1$ entries of $\bar{x}_{nj} := \sum x_{nij}/n$, $1 \le j \le p$ and $W := X - \bar{X}$, the centered design matrix. For any matrix D of order $n \times p$, let

$$(2.1) \qquad A_d := \tau_n^{-1}(D^tD)^{1/2}, \quad B_d := \tau_n(D^tD)^{1/2},$$

as long as the definitions make sense. Define the following process based on residual ranks and weights corresponding to centered design as

$$\mathcal{Z}(u, \Delta) := \sum_i (x_{ni} - \bar{x})\, I(R_{i\Delta} \le nu), \; 0 \le u \le 1, \; \Delta \in R^p.$$

We are now ready to state the AUL theorem of the linear rank statistics S. Recall conditions (A.5) and (A.6) from Theorem 1.2.1.

Theorem 2.1. *In addition to* (1.1.1), (1.1.2), (A.5) *and* (A.6) *assume that the following hold:*

(W.1) $(W^tW)^{-1}$ exists for all $n \ge p$.

(W.2) $n \max_{1 \le i \le n} x_{ni}^t(W^tW)^{-1}x_{ni} = O(1)$.

(L.2) $\psi \in \Psi := \{g\colon [0, 1] \to R^1$, g *is nondecreasing, right continuous,* $g(1) - g(0) = 1\}$.

Then $\forall\, b \in (0, \infty)$,

$$(2.2) \qquad \sup_{0 \le u \le 1,\, s \in N(b)} \left\| B_w^{-1}[\mathcal{Z}(u, \beta + A_w^{-1} s) - \hat{\mathcal{Z}}(u)] - s\, f(F^{-1}(u)) \right\| = o_p(1),$$

$$(2.3) \qquad \sup_{\psi \in \Psi,\, s \in N(b)} \left\| B_w^{-1}[S(\beta + A_w^{-1} s) - \hat{S}] + s \int f\, d\psi(F) \right\| = o_p(1),$$

where,

(2.4) $$\hat{\mathcal{Z}}(u) := \sum_i (x_{ni} - \bar{x})\, [I(\epsilon_i \le F^{-1}(u) - u], \quad 0 \le u \le 1,$$

$$\hat{S} := \sum_i (x_{ni} - \bar{x})\, [\psi(F(\epsilon_i)) - \bar{\psi}], \qquad \bar{\psi} := \int_0^1 \psi(u)du. \qquad //$$

The following lemma gives an approximation to $B_w^{-1}\hat{S}$ in terms of the Hermite polynomials that is useful in determining the limiting distributions of the above estimators.

Lemma 2.1. *Let* $\{\eta_i, i \ge 1\}$ *be a stationary, mean zero, unit variance Gaussian process with correlation* $\rho(k) := E(\eta_1 \eta_{1+k})$, $k \ge 1$. *Suppose* $\epsilon_i := G(\eta_i)$, $i \ge 1$ *and assume that* (1.1.2), (W.1), (W.2) *and* (L.2) *hold. Then,*

(2.5) $$B_w^{-1}\hat{S} = S_w J_m(\psi) + o_p(1),$$

where,

$$J_q(\psi) := E\, \psi(F(\epsilon))\, H_q(\eta) = -\int_0^1 J_q(F^{-1}(u))\, d\psi(u),\ q \ge 1,$$

and for any $n \times p$ *matrix* D,

(2.6) $$S_d := B_d^{-1} \sum_i d_{ni} H_m(\eta_i)/m!. \qquad //$$

The next corollary gives the asymptotic representations of the suitably normalized R-estimators defined in (1.2) and (1.3). These representations are obtained from (2.3) in a routine fashion as in the proof of Theorem 4.1 of Jureckova (1971) and Theorem 3.3 of Jaeckel (1972). We omit the details for the sake of brevity.

Corollary 2.1. *Under the assumptions of Theorem* 2.1 *and condition* (L.1),

(2.7) $$A_w\, (\hat{\beta}_J - \beta) = A_w\, (\hat{\beta}_{JU} - \beta) + o_p(1)$$

$$= \{\textstyle\int f\, d\psi(F)\}^{-1}\, B_w^{-1}\hat{S} + o_p(1),$$

(2.8) $$= \{\textstyle\int f\, d\psi(F)\}^{-1}\, S_w J_m(\psi) + o_p(1). \qquad //$$

Remark 2.1. In the i.i.d. errors case Koul (1992, Corollary 4.4.1) derived the asymptotic representation of $\hat{\beta}_J$, under the conditions (W.1), (W*.2), (A.5), (L.1) - (L.2) where (W*.2) is a slightly weaker condition than (W.2), namely,

(W*.2) $$\max_{1 \le i \le n} w_{ni}^t (W^t W)^{-1} w_{ni} = o(1).$$

Using the AUL of S, Koul obtained that

$$(W^t W)^{1/2}(\hat{\beta}_J - \beta) = \{\int f \, d\psi(F)\}^{-1} (W^t W)^{-1/2} \hat{S} + o_p(1),$$

which converges in distribution to $N(0, \{\int f \, d\psi(F)\}^{-2} \int_0^1 [\psi(u) - \bar{\psi}]^2 \, du \; I_{p \times p})$.

Note that the limiting representations in the LRD case differ from those of the i.i.d. errors case in two fundamental ways. Firstly, they have different normalizations. If, for example, $\lim_n W^t W/n$ exists and is positive definite then $(W^t W)^{1/2}$ is of the order of $n^{1/2}$ whereas A_w is of the order of $n^{m\theta/2}/L^{m/2}(n)$. Secondly, unlike the i.i.d. errors case, the limiting distribution may not be always normal. The value of m is very crucial for determining the limiting distribution.

If, either G is strictly monotone and continuous or G is an odd function with the property that $\{x \in R: G(x) \le 0\}$ equals either $(-\infty, 0]$ or $(0, \infty]$ then m=1 (See Remark 1.2.2). In such cases the first approximation of $A_w(\hat{\beta}_J - \beta)$ is exactly $N(0, \sigma_J^2 \Gamma_n)$ where $\Gamma_n := \tau_n^{-2}(W^t W)^{-1/2} W^t R W (W^t W)^{-1/2}$, R the dispersion matrix of $(\eta_1, .. \eta_n)^t$ and $\sigma_J^2 := \{\int f \, d\psi(F)\}^{-2} \; [E\psi(F(\epsilon))\eta]^2$. Yajima's (1991) results can be used to calculate the limit of Γ_n under some additional conditions on the design. //

Remark 2.2. Conditions (W.1) and (W.2) are satisfied by many designs, in particular, by polynomial designs with $x_{nij} = i^j$, $1 \le i \le n$, $1 \le j \le p$ and by trigonometric designs with $x_{nij} = \cos(i\mu_j)$ or $\sin(i\mu_j)$, $\mu_j \neq \mu_k$ for $j \neq k$. //

2.3. Proofs.

To begin with we introduce some more processes that will be useful in the sequel. Accordingly, define

$$F_n(y, \Delta) := n^{-1}\sum_i I(Y_{ni} - x_{ni}^t\Delta \le y), \quad y \in I,$$

$$T_n(u, \Delta) := \sum_i (x_{ni} - \bar{x}) I(Y_{ni} - x_{ni}^t\Delta \le F^{-1}(u)) \quad 0 \le u \le 1, \ \Delta \in R^p.$$

The basic idea of the proof of (2.2) can be sketched as follows. Note that

$$(3.1) \qquad \mathcal{Z}(u, \beta + A_w^{-1} s) = \sum_i (x_{ni} - \bar{x}) I[\text{Rank of}(\epsilon_i - x_{ni}^t A_w^{-1} s) \le nu]$$

$$= \sum_i (x_{ni} - \bar{x}) I[F_{ns}(\epsilon_i - x_{ni}^t A_w^{-1} s) \le u],$$

where $F_{ns}(.)$ is the empirical distribution function of $\{\epsilon_i - x_{ni}^t A_w^{-1} s, \ 1 \le i \le n\}$. Therefore $B_w^{-1}\mathcal{Z}(u, \beta + A_w^{-1} s)$ can be approximated by the process $B_w^{-1}\dot{\mathcal{Z}}(u, s)$ (defined in (3.5) below) in the sense of (3.6). Now use (3.4) below along with the AUL of the T_n process to conclude the AUL of the $\mathcal{Z}$ process. To that effect, the following preparatory lemma states the AUL of the F_n and T_n processes which are interesting for their own sake.

Lemma 3.1. *Assume that* (1.1.1), (1.1.2), (W.1), (W.2), (A.5) and (A.6) *hold. Then for every* $b \in (0, \infty)$,

$$(3.2) \ \sup_{y \in I, s \in N(b)}\{|\sigma_n^{-1}[F_n(y, \beta + A_w^{-1} s) - F_n(y, \beta)] - n^{1/2}\, \bar{x}^t D_w^{-1} s\, f(y)|\} = o_p(1),$$

$$(3.3) \qquad \|B_w^{-1}[T_n(u, \beta + A_w^{-1} s) - T_n(u, \beta)] - s\, f(F^{-1}(u))\| = \bar{o}_p(1),$$

and

$$(3.4) \qquad FF_{ns}^{-1}(u) - u = \bar{o}_p(1),$$

where

$\sigma_n := n^{-1/2}\tau_n \rightarrow 0$ *and* $\bar{o}_p(1)$ *denotes a sequence of stochastic processes converging to zero uniformly in* $0 \le u \le 1$, $s \in N(b)$, *in probability.* //

Proof. We first consider (3.3). Observe that for fixed $s \in R^p$,

$$B_w^{-1}T_n(u, \beta + A_w^{-1} s) = \tau_n^{-1}\sum_i (W^tW)^{-1/2}(x_{ni} - \bar{x})I[\epsilon_i \le F^{-1}(u) + x_{ni}^t A_w^{-1} s]$$

is a vector of V_n processes with $\gamma_{ni} :=$ j^{th} coordinate of $(W^tW)^{-1/2}(x_{ni} - \bar{x})$ and $\xi_{ni} = x_{ni}^t A_w^{-1} s$, $1 \le i \le n$. Notice that (W.1) entails (A.1) and this choice of $\{\gamma_{ni}\}$ satisfies (A.2). Note also that

$$\max_{1 \le i \le n} | x_{ni}^t A_w^{-1} s| \le \tau_n \max_{1 \le i \le n} \left\| x_{ni}^t (W^tW)^{-1/2} \right\| \, \| s \|$$

$$= n^{1/2} \max_{1 \le i \le n} \left\| x_{ni}^t (W^tW)^{-1/2} \right\| O(n^{-m\theta/2}L^{m/2}(n)) = o(1),$$

so that (A.3) holds $\forall s \in N(b)$. Finally, to verify (A.4), observe that, by the Cauchy-Schwarz inequality applied to the second step,

$$\sum_i \left\| (W^tW)^{-1/2}(x_{ni} - \bar{x}) x_{ni}^t (W^tW)^{-1/2} s \right\|$$

$$\le \max_{1 \le i \le n} | x_{ni}^t (W^tW)^{-1/2} s| \sum_i \left\| (W^tW)^{-1/2}(x_{ni} - \bar{x}) \right\|$$

$$\le \max_{1 \le i \le n} \left\| x_{ni}^t (W^tW)^{-1/2} \right\| \, \| s \| \, n^{1/2} \, [\sum_i \left\| (W^tW)^{-1/2}(x_{ni} - \bar{x}) \right\|^2]^{1/2}$$

$$= \max_{1 \le i \le n} \left\| x_{ni}^t (W^tW)^{-1/2} \right\| \, \| s \| \, n^{1/2} \, p^{1/2} = O(1).$$

Therefore, by (1.2.7) and the fact that $B_w^{-1}\sum_i (x_{ni} - \bar{x}) x_{ni}^t \; A_w^{-1} = I_{p \times p}$, it follows that for all $s \in R^p$,

$$\sup_{u \in [0,\, 1]} \left\| B_w^{-1}[T_n(u, \beta + A_w^{-1} s) - T_n(y, \beta)] - s\, f(F^{-1}(u)) \right\| = o_p(1).$$

The uniform convergence over $s \in \mathcal{N}(b)$ is achieved by exploiting the monotonicity of the indicator function and the d.f. F along with the compactness of $\mathcal{N}(b)$ as in Theorem 2.1 of Koul (1991). This completes the proof of (3.3). Assertion (3.2) can be proved similarly by taking $\gamma_{ni} = n^{-1/2}$, $\forall\ 1 \le i \le n$ and ξ_{ni} as before.

For (3.4), note that $|FF_{ns}^{-1}(u) - u| \le |FF_{ns}^{-1}(u) - F_{ns}F_{ns}^{-1}(u)| + |F_{ns}F_{ns}^{-1}(u) - u|$. The first term is $\bar{o}_p(1)$ from (3.2) and the fact that $F_{ns}(y) = F_n(y, \beta + A_w^{-1} s)$ $\forall y \in I$. The second term is at most $1/n$. Hence (3.4) follows. //

***Proof of Theorem* 2.1.** We first prove (2.2). Define

$$\tilde{\mathfrak{Z}}(u, s) := \sum_i (x_{ni} - \bar{x}) I[\epsilon_i \le F_{ns}^{-1}(u) + x_{ni}^t A_w^{-1} s], \qquad 0 \le u \le 1,\ s \in R^p. \tag{3.5}$$

From representation (3.1) and by the fact that

$$P[\epsilon_i - x_{ni}^t A_w^{-1} s = \epsilon_j - x_{nj}^t A_w^{-1} s\ ,\ \text{for some } i, j;\ 1 \le i \ne j \le n] = 0, \tag{3.6}$$

we obtain that w.p.1

$$\left\| B_w^{-1}[\mathfrak{Z}(u, \beta + A_w^{-1} s) - \tilde{\mathfrak{Z}}(u, s)] \right\| \le \tau_n^{-1} \max_{1 \le i \le n} \left\| x_{ni}^t (W^t W)^{-1/2} \right\| = o(1),$$
$$\forall\ 0 \le u \le 1,\ s \in R^p.$$

Therefore, it suffices to prove that

$$B_w^{-1}[\tilde{\mathfrak{Z}}(u, s) - \hat{\mathfrak{Z}}(u)] - s\, f(F^{-1}(u)) = \bar{o}_p(1). \tag{3.7}$$

Now by (3.3), $B_w^{-1}\tilde{\mathfrak{Z}}(u, s)$

$$= B_w^{-1} T_n(FF_{ns}^{-1}(u), \beta + A_w^{-1} s) = B_w^{-1} T_n(FF_{ns}^{-1}(u), \beta) + s\, fF^{-1}FF_{ns}^{-1}(u) + \bar{o}_p(1).$$

Therefore using $B_w^{-1}\hat{\mathfrak{Z}}(u) = B_w^{-1} T_n(u, \beta)$, the uniform continuity of the function $f o F^{-1}$ and (3.4) it is clear that (3.7) will follow from the *tightness* of $\{B_w^{-1} T_n(u, \beta),\ u \in [0, 1]\}$. Therefore, it remains to prove that $\forall\ \alpha > 0,\ \exists\ \delta > 0 \ni$

$$\lim\sup_n P[\sup_{|u-v|<\delta} \left\| B_w^{-1}[T_n(u, \beta) - T_n(v, \beta)] \right\| > \alpha\ ] < \alpha. \tag{3.8}$$

But (1.2.4) applied p times, with the choice of $\{\gamma_{ni}\}$ and $\{\xi_{ni}\}$ as in the proof of (3.3), yields that

$$\text{(3.9)} \qquad \sup_{u \in [0,\, 1]} \left\| B_w^{-1} T_n(u, \beta) - J_m(F^{-1}(u)) S_w \right\| = o_p(1),$$

where $E\,\|S_w\|^2 = O(1)$.

Note that for $u \le v \in [0, 1]$, the Cauchy-Schwarz inequality yields that

$$\text{(3.10)} \quad |J_m(F^{-1}(v)) - J_m(F^{-1}(u))| \le E^{1/2}[I(u<F(\epsilon) \le v)]^2\, E^{1/2} H_m^2(\eta) \le \{m!(v-u)\}^{1/2}.$$

Hence (3.8) follows from (3.9) and (3.10). This proves (2.2) also. (2.3) follows from (2.2) using an argument similar to the proof of Koul (1992, Theorem 3.2.2). //

***Proof of* (2.5).** Let $\mathfrak{Z}(u) := \sum_i (x_{ni} - \bar{x})\, [I(\epsilon_i < F^{-1}(u) - u]$, $0 \le u \le 1$. Note that because of the continuity of F,

$$\text{(3.11)} \qquad P[\, \hat{\mathfrak{Z}}(u) = \mathfrak{Z}(u), \text{ for some } u \in (0, 1)\,] = 0$$

Since $\hat{S} = -\int_0^1 \mathfrak{Z}(u)\, d\psi(u)$, (2.5) follows from (3.11) and (3.9) and the boundedness of the function ψ. //

CHAPTER 3

MINIMUM DISTANCE ESTIMATION

2.1 Introduction.

In minimum distance (m.d.) method of estimation, one estimates the unknown parameter by a minimizer of some discrepancy measure between a function of observations and that of a family of underlying distributions. Wolfowitz (1957) discussed m. d. estimation as a general principle of estimation that can be applied in many statistical problems and showed that under *identification and continuity assumptions*, m.d. estimators (m.d.e.) are strongly consistent. Research on m. d. estimation proliferated during the mid seventies and eighties. Beran (1978) discussed asymptotic normality and robustness of the one sample location estimators obtained by minimum Hellinger distance. Parr and Schucanny (1980) and Millar (1981) discussed asymptotic normality and robustness of a large class of m.d.e. that includes Cramer-Von Mises type distances in the one sample location model. Boos (1981) discussed minimum Cramer-Von Mises type distance estimation in the one and two sample models and its application to the 'goodness of fit' tests. The above authors, in some way or another, viewed m.d.e. as a functional defined on an appropriate subset of univariate distribution functions, that satisfies Frechet differentiability condition and proved the asymptotic normality of these estimators by techniques delineated in Serfling (1980, Chapter 6).

Koul and Dewet (1983) and Koul (1985a, 1985b) visioned appropriate extensions of m.d.e. from the one and two sample location models to the multiple linear regression model via weighted residual empirical processes. Their technique of proving the asymptotic normality and qualitative robustness of the m.d.e. is completely different from that of the predecessor's work. It uses the ***asymptotic uniform quadraticity*** of the Cramer-Von Mises type statistics based on weighted empirical processes.

While most of the asymptotic literature on m.d.e. assume independent observations, little seems to be known for the dependent observations. Motivated by the important application of linear models with **LRD** errors, we study in this chapter the large sample behavior of m.d.e. of the regression parameters under the model (1.1.1). In Section 2 we derive the limiting representations of the normalized m.d.e.'s. The following empirical processes with weights $\{\mathbf{w}_{ni} := \mathbf{x}_{ni} - \bar{\mathbf{x}}, 1 \le i \le n\}$ and $\{\mathbf{x}_{ni}, 1 \le i \le n\}$ are useful in this chapter.

$$(1.1)\quad \mathbf{V}_{\mathbf{w}}(\Delta, y) := \sum_i \mathbf{w}_{ni} I(Y_{ni} \le \mathbf{x}_{ni}^t \Delta + y) = \sum_i \mathbf{w}_{ni} I(\epsilon_i \le \mathbf{x}_{ni}^t(\Delta - \beta) + y)$$

$$\mu_{\mathbf{w}}(\Delta, y) := \sum_i \mathbf{w}_{ni} F(\mathbf{x}_{ni}^t(\Delta - \beta) + y)$$

$$\mathbf{U}_{\mathbf{w}}(\Delta, y) := \sum_i \mathbf{w}_{ni}\{I(\epsilon_i \le \mathbf{x}_{ni}^t(\Delta - \beta) + y) - F(\mathbf{x}_{ni}^t(\Delta - \beta) + y)\},\ \Delta \in \mathbf{R}^p,\ \ y \in \mathbf{R}^1.$$

Similarly, define $\mathbf{V}_{\mathbf{x}}$, $\mu_{\mathbf{x}}$ and $\mathbf{U}_{\mathbf{x}}$ with weights $\{\mathbf{x}_{ni}\}$.

For linear regression model, Williamson-Koul minimum distance estimator of the regression parameter β is defined by

$$(1.2)\qquad \hat{\beta}_K := \operatorname{argmin}_{\Delta \in \mathbf{R}^p} M_w(\Delta),$$

where

$$M_w(\Delta) := \tau_n^{-2} \int \left\| (\mathbf{W}^t\mathbf{W})^{-1/2} \mathbf{V}_w(\Delta, y) \right\|^2 dy.$$

Under the assumption that the error d.f. F is symmetric around zero Koul (1985b) defined, for each H, a minimum distance estimator of β by

(1.3) $$\beta_K^+ := \operatorname{argmin}_{\Delta \in R^p} M_x(\Delta),$$

where

$$M_x(\Delta) := \tau_n^{-2} \int \left\| (X^tX)^{-1/2} \{ \sum_i x_{ni} \{ I(Y_{ni} - x_{ni}^t \Delta \le y) - I(-Y_{ni} + x_{ni}^t \Delta < y) \} \right\|^2 dH(y),$$

and H is a nondecreasing right continuous function from R^1 to R^1.

Note that $\hat{\beta}_K$ and β_K^+ are the estimators $\hat{\beta}$ and β^+ respectively defined in Koul (1985b) for the independent errors case. The motivation for considering these estimators and its finite sample properties are discussed in Koul (1985b). In particular, as noted by Fine (1966), for $p = 1$ with $x_{ni} = 1$, $1 \le i \le n_1$, and $x_{ni} = 0$, $n_1+1 \le i \le n$, $\hat{\beta}_K$ reduces to the two sample Hodges-Lehmann estimator of the location parameter. Also, for $x_{ni} = 1$, $H(x) = x$ [$H(x) = I(0 \le x)$], β_K^+ is the Hodges-Lehmann estimator of the one sample location parameter [median estimator] (see Koul 1985b).

3.2. Asymptotic representation of minimum distance estimators.

In this section we first derive the asymptotic uniform quadraticity of $M_w(\Delta)$. To that effect, apart from (W.1) and (W.2) in Section (2.2.2), we need the following assumptions on the model (1.1.1) and (1.1.2):

(W.3) For every $s \in R^p$, $\int \left\| B_w^{-1} \mu_w(\beta + A_w^{-1} s) - s f(y) \right\|^2 dy = o(1)$,

(D.1) The d.f. F of ϵ has a continuous density f.

(D.2) $\int f^2 \, dy < \infty$.

(D.3) $\int F(1-F) \, dy < \infty$.

Define,

$$Q_w(\Delta) := \int \left\| B_w^{-1} V_w(\beta, y) + B_w^{-1} W^t X (\Delta - \beta) f(y) \right\|^2 dy, \qquad \Delta \in R^p.$$

Theorem. 2.1. *In addition to* (1.1.1), (1.1.2) *assume that* (W.1) - (W.3), (D.1) - (D.3) *hold. Then, for all* $b \in (0, \infty)$,

$$(2.1) \qquad E \sup_{\mathbf{s} \in \mathcal{N}(b)} |M_w(\beta + A_w^{-1} s) - Q_w(\beta + A_w^{-1} s)| = o(1). \qquad //$$

Proof. The technique of the proof is similar to that of Theorem 2.1 of Koul (1985a). As there, it is enough to show that $\forall\, b \in (0, \infty)$,

$$(2.2) \qquad E \sup_{\mathbf{s} \in \mathcal{N}(b)} \int \left\| B_w^{-1} U_w(\beta, y) + \mathbf{s}\, f(y) \right\|^2 dy = O(1),$$

$$(2.3) \qquad \sup_{\mathbf{s} \in \mathcal{N}(b)} \int \left\| B_w^{-1} \mu_w(\beta + A_w^{-1} s) - \mathbf{s}\, f(y) \right\|^2 dy = o(1),$$

$$(2.4) \qquad E \sup_{\mathbf{s} \in \mathcal{N}(b)} \int \left\| B_w^{-1} U_w(\beta + A_w^{-1} s, y) - B_w^{-1} U_w(\beta, y) \right\|^2 dy = o(1).$$

***Proof of* (2.2).** Using $\|\mathbf{a}+\mathbf{b}\|^2 \le 2\,(\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2)$ for $\mathbf{a}, \mathbf{b} \in R^p$ and Fubini,

$$\text{l.h.s. of (2.2)} \le 2 \int E \left\| B_w^{-1} U_w(\beta, y) \right\|^2 dy + 2\, b^2 \int f^2\, dy.$$

Now, $\displaystyle \int E \left\| B_w^{-1} U_w(\beta, y) \right\|^2 dy = \tau_n^{-2} \sum_{j=1}^{p} \int E\Big[\sum_i \alpha_{wnij} \{I(\epsilon_i \le y) - F(y)\}\Big]^2 dy,$

where, for any matrix $\mathbf{D}$, α_{dnij} := the jth entry of the vector $(\mathbf{D}^t\mathbf{D})^{-1/2}\, \mathbf{d}_{ni}$.

Using the Hermite expansion of $\{I(\epsilon_i \le y) - F(y),\ y \in \mathbf{I}\}$ in (1.2.3) and the properties stated in (1.2.1) the integrand of the jth summand is equal to

$$= \tau_n^{-2} \sum_i \sum_r \alpha_{wnij}\, \alpha_{wnrj} \sum_{q \ge m} \sum_{q' \ge m} J_q(y)/q!\ J_{q'}(y)/q'!\ EH_q(\eta_i)\, H_{q'}(\eta_r)$$

$$= \tau_n^{-2} \sum_i \sum_r \alpha_{wnij}\, \alpha_{wnrj} \sum_{q \ge m} [J_q^2(y)/q!\,]\rho^q(i-r).$$

$$\le \sum_{q \ge m} J_q^2(y)/q!\quad n \max_{1 \le i \le n} \alpha^2_{wnij}\quad \sum\sum |\rho^m(i-r)| / n\tau_n^2$$

$$(2.5) \quad \leq F(y)(1-F(y))\, n \max_{1 \leq i \leq n} w_{ni}^t (W^t W)^{-1} w_{ni}\; n^{-(2-m\theta)} L^{-m}(n) \sum\sum |\rho^m(i-r)|$$

$$(2.6) \quad \leq K\, F(y)(1 - F(y)).$$

Here, (2.5) is obtained from $\sum_{q \geq m} J_q^2(y)/q! = E[I(\epsilon_i \leq y)-F(y)]^2 = F(y)(1-F(y))$ while (2.6) from (F.1) and (W.2). Now (2.2) follows from the assumptions (D.2) and (D.3). The assertion (2.3) follows from (W.3) and a monotonicity argument of Koul (1985a, Theorem 2.1). //

***Proof of* (2.4).** For fixed $s \in R^p$,

$$E\int \left\| B_w^{-1} U_w(\beta + A_w^{-1} s,\ y) - B_w^{-1} U_w(\beta, y) \right\|^2 dy$$

$$\leq 2\int E \left\| B_w^{-1} S_w(\beta + A_w^{-1} s,\ y) - B_w^{-1} S_w(\beta,\ y) \right\|^2 dy$$

$$+ 2\int E \left\| B_w^{-1} \sum_i w_{ni}\{J_m(x_{ni}^t A_w^{-1} s + y) - J_m(y)\}(m!)^{-1} H_m(\eta_i) \right\|^2 dy$$

$$= 2\,[T_1 + T_2],$$

where, for $\Delta \in R^p$, $y \in R^1$,

$$S_w(\Delta, y) := \sum_i w_{ni}\{I(\epsilon_i \leq x_{ni}^t(\Delta-\beta)+y) - F(x_{ni}^t(\Delta-\beta)+y) - J_m(x_{ni}^t(\Delta-\beta)+y)(m!)^{-1} H_m(\eta_i)\}$$

From (1.2.3), T_1 is equal to

$$\sum_{j=1}^{p} \int \tau_n^{-2}\, E[\sum_i \alpha_{wnij} \sum_{q \geq m+1}^{\infty} J_q(y,\ x_{ni}^t A_w^{-1} s + y)(q!)^{-1}\ H_q(\eta_i)]^2\, dy,$$

where for a function $h : R^1 \to R^1$, $h(x, y)$ stands for $h(y) - h(x)$, $x \leq y$.

Now using the properties in (1.2.1) as in the proof of (2.2) the jth integrand of the above expression

$$\leq \tau_n^{-2} \sum_i \sum_r |\alpha_{wnij}\, \alpha_{wnrj}|\, \left|\sum_{q \geq m+1} J_q(y, x_{ni}^t A_w^{-1} s+y) J_q(y, x_{nr}^t A_w^{-1} s+y)/q!\right| |\rho^m(i-r)|$$

$$\le \tau_n^{-2} \sum_i \sum_r |\alpha_{wnij}\, \alpha_{wnrj}| \, [\sum_{q \ge m+1} J_q^2(y,\ x_{ni}^t A_w^{-1} s + y)/q!]^{1/2}$$

$$[\sum_{q \ge m+1} J_q^2(y,\ x_{nr}^t A_w^{-1} s + y)/q!]^{1/2} |\rho^m(i-r)|$$

$$\le \max_{1 \le i \le n} F(y,\ x_{ni}^t A_w^{-1} s + y)\, n \max_{1 \le i \le n} w_{ni}^t (W^t W)^{-1} w_{ni}\, n^{-(2-m\theta)} L^{-m}(n) \sum_i \sum_r |\rho^m(i-r)|.$$

$$\le K \max_{1 \le i \le n} F(y,\ x_{ni}^t A_w^{-1} s + y)$$

Using (1.2.1) once more, it is easy to see that the integrand in T_2 is bounded by $K \max_{1 \le i \le n} J_m^2(y,\ x_{ni}^t A_w^{-1} s + y)$ which by the Cauchy-Schwarz inequality is bounded by $K \max_{1 \le i \le n} F(y,\ x_{ni}^t A_w^{-1} s + y)$.

Note that $\max_{1 \le i \le n} | x_{ni}^t A_w^{-1} s| \le \tau_n \max_{1 \le i \le n} \left\| x_{ni}^t (W^t W)^{-1/2} \right\| \, \|s\|$

$$= n^{1/2} \max_{1 \le i \le n} \left\| x_{ni}^t (W^t W)^{-1/2} \right\| O(n^{-m\theta/2} L^{m/2}(n)) = o(1), \tag{2.7}$$

by(W.2) and the fact that for any slowly varying function V and $\forall \delta > 0$, $V(n)/n^\delta = o(1)$. Therefore, with $\alpha_n := \max_{1 \le i \le n} | x_{ni}^t A_w^{-1} s|$, Fubini's theorem yields that,

$$\int \max_{1 \le i \le n} |F(x_{ni}^t A_w^{-1} s + y) - F(y)|\, dy \le \int [\int_{y-\alpha_n}^{y+\alpha_n} f(t)\, dt]\, dy = 2\,\alpha_n = o(1).$$

Again the uniform convergence in (2.4) is achieved in a routine fashion, see, e.g., Koul (1985a). This completes the proof Theorem 2.1. //

The following corollary gives the asymptotic representations of the suitably normalized minimizer of $M_w(\Delta)$.

Corollary 2.1. *Under the assumptions of Theorem* 2.1,

$$A_w(\hat\beta_K - \beta) = - \left(\int f^2\, dx\right)^{-1} B_w^{-1} \int U_w(\beta,\ y)\, dF(y) + o_p(1), \tag{2.8}$$

$$= - \left(\int f^2\, dx\right)^{-1} S_w \int J_m(y)\, dF(y) + o_p(1), \tag{2.9}$$

$$= \left(\int f^2\, dx\right)^{-1} S_w E F(\epsilon) H_m(\eta) + o_p(1).$$

//

Proof. Proof of (2.8) is routine from the asymptotic uniform quadraticity result (2.1) and hence omitted. For (2.9) note that

$$\left\| \int [B_w^{-1} U_w(\beta, y) - S_w J_m(y)] \, dF(y) \right\|^2$$
$$\leq \sup_{y \in I} \left\| B_w^{-1} U_w(\beta, y) - S_w J_m(y) \right\|^2 = o_p(1),$$

by (1.2.4) when applied p times, j th time with $\gamma_{ni} = \alpha_{wnij}$, $1 \leq j \leq p$. //

Remark 2.1. In the independent errors case Koul (1985b) derived the asymptotic representation of $\hat{\beta}_K$, which when specialized to i.i.d. errors case states that under the conditions (W.1), (W*.2) (W.3) and (D.1)-(D.3)

$$(W^tW)^{1/2}(\hat{\beta}_K - \beta) = -(\int f^2 \, dx)^{-1}(W^tW)^{-1/2} \int U_w(\beta, y) \, dF(y) + o_p(1),$$

where (W*.2) is as in Remark 2.2.1. Therefore, remarks similar to that of Remark 2.2.1 is also applicable here. Also note that condition (W.3) is satisfied in the presence of other conditions when for example, the density f is uniformly continuous.

To cite an example of error distribution under which the limiting distribution of $\hat{\beta}_K$ is nonnormal, consider $G(x) = x^2 - 1$. In this case $m = 2$ (see Dehling and Taqqu, 1989). Hence from the representation (2.9) it is clear that in the case of the two sample (n_1, n_2) location model for the centered chisquare random variables with design vector, $x_{ni} = 1$, $1 \leq i \leq n_1$ and $x_{ni} = 0$, $n_1+1 \leq i \leq n := n_1+n_2$, $n^{\theta} L^{-1}(n)(\hat{\beta}_K - \beta)$, the normalized two sample Hodges-Lehmann estimator of the location parameter converges in distribution to the random variable $[\lambda(1-\lambda)]^{-1}\{Z_2(\lambda) - \lambda Z_2(1)\}$, where $\lambda := \lim(n_1/n)$ is assumed to exist, $0<\lambda<1$ and $\{Z_2(\lambda) : \lambda \in [0,1]\}$ is the Rosenblatt process (see Dehling and Taqqu, 1989). Clearly, the limiting distribution is nonnormal. //

Now we turn to the asymptotic representations of the m.d.e. β_K^+ under the symmetric error distribution. For the sake of clarity we assume that the integrating measure H is symmetric around zero. The techniques of proofs are similar to those of Theorem 2.1 and Corollary 2.1 and are omitted.

To derive the asymptotic uniform quadraticity of the dispersion $M_X(\mathbf{b})$ and the limiting representation of β_K^+ we assume the following conditions on the design matrix and the underlying distribution:

(X.1) $(\mathbf{X}^t\mathbf{X})^{-1}$ exists for all $n \geq p$.

(X.2) $n \max_{1 \leq i \leq n} \mathbf{x}_{ni}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}_{ni} = O(1)$.

(X.3) For every $s \in R^p$, $\int \left\| B_X^{-1} \left[J_X(\beta + A_X^{-1}s, \; y) - J_X(\beta, \; y) \right] - \mathbf{b}\, f(y) \right\|^2 dH(y) = o(1)$.

(D*.1) The d.f. F of ϵ has a continuous symmetric density f.

(D*.2) (i) $0 < \int_0^\infty f \, dH < \infty$, (ii) $0 < \int_0^\infty f^2 \, dH < \infty$.

(D*.3) $0 < \int_0^\infty (1-F) \, dH < \infty$

(D*.4) $\lim_{s \to 0} \int f(y+s) \, dH(y) = \int f(y) \, dH(y)$.

Theorem 2.2. *In addition to* (1.1.1) *and* (1.1.2) *assume that* (X.1)-(X.3) *and* (D*.1)-(D*.4) *hold. Then*

$$(2.10) \quad A_X(\beta_K^+ - \beta) = -\left(2\int f^2 \, dH\right)^{-1} B_X^{-1} \int \{U_X(\beta, y) + U_X(\beta, -y)\} f(y) \, dH(y) + o_p(1).$$

$$= -\left(2\int f^2 \, dH\right)^{-1} S_X \int [J_m(y) + J_m(-y)] f(y) \, dH(y) + o_p(1).$$

$$= \left(2\int f^2 \, dH\right)^{-1} S_X \, E\rho(\epsilon) H_m(\eta) + o_p(1),$$

where,

$$\rho(x) \ := \rho_0(x) - \rho_0(-x), \ \ \rho_0(x) := \int_{-\infty}^{x} f \, dH. \qquad //$$

Remark 2.1 From representation (2.10) it follows that in the case of one sample location model for symmetry, the normalized Hodges-Lehmann estimator $n^{m\theta/2} \, L^{-m/2}(n)(\beta_K^+ - \beta)$ converges in distribution to the random variable $(\int f^2 \, dx)^{-1} \ EF(\epsilon)H_m(\eta) \ Z_m(1)$, where $\{Z_m(t) \, : \, t \in [0, \, 1]\}$ is the m^{th} Hermite process as defined in **DT**. //

CHAPTER 4

REGRESSION QUANTILES AND RELATED PROCESSES

4.1. Introduction

In their fundamental paper, Koenker and Basset (1978) (**KB**) introduced regression quantiles as a natural extension of the notion of sample quantiles and order statistics from the one sample location model to the multiple linear regression model. Most of the asymptotic literature on these assume either independent errors or weakly dependent errors as in Portnoy (1991). In this chapter, we study the large sample behavior of the regression quantiles processes under the dependent setup (1.1.1). We also investigates the asymptotic behavior of the regression rank-score processes, L- and linear regression rank-score statistics of Jureckova (1992a) and Gutenbrunner and Jureckova (1992) under the LRD setup. In particular we obtain the AUL of the regression rank-score statistics under the LRD errors.

For an $\alpha \in (0,\ 1)$, KB defined an α^{th} regression quantile as any member $\hat{\beta}_n(\alpha)$ of the set

$$(1.1) \qquad \hat{\mathbf{B}}_n(\alpha) := \{\mathbf{b} \in \mathbf{R}^p\colon \sum_{i=1}^{n} h_\alpha(Y_{ni} - \mathbf{x}_{ni}^t \mathbf{b}) = \text{minimum}\},$$

where $h_\alpha(u) := \alpha\ uI(u>0) - (1-\alpha)\ uI(u \le 0)$, $u \in \mathbf{R}^1$, $0 \le u \le 1$. Note that, $h_{1/2}(u) = |\ u\ |\ /2$ and so, $\hat{\beta}_n(1/2)$ reduces to the well-known ***least absolute deviation*** (LAD) estimator of β.

Theorem 3.1 of **KB** gives the following linear programming version of the above minimization problem (1.1):

$$\text{minimize } \alpha\, \mathbf{1}_n^t\, \mathbf{r}^+ + (1 - \alpha)\, \mathbf{1}_n^t\, \mathbf{r}^- \tag{1.2}$$

$$\text{subject to } \mathbf{Y}_n - \mathbf{X}\mathbf{b} = \mathbf{r}^+ - \mathbf{r}^- , \quad (\mathbf{b}, \mathbf{r}^+, \mathbf{r}^-) \in \mathrm{R}^p \times \mathrm{R}_+^n \times \mathrm{R}_+^n ,$$

where $\mathbf{1}_n^t := [1,.....1]_{1 \times n}$ and Y_n is the response vector. By the linear programming theory, $\hat{\mathbf{B}}_n(\alpha)$ is the convex hull of one or more basic solutions of the form

$$\mathbf{b}_\mathbf{h} = \mathbf{X}_\mathbf{h}^{-1}\mathbf{Y}_\mathbf{h} \tag{1.3}$$

where h is a subset of $\{1,2...n\}$ of size p and $\mathbf{X}_\mathbf{h}$ ($\mathbf{Y}_\mathbf{h}$) denotes the sub-design matrix (sub design vector) with rows $\mathbf{x}_{ni}^t$, $i \in \mathbf{h}$ (co-ordinates Y_{ni}, $i \in \mathbf{h}$).

In general, one can choose $\hat{\beta}_n(.)$ from (1.1) in such a way that it is a stochastic process called *regression quantile process* that has sample path in $[D(0, 1)]^p$. There will also be 'break-points' $0 = \alpha_0 < \alpha_1 < .... < \alpha_{J_n} = 1$, such that $\hat{\beta}_n(.)$ is constant over each interval (α_i, α_{i+1}), $0 \le i \le J_n-1$. See Gutenbrunner and Jureckova (1992) (GJ) and references therein for more on this.

The corresponding dual program, mentioned in the appendix of **KB** is the following:

$$\text{maximize } \mathrm{Y}_n^t\mathbf{a} \text{ with respect to } \mathbf{a} \tag{1.4}$$

$$\text{subject to } \mathbf{X}_n^t\mathbf{a} = (1\text{-}\alpha)\, \mathbf{X}_n^t\mathbf{1}_n, \quad \mathbf{a} \in [0, 1]^n. \tag{1.5}$$

GJ investigated the statistical properties of the optimal solution of (1.4) when the errors are independent. Note that a maximizer of (1.4) also maximizes $\mathbf{a}^t(\mathbf{Y}_n - \mathbf{X}\mathbf{t})$ subject to (1.5), for any $\mathbf{t} \in \mathrm{R}^p$. Now choose $\mathbf{t} = \hat{\beta}_n(\alpha)$, where for some p-dimensional subset $\mathbf{h}_n(\alpha)$ of $\{1,....n\}$ an optimal solution for (1.2) is given by $\hat{\beta}_n(\alpha) = \mathbf{X}_{\mathbf{h}_n(\alpha)}^{-1}\, \mathbf{Y}_{\mathbf{h}_n(\alpha)}$. Then one particular solution of (1.4) corresponding to this choice of $\hat{\beta}_n(\alpha)$ can be given as follows:

For i not in $\mathbf{h}_n(\alpha)$, let

$$\hat{a}_{ni}(\alpha) = 1, \quad Y_{ni} - \mathbf{x}_{ni}^t \hat{\beta}_n(\alpha) > 0, \tag{1.6}$$
$$= 0, \quad Y_{ni} - \mathbf{x}_{ni}^t \hat{\beta}_n(\alpha) < 0,$$

and for i in $\mathbf{h}_n(\alpha)$, $\hat{a}_{ni}(\alpha)$ is the solution of the p linear equations:

$$\sum_{j \in \mathbf{h}_n(\alpha)} \mathbf{x}_{nj} \hat{a}_{nj}(\alpha) = (1 - \alpha) \sum_{j=1}^{n} \mathbf{x}_{nj} - \sum_{j=1}^{n} \mathbf{x}_{nj} \, I(Y_{nj} - \mathbf{x}_{nj}^t \, \hat{\beta}_n(\alpha) > 0). \tag{1.7}$$

GJ call $\hat{\mathbf{a}}_n(\alpha)$ the regression rank-scores for each $\alpha \in (0, 1)$. When p =1 and $x_{ni} = 1$, $1 \le i \le n$, **GJ** observe that

$$\hat{a}_{ni}(\alpha) = 1, \qquad \alpha < (R_{ni} - 1)/n \tag{1.8}$$
$$= R_{ni} - n\alpha, \quad (R_{ni} - 1)/n \le \alpha \le R_{ni}/n$$
$$= 0, \qquad R_{ni}/n < \alpha,$$

where $R_{ni} := \text{Rank}(Y_{ni})$. Hence in the one sample location model, the processes $\{\hat{a}_{ni}(.),\ 1 \le i \le n\}$ reduces to the familiar rank process (See Hajek and Sidak, 1967, Section V.3.5). **GJ** observed that $\hat{\mathbf{a}}_n(.)$ satisfying (1.6) and (1.7) can be chosen in such a way that it has piecewise linear paths in $[C(0, 1)]^n$ and $\hat{\mathbf{a}}_n(0) = \mathbf{1}_n = \mathbf{1}_n - \hat{\mathbf{a}}_n(1)$. See **GJ** also for a discussion on the generalization of the duality of order statistics and rank process from the one sample location model to the linear regression model by the **RQ** (:= regression quantiles) and **RR** (:= regression rank-scores) processes.

Using these processes one can construct various statistics such as L-statistics and **RR** statistics that are useful to make inference about the regression parameter $\boldsymbol{\beta}$. In the context of i.i.d. errors, different types of L-estimators were proposed by Ruppert and Carrol (1980), Koenker and Portnoy (1987), Portnoy and Koenker (1989) and GJ. Among them, Koenker and Portnoy (1987) and GJ considered smoothed L-estimators that are asymptotically equivalent to the well known M-estimators but have the added advantage of being invariant with respect to scale and reparametrization of the design. In this paper we shall consider these type of L-estimators and will observe that under appropriate conditions, the above asymptotic equivalence continues to hold even in the LRD setup.

As pointed out by Jureckova (1992a), one of the major advantages of using RR statistics based on the residual is that the corresponding estimators of the some of the components of β, when others are treated as nuisance parameters do not require the estimation of the nuisance parameters. The basic result needed to study these estimators is the AUL of the RR statistics based on residuals as given in Jureckova (1992a) for the i.i.d. errors. The corresponding result under (1.1.1) and (1.1.2) is given in section 4. Section 2 obtains the joint asymptotic distribution of the finite number of suitably normalized RQ 's and asymptotic representations of RQ and RR processes. Section 3 applies this results to yield the asymptotic behavior of L- and RR statistics. The proofs heavily depend upon the *uniform closeness result* of Chapter 1.

4.2. Theorems and Proofs.

To find out the asymptotic distribution of RQ, we first define the following minimum-distance type estimator of β by

$$\hat{\beta}_{md}(\alpha) := \operatorname{argmin}_{\Delta} \left\| (X^t X)^{-1/2} \mathcal{T}(\Delta, \alpha) \right\|^2, \tag{2.1}$$

where,

$$\mathcal{T}(\Delta, \alpha) := \sum_i \mathbf{x}_{ni}\{I(Y_{ni} - \mathbf{x}_{ni}^t\Delta \le 0) - \alpha\}, \qquad 0 \le \alpha \le 1,\ \Delta \in \mathbb{R}^p.$$

By the continuity of the d.f. F, $\mathcal{T}(\Delta, \alpha)$ is an almost everywhere differential of the function $\sum_i h_\alpha(Y_{ni} - \mathbf{x}_{ni}^t\Delta)$ with respect to Δ. Therefore, intuitively, minimizer of $\sum_i h_\alpha(Y_{ni} - \mathbf{x}_{ni}^t\Delta)$ and $\hat{\beta}_{md}(\alpha)$ should be asymptotically equivalent. The following lemma is enunciated towards this direction. It also gives the joint asymptotic representation of the finite number of **RQ**. To state it, we need to introduce

$$\beta(\alpha) := \beta + F^{-1}(\alpha)\mathbf{c}_1,\ \mathbf{c}_1 := (1, 0, \ldots, 0)^t, \text{ and } q(\alpha) := f(F^{-1}(\alpha)),\ 0 \le \alpha \le 1.$$

Also, recall the definition of $\mathbf{S_X}$ from (3.2.8), conditions (X.1), (X.2) from Theorem 3.2.2 and conditions (A.7), (A.8) from Theorem 1.2.2. Moreover, the following design condition is assumed:

(X.0) **The first column of the design matrix X consists of one only.**

Lemma 2.1. *Assume that* (1.1.1) (1.1.2) *and* (X.0) - (X.2) *hold. Then*

(i) $$\sup_{\alpha \in [0,1]} \left\| \mathbf{B}_X^{-1}\mathcal{T}(\beta(\alpha), \alpha) - J_m(F^{-1}(\alpha))\mathbf{S_X} \right\| = o_p(1),$$

(ii) *If, in addition,* (A.7) *holds then* $\forall\ b \in (0, \infty)$ *and* $\forall\ \alpha \in (0, 1)$,

$$\sup_{\mathbf{s} \in N(b)} \left\| \mathbf{B}_X^{-1}[\mathcal{T}(\beta(\alpha) + \mathbf{A}_X^{-1}\mathbf{s}, \alpha) - \mathcal{T}(\beta(\alpha), \alpha)] - \mathbf{s}\, q(\alpha) \right\| = o_p(1),$$

(iii) *If, in addition,* (A.7) *and* (A.8) *hold then* $\forall\ a \in (0, 1/2]$ *and* $\forall\ b \in (0, \infty)$,

$$\sup_{\alpha \in [a, 1-a]\ \mathbf{s} \in N(b)} \left\| \mathbf{B}_X^{-1}[\mathcal{T}(\beta(\alpha) + \mathbf{A}_X^{-1}\mathbf{s}, \alpha) - \mathcal{T}(\beta(\alpha), \alpha)] - \mathbf{s}\, q(\alpha) \right\| = o_p(1).$$

(iv) *Under the conditions of* (ii), *for every* $\alpha \in (0, 1)$,

(a) $\mathbf{A_x}(\hat{\boldsymbol{\beta}}_{md}(\alpha) - \boldsymbol{\beta}(\alpha)) = - q^{-1}(\alpha)\, \mathbf{B}_x^{-1}\mathfrak{T}(\boldsymbol{\beta}(\alpha), \alpha) + o_p(1)$

$$= - q^{-1}(\alpha)\, \mathbf{S_x}\, \mathrm{J}_m(\mathrm{F}^{-1}(\alpha)) + o_p(1).$$

(b) $\mathbf{A_x}(\hat{\boldsymbol{\beta}}_n(\alpha) - \hat{\boldsymbol{\beta}}_{md}(\alpha)) = o_p(1).$

(c) *Consequently, for any* $0 < \alpha_1 < \alpha_2, \ldots, < \alpha_k < 1$,

$$[\mathbf{A_x}(\hat{\boldsymbol{\beta}}_n(\alpha_1) - \boldsymbol{\beta}(\alpha_1)), \ldots\ldots, \mathbf{A_x}(\boldsymbol{\beta}_n(\alpha_k) - \boldsymbol{\beta}(\alpha_k))] \tag{2.2}$$

$$= - [\, q^{-1}(\alpha_1)\, \mathrm{J}_m(\mathrm{F}^{-1}(\alpha_1)), \ldots\ldots, q^{-1}(\alpha_k)\, \mathrm{J}_m(\mathrm{F}^{-1}(\alpha_k))] \oplus \mathbf{S_x} + o_p(1),$$

where, for any two matrices **A** and **B**, $\mathbf{A} \oplus \mathbf{B}$ *stands for its Kronecker product.* //

Proof. (i) follows from (1.2.4) when applied p times, j^{th} time to $\gamma_{ni} = j^{th}$ coordinate of $\mathbf{D}_x^{-1}\mathbf{x}_{ni}$, $1 \le i \le n$, $1 \le j \le p$.

(ii) Note that

$$\mathbf{B}_x^{-1}[\mathfrak{T}(\boldsymbol{\beta}(\alpha) + \mathbf{A}_x^{-1}\mathbf{s}, \alpha) - \mathfrak{T}(\boldsymbol{\beta}(\alpha), \alpha)] - \mathbf{s}\, q(\alpha)$$

$$= \tau_n^{-1} \sum_i \mathbf{D}_x^{-1}\mathbf{x}_{ni}\{\mathrm{I}(\epsilon_i \le \mathrm{F}^{-1}(\alpha) + \mathbf{x}_{ni}^t\mathbf{A}_w^{-1}\mathbf{s}) - \mathrm{I}(\epsilon_i \le \mathrm{F}^{-1}(\alpha))\} - \mathbf{s}\, q(\alpha)$$

Now (ii) and (iii) follow from (1.2.9) and (1.2.10) of Theorem 1.2.2 respectively, as in the proof of (2.3.3).

(iv)(a) Note that $\forall\ \alpha \in (0, 1)$, $\forall\ \boldsymbol{\theta} \in \mathrm{R}^p$, $\boldsymbol{\theta}^t\mathbf{B}_x^{-1}\mathfrak{T}(\boldsymbol{\beta}(\alpha) + r\mathbf{A}_x^{-1}\boldsymbol{\theta}, \alpha)$ is an increasing function of $r > 0$. This fact together with an argument like the one given in the proof of Theorem 2.1 below implies (iv)(a) in a routine fashion.

(b) First we show that for every $0 < \alpha < 1$,

$$\left\| \mathbf{B}_x^{-1}\mathfrak{T}(\hat{\boldsymbol{\beta}}_n(\alpha), \alpha) \right\| = o_p(1). \tag{2.3}$$

From Theorem 3.3 of KB, we have the following algebraic identity:

$$[\sum_i \mathbf{x}_{ni}^t\{I(Y_{ni}-\mathbf{x}_{ni}^t\hat{\beta}_n(\alpha)\le 0)-\alpha\} - \sum_{i\in \mathbf{h}_n(\alpha)} \mathbf{x}_{ni}^t\{I(Y_{ni}-\mathbf{x}_{ni}^t\hat{\beta}_n(\alpha)\le 0)-\alpha\}]\mathbf{X}_{\mathbf{h}_n(\alpha)}^{-1} = \mathbf{w}_n^t(\alpha),$$

where, $\mathbf{h}_n(\alpha)$ is as in (1.6) and each element of the $p\times 1$ vector $\mathbf{w}_n(\alpha)$ belongs to the interval $[\alpha -1, \alpha]$. Hence

$$\mathbf{B}_x^{-1}\mathcal{T}(\hat{\beta}_n(\alpha), \alpha) = \mathbf{B}_x^{-1} \sum_{i\in \mathbf{h}_n(\alpha)} \mathbf{x}_{ni}(1-\alpha) + \mathbf{B}_x^{-1}\mathbf{X}_{\mathbf{h}_n(\alpha)}^t \mathbf{w}_n(\alpha).$$

Therefore, (2.2) follows from (2.3) and (X.2) by noting that the right hand side is bounded by $2\,p \max_{1\le i\le n} \|\mathbf{x}_{ni}^t\mathbf{B}_x^{-1}\| = o(1)$.

Now the claim in (b) follows along the same line as in the proof of (2.5) below. The claim about (c) follows from (b) with the help of Cramer-Wold device. //

Remark 2.1. An interesting special case of (2.2) is when $k=1$ and $\alpha_1 = 1/2$. Here regression quantile reduces to the celebrated LAD (*least absolute deviation*) estimator. As an example of errors when the limiting distribution of the LAD estimator is nonnormal, consider the model (1.1.1) when the error r.v. is a chi-square centered at its median, i.e., $G(x) = x^2 - v$, where v is the median of χ_1^2 r.v. Then, as shown in DT, m equals two. Hence in the one sample location model with chi-square errors centered at its median, it follows that the asymptotic distribution of the LAD estimator is the same as that of $v.\{n^{1-\theta}L(n)\}^{-1}\sum_i(\eta_i^2 - 1)$ which converges weakly to the r.v. $v.Z_2(1)$ where $\{Z_2(\alpha),\ \alpha\in[0,1]\}$ is the Rosenblatt process as in example 2 of DT. //

Remark 2.2. A very intriguing phenomenon is observed regarding the asymptotic efficiency of different estimators when the errors are exactly Gaussian, i.e., $G(x) = x$. In this case, m equals one and all the estimators discussed in this thesis are asymptotically normally distributed. Moreover, (2.2) implies that the

asymptotic dispersion of the regression quantiles $\{\hat{\beta}_n(\alpha);\ \alpha \in (0, 1)\}$ **is independent of** α. **In addition, it is same as the** asymptotic dispersion of the ***least squares estimator***, M-estimators and that of Koul's m. d.e. β_K^+, irrespective of the scores - functions (in the case of M-estimators) and integrating measure H (in the case of m.d.e.). Similarly, in this situation, R-estimators and Koul's m.d.e. $\hat{\beta}_K$ continue to have the same asymptotic dispersion irrespective of the score function ζ. **This is in complete contrast with the i.i.d. errors case.** Finding an estimator with better asymptotic efficiency in the case of Gaussian errors is still an open problem. //

Remark 2.3. Recall that in the i.i.d. errors case, under (X.0), (X.1), the assumption that $\max_{1 \le i \le n} \mathbf{x}_{ni}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}_{ni} = o(1)$, and (A.7) one obtains that $\forall\ \alpha \in (0, 1)$,

$$\text{(2.4)} \qquad \mathbf{D}_X(\hat{\beta}_n(\alpha) - \beta(\alpha)) \Rightarrow \mathbf{N}[0, \mathbf{I}_{p \times p}\ \alpha\ (1 - \alpha)\ q^{-2}(\alpha)] \text{ r.v.}$$

This essentially follows from the work of Koul (1992, section 5.4) by viewing $\hat{\beta}_n(\alpha)$ as an M-estimator corresponding to the convex score function h_α. Under stronger conditions on the design matrix it also follows from **KB**. Hence, remark similar to that of Remark 2.2.1 is also applicable here by comparing (2.2) and (2.4).

In the following theorem, we depict the asymptotic representation of the regression quantiles process uniformly over the compact subset of (0, 1). The basic idea for its proof can be found in Jureckova (1971) and Koul (1985). Here, for an $\mathbf{R}^p$ valued stochastic process $\{\mathbf{X}_n(\alpha),\ \alpha \in [0, 1]\}$, $\|\mathbf{X}_n\|_a := \sup\{\ \|\mathbf{X}_n(\alpha)\|\ ;\ a \le \alpha \le 1\text{-}a\}$ and we say that $\mathbf{X}_n = O_p^*(1)$ ($o_p^*(1)$) if $\|\mathbf{X}_n\|_a = O_p(1)$ ($o_p(1)$) for every $a \in (0, 1/2]$.

Theorem 2.1. *Assume that* (1.1.1), (1.1.2), (X.0) - (X.2) *and* (A.7) - (A.8) *hold. Then*

$$\text{(2.5)} \qquad \mathbf{A}_X(\hat{\beta}_{md}(\alpha) - \beta(\alpha)) + q^{-1}(\alpha)\ \mathbf{B}_X^{-1}\mathcal{T}(\beta(\alpha), \alpha) = o_p^*(1),$$

(2.6) $$A_x(\hat{\beta}_n(\alpha) - \hat{\beta}_{md}(\alpha)) = o_p^*(1). \qquad //$$

Proof. The proof will be given in several steps. Let $D_n(\alpha)$ denote the set of minimizers of $\|D_x^{-1}\, \mathcal{T}(\Delta, \alpha)\|$. Note that $\|D_x^{-1}\, \mathcal{T}(\Delta, \alpha)\|$ can take at most 2^n possible values and the set $D_n(\alpha)$ is nonempty for each α in (0, 1/2]. Now define $\hat{\Delta}_n(\alpha)$ by $A_x[\hat{\Delta}_n(\alpha) - \beta(\alpha)] := -\, q^{-1}(\alpha)\, B_x^{-1}\mathcal{T}(\beta(\alpha), \alpha)$. Therefore, to prove (2.5), it is enough to show that for every a in (0, 1/2],

(2.7) $$\sup_{\alpha \in I_a} \sup_{\Delta \in D_n(\alpha)} \|A_x[\Delta - \beta(\alpha)] - A_x[\hat{\Delta}_n(\alpha) - \beta(\alpha)]\| = o_p(1).$$

***Step* 1.** $\|B_x^{-1}\, \mathcal{T}(\hat{\Delta}_n(\alpha), \alpha)\| = o_p^*(1).$

Proof. Follows from the above Lemma (2.1)(iii) by observing that

$$A_x[\hat{\Delta}_n(\alpha) - \beta(\alpha)] = O_p^*(1).$$

***Step* 2.** $\inf_{\Delta \in R^p} \|B_x^{-1}\, \mathcal{T}(\Delta, \alpha)\| = o_p^*(1).$

Proof. Immediate from *Step* 1.

***Step* 3.** *For a* b>0, *define*

$O_b(\alpha) := \{\Delta \in D_n(\alpha);\ \|A_x(\Delta - \beta(\alpha))\| \le b\}$ *and*

$\bar{O}_\delta(\alpha) := \{\Delta \in D_n(\alpha);\ \|A_x(\Delta - \beta(\alpha))\| > b\}$

Then $\forall$ M > 0,

(2.8) $$P[\sup_{\alpha \in I_a} \sup_{\Delta \in O_b(\alpha)} \|A_x(\Delta - \hat{\Delta}_n(\alpha))\| > M,\ O_b(\alpha) \ne \phi\ \forall \alpha \in I_a] = o(1).$$

Proof. Note that $A_x(\Delta - \hat{\Delta}_n(\alpha)) = A_x[\Delta - \beta(\alpha)] - (-q^{-1}(\alpha))\, [B_x^{-1}\mathcal{T}(\beta(\alpha), \alpha)]$.

Therefore, the left hand side of (2.8) is less than

$$P[\sup_{\alpha \in I_a} \sup_{\Delta \in O_b(\alpha)} \|A_x[\Delta - \beta(\alpha)] - (-q^{-1}(\alpha))\, [B_x^{-1}\mathcal{T}(\beta(\alpha), \alpha)]))\| > M,$$

$$\sup_{\alpha \in I_a} \sup_{\Delta \in O_b(\alpha)} \|q^{-1}(\alpha) B_x^{-1}\mathcal{T}(\Delta, \alpha)\| < M/2,\ O_b(\alpha) \ne \phi\ \forall \alpha \in I_a]$$

$$+\ P[\sup_{\alpha \in I_a} \sup_{\Delta \in O_b(\alpha)} \|q^{-1}(\alpha) B_x^{-1}\mathcal{T}(\Delta, \alpha)\| \ge M/2,\ O_b(\alpha) \ne \phi\ \forall \alpha \in I_a]$$

$$\leq P[\sup_{\alpha \in I_a} \sup_{\Delta \in O_b(\alpha)} \left\| q^{-1}(\alpha) B_x^{-1}[\mathcal{T}(\Delta, \alpha) - \mathcal{T}(\beta(\alpha), \alpha)] - A_x[\Delta - \beta(\alpha))\right\| > M/2,$$

$$O_b(\alpha) \neq \phi \ \forall \alpha \in I_a]$$

$$+ P[\sup_{\alpha \in I_a} \sup_{\Delta \in O_b(\alpha)} \left\| q^{-1}(\alpha) B_x^{-1} \mathcal{T}(\Delta, \alpha)\right\| \geq M/2, O_b(\alpha) \neq \phi \ \forall \alpha \in I_a].$$

Now, the first and second terms are o(1) by Lemma (2.1)(iii) and *Step* 2 respectively.

***Step* 4.** *Given* ϵ, M >0, $\exists\ \delta>0$ *and* n_0 *such that* $\forall\ n \geq n_0$,

(2.9) $P[\inf_{\alpha \in I_a} \inf_{\|s\| \geq \delta} \left\| B_x^{-1} \mathcal{T}(\beta(\alpha) + A_x^{-1}s, \alpha)\right\| > M] > 1 - \epsilon.$

Proof. The polar representation of vectors, the Cauchy-Schwarz inequality and the fact that $\forall\ \alpha \in (0, 1)$ and $\forall\ \boldsymbol{\theta} \in R^p$, $\boldsymbol{\theta}^t B_x^{-1}[\mathcal{T}(\beta(\alpha) + rA_x^{-1}\boldsymbol{\theta}, \alpha)$ is a monotone increasing function of $r > 0$, yields that for any $\delta > 0$,

$$\inf_{\|s\| \geq \delta} \left\| B_x^{-1} \mathcal{T}(\beta(\alpha) + A_x^{-1}s, \alpha)\right\|$$

$$\geq \inf_{r \geq \delta} \inf_{\|\boldsymbol{\theta}\| = 1} \boldsymbol{\theta}^t B_x^{-1} \mathcal{T}(\beta(\alpha) + A_x^{-1}\boldsymbol{\theta}\, r, \alpha) \geq \inf_{\|\boldsymbol{\theta}\| = 1} \boldsymbol{\theta}^t B_x^{-1} \mathcal{T}(\beta(\alpha) + A_x^{-1}\boldsymbol{\theta}\, \delta, \alpha).$$

Therefore, using Lemma 2.1(iii), for all sufficiently large n, the left hand side of (2.9) is not less than

$$P[\inf_{\alpha \in I_a} \inf_{\|\boldsymbol{\theta}\| = 1} \boldsymbol{\theta}^t B_x^{-1} \mathcal{T}(\beta(\alpha) + A_x^{-1}\boldsymbol{\theta}\, \delta, \alpha) > M]$$

$$\geq P[\inf_{\alpha \in I_a} \inf_{\|\boldsymbol{\theta}\| = 1} [\boldsymbol{\theta}^t B_x^{-1} \mathcal{T}(\beta(\alpha), \alpha) - \delta\, q(\alpha)] > M + 1] - \epsilon/2$$

$$\geq P[\ \delta \inf_{\alpha \in I_a} q(\alpha) - \inf_{\alpha \in I_a} \inf_{\|\boldsymbol{\theta}\| = 1} \boldsymbol{\theta}^t B_x^{-1} \mathcal{T}(\beta(\alpha), \alpha) > M + 1] - \epsilon/2.$$

Now, using $\inf_{\alpha \in I_a} q(\alpha) > 0$ and $\inf_{\alpha \in I_a} \inf_{\|\boldsymbol{\theta}\| = 1} \boldsymbol{\theta}^t B_x^{-1} \mathcal{T}(\beta(\alpha), \alpha) = O_p(1)$, we can choose δ sufficiently large so that the above probability is not less than $1 - \epsilon/2$.

***Step* 5.** *Given* ϵ, M >0, $\exists$ n_0 *such that* $\forall$ $n \geq n_0$,

$$\text{(2.10)} \quad P[\sup_{\alpha \in I_a} \sup_{\Delta \in D_n(\alpha)} \| A_x[\Delta - \beta(\alpha)] - A_x[\hat{\Delta}_n(\alpha) - \beta(\alpha)] \| > M] < \epsilon.$$

Proof. By *Step* 4, choose δ and n large such that P[$\bar{O}_\delta(\alpha) \neq \phi$ for some α in I_a] < $\epsilon/3$. Then the left hand side of (2.10) is less than

$$P[\sup_{\alpha \in I_a} \sup_{\Delta \in D_n(\alpha)} \| A_x[\Delta - \beta(\alpha)] - A_x[\hat{\Delta}_n(\alpha) - \beta(\alpha)] \| > M, \ O_b(\alpha) \neq \phi \ \forall \alpha \in I_a]$$

$$+ P[\bar{O}_\delta(\alpha) \neq \phi \ \text{ for some } \alpha \text{ in } I_a]$$

$$\leq P[\sup_{\alpha \in I_a} \sup_{\Delta \in O_\delta(\alpha)} \| A_x[\Delta - \beta(\alpha)] - A_x[\hat{\Delta}_n(\alpha) - \beta(\alpha)] \| > M, \ O_\delta(\alpha) \neq \phi \ \forall \alpha \in I_a]$$

$$+ P[\sup_{\alpha \in I_a} \sup_{\Delta \in \bar{O}_\delta(\alpha)} \| A_x[\Delta - \beta(\alpha)] - A_x[\hat{\Delta}_n(\alpha) - \beta(\alpha)] \| > M, \ O_\delta(\alpha) \neq \phi \forall \alpha \in I_a] + \epsilon/3.$$

The first probability is less than $\epsilon/3$ by *Step* 3. The second probability is less than $\epsilon/3$ by the choice of δ. This completes the proof of (2.5). Assertion (2.6) can be following the same lines by observing that $\| B_x^{-1} \mathcal{T}(\hat{\beta}_n(\alpha), \alpha) \| = o_p^*(1)$. //

We now turn to the **RR** processes. To define these, let $\{c_{ni}, 1 \leq i \leq n\}$ be a triangular array of $p \times 1$ vectors and let $C_{n \times p}$ be the matrix with rows c_{ni}^t, $1 \leq i \leq n$. Define sequences of weighted **RR** processes by

$$\hat{U}_n^c(\alpha) := \sum_i c_{ni} \{ \hat{a}_{ni}(\alpha) - (1 - \alpha) \}, \quad 0 \leq \alpha \leq 1,$$

and an approximating sequence of weighted empirical process

$$U_n^c(\alpha) := \sum_i c_{ni} \{ I[\epsilon_i > F^{-1}(\alpha)] - (1 - \alpha) \}, \quad 0 \leq \alpha \leq 1.$$

For convenience we now recall the following algebraic identity from (5.15) of **GJ**.

$$\begin{aligned} \text{(2.10)} \quad \hat{a}_{ni}(\alpha) - (1 - \alpha) = {} & I[\epsilon_i > F^{-1}(\alpha)] - (1 - \alpha) \\ & - \{ I[\epsilon_i \leq F^{-1}(\alpha) + x_{ni}^t(\hat{\beta}_n(\alpha) - \beta(\alpha)) - I[\epsilon_i \leq F^{-1}(\alpha)] \} \\ & + \hat{a}_{ni}(\alpha) \, I[Y_{ni} = x_{ni}^t(\hat{\beta}_n(\alpha)], \qquad 1 \leq i \leq n, \ 0 \leq \alpha \leq 1. \end{aligned}$$

This identity is useful in approximating $\hat{U}_n^c$ by U_n^c. The following theorem gives the desired asymptotic representation of $\hat{U}_n^c$.

Theorem 2.2. *In addition to* (1.1.1), (1.1.2), (X.0) - (X.2) (A.7) - (A.8) *assume that* (C.1) - (C.2) *hold where,*

(C.1) $(C^tC)^{-1}$ exists for all $n \geq p$. (C.2) $n \max_{1 \leq i \leq n} c_{ni}^t (C^tC)^{-1} c_{ni} = O(1)$.

Then the regression rank-scores process admits the following representation:

(2.11) $$B_c^{-1}[\hat{U}_n^c(\alpha) - U_n^c(\alpha)] = B_c^{-1} C^t X A_x^{-1} A_x(\hat{\beta}_n(\alpha) - \beta(\alpha))\, q(\alpha) + o_p^*(1),$$

and

(2.12) $$B_c^{-1} U_n^c(\alpha) = - S_c J_m(F^{-1}(\alpha)) + o_p^*(1).$$

Consequently,

(2.13) $$B_c^{-1}\hat{U}_n^c(\alpha) = - [S_c - B_c^{-1} C^t X A_x^{-1} S_x]\, J_m(F^{-1}(\alpha)) + o_p^*(1). \qquad //$$

Remark 2.4. Let $\mathfrak{g} := B_c^{-1} C^t X A_x^{-1}$. Then from its definition $\mathfrak{g} = D_c^{-1} C^t X D_x^{-1}$. Hence by the Cauchy-Schwarz inequality $\mathfrak{g}$ is bounded. //

Proof of the Theorem. Let $E_n(\alpha) := A_x[\hat{\beta}_n(\alpha) - \beta(\alpha)]$. From (2.10) we obtain that

$$B_c^{-1}\hat{U}_n^c(\alpha) = B_c^{-1} U_n^c(\alpha) - \mathfrak{g}\hat{E}(\alpha)\, q(\alpha)$$

$$-[B_c^{-1} \sum_i c_{ni}\{I[\epsilon_i \leq F^{-1}(\alpha) + x_{ni}^t A_x^{-1} E_n(\alpha) - I[\epsilon_i \leq F^{-1}(\alpha)]\} - \mathfrak{g} E_n(\alpha)\, q(\alpha)]$$

$$+ B_c^{-1} \sum_{i \in \mathfrak{h}_n(\alpha)} c_{ni}\hat{a}_{ni}(\alpha)\, I[Y_{ni} = x_{ni}^t(\hat{\beta}_n(\alpha)]$$

$$= B_c^{-1} U_n^c(\alpha) - \mathfrak{g}\, E_n(\alpha)\, q(\alpha) - R_1(\alpha) + R_2(\alpha), \quad \text{(say)}.$$

By Theorem 2.1, (2.6) and Remark 2.4, $R_1 = o_p^*(1)$. By (C.2), $\sup_{\alpha \in [0,1]} \|R_2(\alpha)\| = o(1)$ almost surely. Hence (2.11) follows. The relation (2.12) follows as in the proof of Lemma 2.1(i).

Remark 2.5. As in Remark 2.1, the nature of the approximating process in (2.13) is quite different from that in the i.i.d. errors case. The leading r.v.

$$(2.14) \qquad Z := -[S_c - B_c^{-1}C^tXA_x^{-1}S_x]\ J_m(F^{-1}(\alpha))$$

in the right hand side of (2.13) is a product of a random quantity, independent of α, and a nonrandom continuous function of α, which also depends on the Hermite rank m. If m =1, then **Z** is a multivariate normal r.v. with mean **0** and dispersion matrix proportional to

$$(2.15) \qquad B_c^{-1}C^t(I - X(X^tX)^{-1}X)\ R\ (I - X(X^tX)^{-1}X)CB_c^{-1},$$

where **R** is the dispersion matrix of $(\eta_1, \eta_2 \ldots \eta_n)'$. For m other than one the limiting distribution may not be normal. Also note that unlike the i.i.d. errors case the limiting distribution may not be distribution free. //

4.3. L-estimators and regression rank scores statistics.

In this section we derive the asymptotic distribution of smoothed L-estimators based on **RQ** processes. For a finite signed measure ν on (0, 1) with compact support, an L-estimator of β is defined by

$$(3.1) \qquad \mathbf{T}_n = \int_0^1 \hat{\beta}_n(\alpha)\ d\nu(\alpha).$$

Note that for ν with $d\nu(\alpha) = I(a \le \alpha \le 1\text{-}a)d\alpha$, $\mathbf{T}_n$ reduces to an analog of the trimmed mean. The following theorem is an immediate consequence of Theorem 2.1.

Theorem 3.1. *Under the assumptions of Theorem 2.1, with* $\mu := \nu(0, 1)$,

$$(3.2) \qquad A_x[\mathbf{T}_n - \beta\mu - e_1\int_0^1 F^{-1}(\alpha)d\nu(\alpha)] = -S_x\int_0^1 J_m(F^{-1}(\alpha))\ q^{-1}(\alpha)\ d\nu(\alpha) + o_p(1). \quad //$$

Remark 3.1. Consider the linear model (1.1.1) where now $\{\epsilon_i\}$ are i.i.d. F. Let ψ be an absolutely continuous function from R^1 to R^1 such that $\int \psi dF = 0$, $0 < \int \psi^2 dF < \infty$ and its almost everywhere derivative ψ' satisfies $0 < \int \psi' dF$, $\int (\psi')^2 dF < \infty$. Then from GJ, it can be deduced that if ϵ has a continuous positive density f, then the normalized M-estimator corresponding to ψ and an L-estimator corresponding to the measure ν given by the relation $d\nu(\alpha) = \psi'(F^{-1}(\alpha))/E\psi'(\epsilon)d\alpha$ coincide asymptotically. The same phenomenon happen in the LRD setup with the same choice of the measure ν.

To explain this further, suppose that the m in (1.2.2) is also equal to the Hermite rank m_1 of $\psi(G(\eta))$, i.e., $m = \inf\{k \geq 1\colon E\psi(G(\eta))H_k(\eta) \neq 0\}$ and ψ is constant outside compact. Koul (1992) showed that under the assumptions that the function $z \to E|\psi'(\epsilon - z) - \psi'(\epsilon)|$ is continuous at zero, (1.1.1), (1.1.2) and (X.0) - (X.2),

(3.3) $$A_x(\tilde{\beta} - \beta) = [E\psi'(\epsilon)]^{-1} S_x J_m(\psi) + o_p(1),$$

where $\tilde{\beta}$ is the M-estimator of β corresponding to ψ and $J_m(\psi) := E\psi(G(\eta))H_m(\eta)$. Note that with $d\nu(\alpha) = [E\psi'(\epsilon)]^{-1}\psi'(F^{-1}(\alpha))\, d\alpha$,

$$\int_0^1 J_m(F^{-1}(\alpha))\, q^{-1}(\alpha) d\nu(\alpha)$$

$$= [E\psi'(\epsilon)]^{-1} EH_m(\eta) \int_0^1 I(G(\eta) \leq F^{-1}(\alpha))\, \psi'(F^{-1}(\alpha))\, q^{-1}(\alpha)\, d\alpha = [E\psi'(\epsilon)]^{-1} J_m(\psi).$$

Hence from (3.2), we readily obtain that

$$A_x(T_n - \beta - e_1[E\psi'(\epsilon)]^{-1} E[\epsilon\psi'(\epsilon)]\,) = -[E\psi'(\epsilon)]^{-1} J_m(\psi)\, S_x + o_p(1). \qquad //$$

Next we turn to the *linear regression rank-scores statistics* as defined in GJ. They are of the form $V_n^c = \sum_i c_{ni}\hat{b}_{ni}$, where $\{\hat{b}_{ni}\}$ are the scores generated by

(3.4) $$\hat{b}_{ni} = -\int_0^1 b(\alpha)\, d\hat{a}_{ni}(\alpha),\ 1 \leq i \leq n,$$

for a function b from (0, 1) to $\mathbf{R}^1$ which is of bounded variation and constant outside a compact sub-interval of (0, 1). From (1.8) it is easy to see that, in the location model, $\hat{b}_{ni}$ reduces to the familiar one sample score

$$\hat{b}_{ni} = n \int_{(R_i - 1)/n}^{R_i/n} b(\alpha)\, d\alpha$$

Theorem 3.2. *Under the assumptions of Theorem* 2.2,

(3.5) $$\mathbf{B}_c^{-1}(\mathbf{V}_n^c - \sum_i \mathbf{c}_{ni}\bar{b}) = -(\mathbf{S}_c - \mathfrak{g}\, \mathbf{S}_x)\tilde{b} + o_p(1),$$

where

(3.6) $$\bar{b} := \int_0^1 b(\alpha)\, d\alpha \quad \textit{and} \quad \tilde{b} := -E\{b(F(G(\eta)))H_m(\eta)\}. \qquad //$$

Proof. Integration by parts yields that $\hat{b}_{ni} = b(0) + \int_0^1 \hat{a}_{ni}(\alpha)\, db(\alpha)$. Hence,

$$\mathbf{B}_c^{-1}\int_0^1 \hat{\mathbf{U}}_n^c\, db(\alpha) = \mathbf{B}_c^{-1}\sum_i \mathbf{c}_{ni}\hat{b}_{ni} - \mathbf{B}_c^{-1}\sum_i \mathbf{c}_{ni}[\, b(0) + \int_0^1 (1 - \alpha) db(\alpha)]$$

$$= \mathbf{B}_c^{-1}(\mathbf{V}_n^c - \sum_i \mathbf{c}_{ni}\bar{b}).$$

Therefore (3.5) follows from (2.13), since by Fubini's Theorem

$$\int_0^1 J_m(F^{-1}(\alpha))\, db(\alpha) = -\, E\{b(F(G(\eta)))H_m(\eta)\}. \qquad //$$

Remark 3.2. Observe that the random coefficient in the leading term on the right hand side of (3.5) is the same as in (2.14). In general, $\tilde{b}$ depends on the unknown G. This in turn implies that unlike the i.i.d. errors case, the sequence of r.v.'s $\{\mathbf{B}_c^{-1}(\mathbf{V}_n^c - \sum_i \mathbf{c}_{ni}\bar{b})\}$ is not asymptotically distribution free. (a.d.f.).

However, if G is strictly increasing with d.f. F, then $G = F^{-1}\Phi$ and in this case $m = 1$, $\tilde{b} = -\, E\, b(\Phi(\eta))\eta$. Hence in this case the r.v.'s are a.d.f. In order for this result to be useful for testing about the slope parameters $\boldsymbol{\beta}$ it is necessary to estimate $\boldsymbol{\theta}$ that appears in $\mathbf{B}_c$ and $\mathbf{R}$ in (2.15). Let $\hat{\boldsymbol{\theta}}_n$ be an estimator of $\boldsymbol{\theta}$ such that $(\hat{\theta}_n - \theta)\log_e n = o_p(1)$. The estimator of θ given by Yajima (1988) is known to satisfy this condition. Let $\hat{\mathbf{B}}_c$ and $\hat{\mathbf{R}}$ denote $\mathbf{B}_c$ and $\mathbf{R}$ respectively after θ is replaced by $\hat{\theta}_n$ in these entities. Then, the sequence of statistics $\hat{\mathbf{R}}^{-1/2}\,\hat{\mathbf{B}}_c^{-1}\,[\mathbf{V}_n^c - \sum_i \mathbf{c}_{ni}\bar{\mathbf{b}}]$ converges in distribution to a $N[0, \mathbf{I}_{p\times p}]$ r.v. and this can be used to test different hypotheses concerning $\boldsymbol{\beta}$.

4.4. Asymptotic uniform linearity of linear regression rank-scores statistics.

This section obtains the AUL of the RR processes and statistics based on residuals. These results are similar to those obtained by Jureckova (1992a) for the i.i.d. errors case. They are useful for testing sub-hypotheses and the estimation of some slope parameters when others are treated as nuisance parameters. See Jureckova (1992b) for other applications of the AUL results.

Accordingly, let $\{\mathbf{r}_{ni},\ 1 \le i \le n\}$ be a triangular array of $p\times 1$ vectors and let $\mathbf{R}_{n\times p}$ be the matrix with rows $\mathbf{r}_{ni}^t$, $1 \le i \le n$ satisfying the following two conditions:

(R.1) $(\mathbf{R}^t\mathbf{R})^{-1}$ exists for all $n \ge p$. (R.2) $\max_{1\le i\le n} \left\| \mathbf{r}_{ni}^t \mathbf{A}_r^{-1} \right\| = o(1)$.

Let $Y_{nit} := Y_{ni} - \mathbf{r}_{ni}^t\mathbf{A}_r^{-1}\mathbf{t}$, $1 \le i \le n$, $\mathbf{t} \in \mathbb{R}^p$ and $\mathbf{Y}_{nt} := [Y_{n1t}, Y_{n2t}, \ldots, Y_{nnt}]^t$. Let $\mathfrak{T}(\boldsymbol{\beta}(\alpha), \alpha, \mathbf{Y}_{nt})$, $\hat{\boldsymbol{\beta}}_{md}(\alpha, \mathbf{Y}_{nt})$, $\hat{\boldsymbol{\beta}}_n(\alpha, \mathbf{Y}_{nt})$ denote $\mathfrak{T}(\boldsymbol{\beta}(\alpha), \alpha)$, $\hat{\boldsymbol{\beta}}_{md}(\alpha)$, $\hat{\boldsymbol{\beta}}_n(\alpha)$ etc. respectively, when $\{Y_{ni}\}$ are replaced by $\{Y_{nit}\}$ in their definitions. The following lemma is similar in spirit to Lemma 2.1 and Theorem 2.1. It gives the asymptotic representation of the regression quantiles processes based on the residuals $\{Y_{nit}\}$ and the proof is similar to that of Lemma 2.1. Here $\bar{o}_p(1)$ $(\bar{O}_p(1))$ denotes a sequence of stochastic processes that converge to zero (bounded) in probability, uniformly over $a \le \alpha \le 1 - a$, $\|\mathbf{t}\| \le L$, $\forall\ a \in (0, 1/2]$, $L \in (0, \infty)$. Also racall the notation $o_p^*(1)$ from Theorem 2.1.

Lemma 4.1 *Assume that* (1.1.1), (1.1.2), (X.0) - (X.2), (A.7), (A.8), (R.1) *and* (R.2) *hold. Then*

(4.1) $$\left\| \mathbf{B}_x^{-1}[\mathfrak{T}(\beta(\alpha), \alpha, \mathbf{Y}_{nt}) - \mathfrak{T}(\beta(\alpha), \alpha)] - \mathbf{B}_x^{-1}\mathbf{X}^t\mathbf{R}\mathbf{A}_r^{-1}\mathbf{t}\, q(\alpha) \right\| = \bar{o}_p(1),$$

(4.2) $$\sup\{ \left\| \mathbf{B}_x^{-1}[\mathfrak{T}(\beta(\alpha) + \mathbf{A}_x^{-1}\mathbf{s}, \alpha, \mathbf{Y}_{nt}) - \mathfrak{T}(\beta(\alpha), \alpha, \mathbf{Y}_{nt})] - \mathbf{s}\, q(\alpha) \right\| \} = o_p(1),$$

where the above supremum is taken over $a \leq \alpha \leq 1 - a$, $\|\mathbf{s}\| \leq K$, $\|\mathbf{t}\| \leq L$.

(4.3) $$\mathbf{A}_x[\hat{\beta}_{md}(\alpha, \mathbf{Y}_{nt}) - \beta(\alpha)] = - q^{-1}(\alpha)\{S_x J_m(F^{-1}(\alpha)) + \mathbf{B}_x^{-1}\mathbf{X}^t\mathbf{R}\mathbf{A}_r^{-1}\mathbf{t}\, q(\alpha)\} + \bar{o}_p(1),$$

(4.4) $$\mathbf{A}_x\, [\hat{\beta}_{md}(\alpha, \mathbf{Y}_{nt}) - \hat{\beta}_n(\alpha, \mathbf{Y}_{nt})] = \bar{o}_p(1),$$

(4.5) $$\mathbf{A}_x[\hat{\beta}_n(\alpha, \mathbf{Y}_{nt}) - \beta(\alpha)] = \bar{O}_p(1).$$ //

The following theorem gives the main result of this section.

Theorem 4.1. *Assume that* (1.1.1), (1.1.2), (X.0) - (X.2), (A.7), (A.8), (R.1), (R.2) *and* (C.1) - (C.3) *hold, where*

(C.3) $$\mathbf{C}^t\mathbf{X} = \mathbf{0}.$$

Then

(4.6) $$\mathbf{B}_c^{-1}\hat{\mathbf{U}}_n^c(\alpha, \mathbf{Y}_{nt}) = \mathbf{B}_c^{-1}\hat{\mathbf{U}}_n^c(\alpha) + \mathbf{B}_c^{-1}\mathbf{C}^t\mathbf{R}\mathbf{A}_r^{-1}\mathbf{t}\, q(\alpha) + \bar{o}_p(1).$$

Moreover, if the score function b *is of bounded variation and constant outside a compact subinterval of* (0, 1), *then* $\forall\ 0 < L < \infty$,

(4.7) $$\sup_{\|\mathbf{t}\| \leq L} \left\| \mathbf{B}_c^{-1}[\mathbf{V}_n^c(\mathbf{Y}_{nt}) - \mathbf{V}_n^c] - \mathbf{B}_c^{-1}\mathbf{C}^t\mathbf{R}\mathbf{A}_r^{-1}\, \mathbf{t} \int_0^1 q(\alpha)\, db(\alpha) \right\| = o_p(1).$$ //

Proof. Let $\hat{\mathbf{E}}_n(\alpha, \mathbf{t}) := \mathbf{A}_x [\hat{\beta}_n(\alpha, \mathbf{Y}_{nt}) - \beta(\alpha)]$. From (1.6), for $1 \le i \le n$,

$$\hat{a}_{ni}(\alpha, \mathbf{Y}_{nt}) = I[Y_{nit} > \mathbf{x}_{ni}^t\, \hat{\beta}_n(\alpha, \mathbf{Y}_{nt})] + \hat{a}_{ni}(\alpha, \mathbf{Y}_{nt})\, I[Y_{nit} = \mathbf{x}_{ni}^t\, \hat{\beta}_n(\alpha, \mathbf{Y}_{nt})]$$

$$= 1 - I[\epsilon_i \le F^{-1}(\alpha) + \mathbf{x}_{ni}^t \mathbf{A}_x^{-1} \hat{\mathbf{E}}_n(\alpha, \mathbf{t}) + \mathbf{r}_{ni}^t \mathbf{A}_r^{-1} \mathbf{t}]$$

$$+ \hat{a}_{ni}(\alpha, \mathbf{Y}_{nt})\, I[Y_{nit} = \mathbf{x}_{ni}^t\, \hat{\beta}_n(\alpha, \mathbf{Y}_{nt})].$$

Hence, $\mathbf{B}_c^{-1}[\hat{\mathbf{U}}_n^c(\alpha, \mathbf{Y}_{nt}) - \mathbf{B}_c^{-1}\hat{\mathbf{U}}_n^c(\alpha)]$

$$= -\mathbf{B}_c^{-1} \sum_i \mathbf{c}_{ni}\{I[\epsilon_i \le F^{-1}(\alpha) + \mathbf{x}_{ni}^t \mathbf{A}_x^{-1} \hat{\mathbf{E}}_n(\alpha, \mathbf{t}) + \mathbf{r}_{ni}^t \mathbf{A}_r^{-1} \mathbf{t}] - I[\epsilon_i \le F^{-1}(\alpha) + \mathbf{x}_{ni}^t \mathbf{A}_x^{-1} \hat{\mathbf{E}}_n(\alpha)]\}$$

$$+ \mathbf{B}_c^{-1} \sum_i \mathbf{c}_{ni} \hat{a}_{ni}(\alpha, \mathbf{Y}_{nt})\, I[Y_{nit} = \mathbf{x}_{ni}^t\, \hat{\beta}_n(\alpha, \mathbf{Y}_{nt})] - \mathbf{B}_c^{-1} \sum_i \mathbf{c}_{ni} \hat{a}_{ni}(\alpha)\, I[Y_{ni} = \mathbf{x}_{ni}^t\, \hat{\beta}_n(\alpha)]$$

$$= -\mathbf{R}_1(\alpha, \mathbf{t}) + \mathbf{R}_2(\alpha, \mathbf{t}) - \mathbf{R}_2(\alpha), \quad \text{(say)}. \tag{4.8}$$

By (C.2), $\mathbf{R}_2(\alpha, \mathbf{t}) = \bar{o}_p(1)$ and $\mathbf{R}_2(\alpha) = o_p^*(1)$. To handle $\mathbf{R}_1(\alpha, \mathbf{t})$, let

$$\mathbf{T}(\alpha, \mathbf{s}, \mathbf{t}) := \mathbf{B}_c^{-1} \sum_i \mathbf{c}_{ni}\{I[\epsilon_i \le F^{-1}(\alpha) + \mathbf{x}_{ni}^t \mathbf{A}_x^{-1}\, \mathbf{s} + \mathbf{r}_{ni}^t \mathbf{A}_r^{-1} \mathbf{t}], \qquad \mathbf{s}, \mathbf{t} \in \mathbf{R}^p.$$

Applying Theorem 1.2.2(iii) p times, j^{th} time with $\gamma_{ni} = j^{th}$ component of $\mathbf{D}_c^{-1}\mathbf{c}_{ni}$, $1 \le j \le p$, and $\xi_{ni} = \mathbf{x}_{ni}^t \mathbf{A}_x^{-1} \mathbf{s} + \mathbf{r}_{ni}^t \mathbf{A}_r^{-1} \mathbf{t}$ to conclude that $\forall$ $a \in (0, 1/2]$, $\mathbf{s}$, $\mathbf{t} \in \mathbf{R}^p$,

$$\mathbf{T}(\alpha, \mathbf{s}, \mathbf{t}) - \mathbf{T}(\alpha, \mathbf{0}, \mathbf{0}) - \mathbf{B}_c^{-1} \sum_i \mathbf{c}_{ni} \mathbf{x}_{ni}^t \mathbf{A}_x^{-1}\, \mathbf{s} - \mathbf{B}_c^{-1} \sum_i \mathbf{c}_{ni} \mathbf{r}_{ni}^t \mathbf{A}_r^{-1} \mathbf{t} = o_p^*(1).$$

In view of (C.3) and an argument similar to the proof of (2.3.2) yields that

$$\sup\{\mathbf{T}(\alpha, \mathbf{s}, \mathbf{t}) - \mathbf{T}(\alpha, \mathbf{0}, \mathbf{0}) - \mathbf{B}_c^{-1} \mathbf{C}^t \mathbf{R} \mathbf{A}_r^{-1}\, \mathbf{t}\, q(\alpha)\} = o_p(1), \tag{4.9}$$

where now the supremum is taken over $a \le \alpha \le 1 - a$, $\|\mathbf{s}\| \le K$, $\|\mathbf{t}\| \le L$. Hence, (4.5) and (4.9) yields that

$$\mathbf{B}_c^{-1} \sum_i \mathbf{c}_{ni} I[\epsilon_i \le F^{-1}(\alpha) + \mathbf{x}_{ni}^t \mathbf{A}_x^{-1} \hat{\mathbf{E}}_n(\alpha, \mathbf{t}) + \mathbf{r}_{ni}^t \mathbf{A}_r^{-1} \mathbf{t}] - \mathbf{T}(\alpha, \mathbf{0}, \mathbf{0}) - \mathbf{B}_c^{-1} \mathbf{C}^t \mathbf{R} \mathbf{A}_r^{-1} \mathbf{t}\, q(\alpha) = \bar{o}_p(1). \tag{4.10}$$

By a similar type of argument,

(4.11) $$B_c^{-1}\sum_i c_{ni} I[\epsilon_i \le F^{-1}(\alpha) + x_{ni}^t A_x^{-1}\hat{E}_n(\alpha) - T(\alpha, 0, 0) = o_p^*(1).$$

Combining (4.8), (4.10) and (4.11) to get (4.6). Assertion (4.7) follows from (4.6) by using *integration by parts* and the assumption that $\sum_i c_{ni} = 0$.

BIBLIOGRAPHY

Boos, D. (1981). Minimum distance estimators for location and goodness of fit. *JASA* **76**, 663-670.

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

Cramer, H. (1946). *Mathematical methods of statistics*. Princeton University Press.

Dehling, H. and Taqqu, M.S. (1989). The empirical process of some long range dependent sequences with an application to U-statistics. *Ann. Statist.*, **17**, 1767-1783.

Dobrushin, R.L. and Major, P. (1979). Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahhrsch.* **55**. 27-52.

Fine, T. (1966). On Hodges-Lehmann shift estimator in the two sample problem. *Ann. Statist.*, **37**, 1814-1818.

Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Ann. Statist.*, **14**. 517-532.

Gastwirth, J.L. and Rubin, H. (1975). The behavior of robust estimators on dependent data. *Ann. Statist.* **3**. 1070-1100.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *J. Time Ser. Anal.* 1, 15-29.

Gutenbrunner, C. and Jureckova, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20**. 305-329.

Hajek, J. and Sidak, Z. (1967). *Theory of rank tests*. Academic Press, N.Y.

Hodges, J.L., Jr. and Lehmann, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34**. 598-611.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika* **68**. 165-176.

Huber, P. (1981). ***Robust Statistics.*** Wiley, N.Y.

Jaeckel, L.A. (1972). Estimating regression coefficient by minimizing the dispersion of residuals. *Ann. Math. Statist.* **43**. 1449-1458.

Jureckova, J. (1971). Nonparametric estimates of regression coefficient. *Ann. Math. Statist.* **42**. 1328-1338.

Jureckova, J. (1992a). Uniform asymptotic linearity of regression rank scores process. *Proceedings of the meeting on Nonparametric Statistics and Related Topics.*Edited by A.K.Md. E.Saleh. Elseviers Sciences Publishers. pp 217-228.

Jureckova, J. (1992b). Estimation in a linear model based on regression rank scores. *J. Nonparametric Statistics.* **1**. 197-204.

Koenker, R. and Basset, G. (1978). Regression quantiles. *Econometrica.* **46**. 33-50.

Koenker, R. and Portnoy, S. (1987). L-estimation for linear models. *JASA* **82**. 851-113.

Koul, H.L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40**, 1950-1979.

Koul, H.L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* **41**, 1768-1773.

Koul, H.L. (1977). Behavior of robust estimators in the regression model with dependent errors. *Ann. Statist.* **5**, 681-699.

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

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

Koul, H.L. (1991). A weak convergence result useful in robust auto regression. *J.Statist. Planning and Inference* **29**, 291-308.

Koul, H.L. (1992). *Weighted Empiricals and Linear Models*. IMS Lecture Notes. 21.

Koul, H.L. (1992a). M- estimators in linear models with long range dependent errors. *Statist. & Prob. Letters.* **14**. 153-164.

Koul, H. and Dewet, T. (1983). Minimum distance estimation in a linear regression model. *Ann. statist.* **11**. 921-932.

Mandelbrot, B.B. and Taqqu, M.S. (1979). Robust R/S of long run serial correlation. *Proc. 42 Session of the International Statist. Institute. Bulletin of the I.S.I.* **48**, *Book 2,* 69-104.

Mandelbrot, B.B. and Van Ness, J.W.(1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**. 422-437.

Millar, P.W. (1981). Robust estimation via minimum distance methods. *Z. Wahhrsch.* **55**. 73-89.

Parr, W.C. and Schucanny, W.R. (1980). Minimum distance and robust estimation. *JASA* **75**. 616-624.

Portnoy, S. (1991). Asymptotic behavior of regression quantiles in non-stationary, dependent cases. *J. Multi. Analys.* **38**. 100-113.

Portnoy, S. and Koenker, R.(1989). Adaptive L-estimation of linear models. *Ann. Statist.* **17**. 362-381.

Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in linear models. *JASA* **75**. 828-838.

Sanson, G. (1959). *Orthogonal Functions*. Interscience Publishers, N.Y.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, N.Y.

Taqqu, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahhrsch.* **31**. 287-302.

Taqqu, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank.

Z. Wahhrsch. **55**. 53-83.

Wolfowitz, J. (1957). Minimum distance method. *Ann. Math. Statist.* **2**. 75-88.

Yajima, Y. (1985). On estimation of a long-memory time series models. *Austral. J. Statist.* **27**. 303-320.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16**. 791-807.

Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19**. 158-177.