Regression Quantile Spacings

Keith Knight*   Chuan Goh†

March 2006‡

1 Introduction

Assume a linear model with independently and identically distributed errors where the conditional α-quantile of the response variables Y_i given the d-dimensional predictor $\boldsymbol{x}_i$ is $\boldsymbol{x}_i^T\boldsymbol{\beta}(\alpha)$. Set $\epsilon_i \equiv Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}(\alpha)$; we will assume that the $\{\epsilon_i\}$ are iid random variables with α-quantile zero.

Recall that the *regression α-quantile* estimator[1] $\hat{\boldsymbol{\beta}}_n(\alpha)$ minimizes the criterion function

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - \boldsymbol{x}_i^T\boldsymbol{\phi}), \tag{1}$$

where $\rho_\alpha(s) \equiv s\left\{\alpha - 1(s<0)\right\}$. Under appropriate regularity conditions, $\hat{\boldsymbol{\beta}}_n(\alpha)$ is asymptotically normally distributed. We are interested in this paper in the limiting behaviour of the regression quantile "spacings" $\hat{\boldsymbol{\beta}}_n\left(\alpha + \frac{m}{n}\right) - \hat{\boldsymbol{\beta}}_n(\alpha)$ for some $m > 0$.

*Department of Statistics, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3; `keith@utstat.utoronto.ca`; `http://www.utstat.utoronto.ca/keith/home.html`

†Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario, Canada M5S 3G7; `goh@economics.utoronto.ca`; `http://www.chuangoh.org`

‡**Embarrassingly preliminary and incomplete. Comments welcome.** Knight acknowledges support from the Natural Sciences and Engineering Research Council of Canada. Goh acknowledges support from the Connaught Fund of the University of Toronto. Both authors are grateful to Roger Koenker for his encouragement. Correspondence may be sent to either author.

[1]Cf. esp. Koenker and Bassett (1978).

We note that in the case of iid or stationary samples, the asymptotic behaviour of normalized spacings of *order statistics* is well-known.[2]

2 Preliminaries

In order to study the asymptotic behaviour of regression quantile spacings, we define the "localized" objective function

$$Z_n(\boldsymbol{u}) \equiv n \sum_{i=1}^{n} \left\{ \rho_{\alpha+\frac{m}{n}} \left(Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_n(\alpha) - \frac{1}{n} \boldsymbol{x}_i^T \boldsymbol{u} \right) - \rho_{\alpha+\frac{m}{n}} \left(Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_n(\alpha) \right) \right\}, \qquad (2)$$

where $\hat{\boldsymbol{\beta}}_n(\alpha)$ is the regression α-quantile estimator minimizing (2). Note that $Z_n(\cdot)$ is minimized at $D_n(m) \equiv n \left\{ \hat{\boldsymbol{\beta}}_n \left(\alpha + \frac{m}{n} \right) - \hat{\boldsymbol{\beta}}_n(\alpha) \right\}$. The limiting behaviour of the sequence $\{D_n(m)\}$ is determined from the limiting behaviour of the sequence of objective functions $\{Z_n\}$; note that in particular, that since $Z_n(\cdot)$ is convex, the finite-dimensional weak convergence of $\{Z_n\}$ to Z implies that $D_n(m) \xrightarrow{d} \arg\min(Z)$.

Define residuals $\hat{\epsilon}_i \equiv Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_n(\alpha)$ for $i = 1, \ldots, n$. Under mild assumptions (cf. e.g., Koenker, 2005, §2.2.1), exactly d residuals will be exactly zero and $\hat{\boldsymbol{\beta}}_n(\alpha)$ is determined by the observations i where $\hat{\epsilon}_i = 0$. Define

$$\mathcal{H}_n \equiv \{ i : \hat{\epsilon}_i = 0 \} ;$$

then $\hat{\boldsymbol{\beta}}_n(\alpha)$ satisfies

$$\sum_{i \notin \mathcal{H}_n} \psi_\alpha(\hat{\epsilon}_i) \boldsymbol{x}_i = \sum_{i \in \mathcal{H}_n} \tau_i \boldsymbol{x}_i,$$

where for $i \in \mathcal{H}_n$, τ_i is a random variable taking values in $[-\alpha, 1-\alpha]$. The asymptotic behaviour of Z_n in (2) depends crucially in very different ways on whether the observations $(\boldsymbol{x}_i^T, Y_i)$ have $i \in \mathcal{H}_n$ or $i \notin \mathcal{H}_n$.

We will start by studying the behaviour of $\{(\boldsymbol{x}_i^T, \tau_i) : i \in \mathcal{H}_n\}$. Lemma 1 below describes the asymptotic behaviour of this sequence under certain conditions, which are described below.

Assumption 1. $\epsilon_i \equiv \epsilon_i(\alpha) = Y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}(\alpha)$ *are iid random variables with density* f_α *continuous at* $x = 0$ *with* $f_\alpha(0) > 0$.

[2] Cf. e.g., Siddiqui (1960); Pyke (1965); David and Nagaraja (2003).

Assumption 2. *For some non-lattice probability measure μ,*

$$\frac{1}{n}\sum_{i=1}^{n} 1(\boldsymbol{x}_i \in B) \to \mu(B),$$

for all sets B with $\mu(\partial B) = 0$.

Assumption 3. $\max_i \boldsymbol{x}_i^T \boldsymbol{x}_i = o(n)$ *and*

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \to \boldsymbol{C} \equiv \int \boldsymbol{x}\boldsymbol{x}^T \mu(d\boldsymbol{x}).$$

Note that Assumptions 1–3 imply the asymptotic normality of the regression α-quantile. To wit:

$$\sqrt{n}\left\{\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right\} \xrightarrow{d} \boldsymbol{W} \sim N\left(0, \frac{\alpha(1-\alpha)}{f_\alpha^2(0)}\boldsymbol{C}^{-1}\right).$$

Lemma 1. *Under Assumptions 1–3,*

$$\left\{(\boldsymbol{x}_i^T, \tau_i) : i \in \mathcal{H}_n\right\} \xrightarrow{d} \left\{(\boldsymbol{\mathcal{X}}_1^T, \mathcal{T}_1), \ldots, (\boldsymbol{\mathcal{X}}_d^T, \mathcal{T}_d)\right\},$$

where $(\mathcal{T}_1, \ldots, \mathcal{T}_d)^T$ are independent Uniform random variables on the interval $[-\alpha, 1-\alpha]$, and $(\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_d)$ have a joint distribution given by

$$\frac{\left|(\boldsymbol{x}^1 \cdots \boldsymbol{x}^d)\right|^2}{d!\,|C|}\mu(d\boldsymbol{x}^1)\cdots\mu(d\boldsymbol{x}^d),$$

where $(\boldsymbol{x}^1 \cdots \boldsymbol{x}^d)$ is the matrix whose jth column is $\boldsymbol{x}^j$. The random vector $(\mathcal{T}_1, \ldots, \mathcal{T}_d)^T$ is independent of $(\boldsymbol{\mathcal{X}}^1, \ldots, \boldsymbol{\mathcal{X}}^d)^T$.

Proof. The proof is simply an elaboration of the proof of asymptotic normality given in Bassett and Koenker (1978). In particular, let $\boldsymbol{W}$ be the limit of $\boldsymbol{W}_n \equiv \sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right)$; we will show that the joint density of $\left(\boldsymbol{\mathcal{X}}_1^T, \ldots, \boldsymbol{\mathcal{X}}_d^T, \mathcal{T}_1, \ldots, \mathcal{T}_d, \boldsymbol{W}^T\right)^T$ is

$$g(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d, \boldsymbol{\tau}^T, \boldsymbol{w}^T) = \frac{\left|(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d)\right|^2}{d!\,|\boldsymbol{C}|^{\frac{1}{2}}}\left\{\frac{f_\epsilon^2(0)}{2\pi\alpha(1-\alpha)}\right\}^{\frac{d}{2}} \exp\left(-\frac{f_\epsilon^2(0)\boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w}}{2\alpha(1-\alpha)}\right)$$

with respect to the dominating measure

$$\mu(d\boldsymbol{x}^1) \times \cdots \times \mu(d\boldsymbol{x}^d)\lambda(d\boldsymbol{\tau}^T)\lambda(d\boldsymbol{w})1\left\{\boldsymbol{\tau} \in (-\alpha, 1-\alpha)^d\right\},$$

where λ denotes Lebesgue measure. As such, the joint density of $\left(\mathcal{X}_1^T, \ldots, \mathcal{X}_d^T, \mathcal{T}_1, \ldots, \mathcal{T}_d\right)^T$ is

$$g(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d, \boldsymbol{\tau}^T) = \frac{\left|(\boldsymbol{x}^1 \cdots \boldsymbol{x}^d)\right|^2}{d!\,|\boldsymbol{C}|}.$$

Let B_1, B_2 and B_3 be subsets of $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$, $(-\alpha, 1-\alpha)^d$ and $\mathbb{R}^d$ respectively. Then

$$\begin{aligned} & \Pr\left[\{\boldsymbol{x}_i : i \in \mathcal{H}_n\} \in B_1, \{\tau_i : i \in \mathcal{H}_n\} \in B_2, \boldsymbol{W} \in B_3\right] \\ = \; & n^{-\frac{d}{2}} \sum_{\boldsymbol{\Omega}_H \in B_1} \int_{B_3} |\boldsymbol{\Omega}_H| \prod_{i \in H} f_\alpha\left(n^{-\frac{1}{2}} \boldsymbol{x}_i^T \boldsymbol{w}\right) P\left\{\boldsymbol{V}_n(\boldsymbol{w}, \boldsymbol{\Omega}_H) \in B_2\right\} \lambda(d\boldsymbol{w}), \end{aligned}$$

where H is a subset of d elements from $\{1, \ldots, n\}$, $\boldsymbol{\Omega}_H$ is the $d \times d$ matrix whose columns are $\{\boldsymbol{x}_i : i \in H\}$ and

$$\boldsymbol{V}_n\left(\boldsymbol{w}, \boldsymbol{\Omega}_H\right) = \sum_{i \notin H} \psi_\alpha\left(\epsilon_i - n^{-\frac{1}{2}} \boldsymbol{x}_i^T \boldsymbol{w}\right) \boldsymbol{\Omega}_H^{-1} \boldsymbol{x}_i.$$

The key element of the proof is provided by the results of Mukhin (1985, Theorem 1). Specifically, we have for $B_2 \subset (-\alpha, 1-\alpha)^d$ that

$$\begin{aligned} & n^{\frac{d}{2}} P\left\{\boldsymbol{V}_n(\boldsymbol{w}, \boldsymbol{\Omega}_H) \in B_2\right\} \\ = \; & \lambda(B_2) \frac{|\boldsymbol{\Omega}_H|}{\{2\pi\alpha(1-\alpha)\}^{\frac{d}{2}} |\boldsymbol{C}|^{\frac{1}{2}}} \exp\left(-\frac{f_\alpha^2(0) \boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w}}{2\alpha(1-\alpha)}\right) + o(1), \end{aligned}$$

and thus we have

$$\begin{aligned} & P\left[\{\boldsymbol{x}_i : i \in \mathcal{H}_n\} \in B_1, \{\tau_i : i \in \mathcal{H}_n\} \in B_2, \boldsymbol{W}_n \in B_3\right] \\ = \; & \lambda(B_2) \sum_{\boldsymbol{\Omega}_H \in B_1} n^{-p} \int_{B_3} \frac{|\boldsymbol{\Omega}_H|^2 f_\alpha^d(0)}{\{2\pi\alpha(1-\alpha)\}^{\frac{d}{2}} |\boldsymbol{C}|^{\frac{1}{2}}} \exp\left(-\frac{f_\alpha^2(0) \boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w}}{2\alpha(1-\alpha)}\right) \lambda(d\boldsymbol{w}) + o(1). \end{aligned}$$

The conclusion follows by noting that

$$n^{-p} \sum_{\boldsymbol{\Omega}_H \in B_1} |\boldsymbol{\Omega}_H|^2 \to \frac{1}{d!} \int_{B_1} \left|(\boldsymbol{x}^1 \cdots \boldsymbol{x}^d)\right|^2 \mu(d\boldsymbol{x}^1) \cdots \mu(d\boldsymbol{x}^d)$$

for any set B_1. $\square$

Note that Lemma 1 implies that $\{\boldsymbol{x}_i : i \in \mathcal{H}_n\}$ are asymptotically drawn from a biased version of the d-fold product of the limiting design measure μ, with samples having higher dispersion (as measured by the determinant of the resulting matrix) being preferred. For example, when $d = 2$ with $\boldsymbol{x} \equiv (1, x)^T$, the limiting measure of the non-intercept component is

$$\frac{|x_1 - x_2|^2}{2\sigma_\mu^2} \mu(dx_1)\mu(dx_2),$$

where σ_μ^2 is the variance of the measure μ. It is worth noting that this result depends on the assumption that $f_\alpha(0) > 0$. In general, the limiting distribution of $\{\boldsymbol{x}_i : i \in \mathcal{H}_n\}$ depends on the behaviour of the distribution function of $Y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}(\alpha)$ near zero; see the Appendix for details.

For observations outside of $\mathcal{H}_n$, we will define the point processes

$$M_n(A \times B) \equiv \sum_{i \notin \mathcal{H}_n} 1\{n\hat{\epsilon}_i \in A, \boldsymbol{x}_i \in B\}. \tag{3}$$

Under the same conditions as in Lemma 1, $\{M_n\}$ will converge in distribution to a Poisson process.

Lemma 2. *Under Assumptions1–3, $\{M_n\}$ converges in distribution (with respect to the vague topology) to a Poisson process M with mean measure*

$$m(A \times B) \equiv f_\alpha(0)\lambda(A)\mu(B),$$

where λ denotes Lebesgue measure.

Proof. We start by defining the point process

$$M_n^*(A \times B) \equiv \sum_{i=1}^{n} 1\{n\epsilon_i \in A, \boldsymbol{x} \in B\},$$

which will in turn converge in distribution to the Poisson process M. We can write

$$M_n(A \times B) = \sum_{i \notin \mathcal{H}_n} 1\left\{n\epsilon_i - \boldsymbol{x}_i^T \boldsymbol{\Omega}_n^{-1} \boldsymbol{\xi}_n \in A, \boldsymbol{x}_i \in B\right\},$$

where $\boldsymbol{\Omega}_n$ is a matrix whose columns are denoted by $\boldsymbol{x}_i$ for $i \in \mathcal{H}_n$ and $\boldsymbol{\xi}_n$ is a vector with components $n\epsilon_i$ for $i \in \mathcal{H}_n$; the conclusion follows by arguing conditionally given $\boldsymbol{\Omega}_n^{-1}\boldsymbol{\xi}_n$ and noting that $\boldsymbol{\Omega}_n^{-1}\boldsymbol{\xi}_n = O_p\left(n^{\frac{1}{2}}\right)$.

The points of the limiting Poisson process M can be represented as $\{(\Gamma_k, \boldsymbol{X}_k) : k \neq 0\}$, where $\Gamma_k = E_1 + \cdots + E_k$ and $\Gamma_{-k} = -(E_{-1} + \cdots E_{-k})$ for an iid sequence $\{E_i\}$ of exponential random variables with mean $\frac{1}{f_\alpha(0)}$, and $\{\boldsymbol{X}_k\}$ is an iid sequence with distribution μ; the two iid sequences are mutually independent. □

3 Main Results

Theorem 1. *Suppose that Assumptions 1–3 hold. Let* $\gamma \equiv \int \boldsymbol{x}\mu(d\boldsymbol{x})$. *Then* $Z_n \stackrel{f-d}{\to} Z$, *where*

$$Z(\boldsymbol{u}) \equiv -m\boldsymbol{u}^T\boldsymbol{\gamma} - \sum_{j=1}^{d} \mathcal{T}_j \boldsymbol{u}^T \mathcal{X}_j + \sum_{j=1}^{d} \rho_{1-\alpha}\left(\boldsymbol{u}^T\mathcal{X}_j\right) + \sum_{k\neq 0} \int_0^{\boldsymbol{X}_k^T\boldsymbol{u}} \left\{1(\Gamma_k \le s) - 1(\Gamma_k < 0)\right\} ds.$$

Thus $n\left\{\hat{\boldsymbol{\beta}}_n\left(\alpha + n^{-1}m\right) - \hat{\boldsymbol{\beta}}_n\left(\alpha\right)\right\} \stackrel{d}{\to} argmin(Z)$.

Proof. Using the identity

$$\begin{aligned}\rho_{\alpha+n^{-1}m}(u-t) - \rho_{\alpha+n^{-1}m}(u) &= -t\left\{\alpha + \frac{m}{n} - 1(u<0)\right\} 1(u \neq 0) + \rho_{\alpha+n^{-1}m}(-t)1(u=0) \\ &\quad + \int_0^t \left\{1(u \le s) - 1(u<0)\right\} ds,\end{aligned}$$

it follows that for each $\boldsymbol{u}$,

$$\begin{aligned}Z_n(\boldsymbol{u}) &= -\frac{m}{n}\sum_{i\notin\mathcal{H}_n} \boldsymbol{x}_i^T\boldsymbol{u} - \sum_{i\in\mathcal{H}_n} \tau_i \boldsymbol{x}_i^T\boldsymbol{u} + \sum_{i\in\mathcal{H}_n} \rho_{\alpha+n^{-1}m}(-\boldsymbol{x}_i^T\boldsymbol{u}) \\ &\quad + \sum_{i=1}^{n} \int_0^{\boldsymbol{x}_i^T\boldsymbol{u}} \left\{1(n\hat{\epsilon}_i \le s) - 1(n\hat{\epsilon}_i < 0)\right\} ds \qquad (4) \\ &\stackrel{d}{\to} -m\int \boldsymbol{u}^T\boldsymbol{x}\mu(d\boldsymbol{x}) - \sum_{j=1}^{d} \mathcal{T}_j \boldsymbol{u}^T \boldsymbol{\mathcal{X}}_j + \sum_{j=1}^{d} \rho_{1-\alpha}(\boldsymbol{u}^T\boldsymbol{\mathcal{X}}_j) \\ &\quad + \sum_{k\neq 0} \int_0^{\boldsymbol{X}_k^T\boldsymbol{u}} \left\{1(\Gamma_k \le s) - 1(\Gamma_k < 0)\right\} ds. \qquad (5)\end{aligned}$$

Here point process convergence is used to get from (4) to (5), where the finite-dimensional weak convergence follows trivially. □

For sufficiently small m, the distribution of $argmin(Z)$ will have positive probability mass at $\mathbf{0}$. The convergence in Theorem 1 gives us

$$\limsup_{n\to\infty} P\left\{\hat{\boldsymbol{\beta}}_n(\alpha + n^{-1}m) = \hat{\boldsymbol{\beta}}_n(\alpha)\right\} \le P\left\{argmin(Z) = \mathbf{0}\right\}.$$

It is however possible to obtain the equality above directly by considering only the relevant part of the objective function Z_n.

Note that $\hat{\boldsymbol{\beta}}_n(\alpha + n^{-1}m) = \hat{\boldsymbol{\beta}}_n(\alpha)$ if and only if $\mathbf{0} \in \partial Z_n(\mathbf{0})$; if we define (as before) $\boldsymbol{\Omega}_n$ to be the matrix whose columns are $\boldsymbol{x}_i$ for $i \in \mathcal{H}_n$ and $\boldsymbol{\tau}_n$ to be the vector of $\{\tau_i : i \in \mathcal{H}_n\}$, then

$$P\left\{\hat{\boldsymbol{\beta}}_n(\alpha + n^{-1}m) = \hat{\boldsymbol{\beta}}_n(\alpha)\right\} = P\left\{m\boldsymbol{\Omega}_n^{-1}\left(\frac{1}{n}\sum_{i\notin\mathcal{H}_n}\boldsymbol{x}_i\right) + \boldsymbol{\tau}_n \in [-\alpha - n^{-1}m, 1-\alpha-n^{-1}m]^d\right\}.$$

However,

$$m\boldsymbol{\Omega}_n^{-1}\left(\frac{1}{n}\sum_{i\notin\mathcal{H}_n}\boldsymbol{x}_i\right) + \boldsymbol{\tau}_n \xrightarrow{d} m\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} + \boldsymbol{T},$$

where $\boldsymbol{T} = (\mathcal{T}_1, \ldots, \mathcal{T}_d)^T$, and $\boldsymbol{\Omega}$ is a matrix with columns $\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_d$. Since the limiting random variable has a continuous distribution, we have

$$\lim_{n\to\infty} P\left\{\hat{\boldsymbol{\beta}}_n(\alpha + n^{-1}m) = \hat{\boldsymbol{\beta}}_n(\alpha)\right\} = P\left\{m\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} + \boldsymbol{T} \in [-\alpha, 1-\alpha]^d\right\}.$$

Using a similar argument, we obtain

$$P\left\{argmin(Z) = \mathbf{0}\right\} = P\left\{m\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} + \boldsymbol{T} \in [-\alpha, 1-\alpha]^d\right\}.$$

This limiting distribution does not depend on $\alpha \in (0,1)$. Note that the right-hand side above can be written as

$$P\left\{m\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} + \boldsymbol{T} + \alpha\boldsymbol{\iota} \in [0,1]^d\right\},$$

where $\boldsymbol{\iota}$ is a vector of ones; the elements of $\boldsymbol{T} + \alpha\boldsymbol{\iota}$ are uniformly distributed on $[0,1]$.

It is also straightforward to derive the limiting distribution of the endpoints of the interval on which $\hat{\boldsymbol{\beta}}_n(\alpha + n^{-1}m) = \hat{\boldsymbol{\beta}}_n(\alpha)$, and hence the limiting distribution of the length of this interval $R_n(\alpha)$. Following Portnoy (1991), we can also define the number of jumps in the interval $[0,1]$ by

$$J_n = \int_0^1 \frac{1}{R_n(\alpha)} d\alpha.$$

Define

$$\begin{aligned}
\Lambda_n(\alpha) &\equiv \inf\left\{m < 0 : \hat{\boldsymbol{\beta}}_n(\alpha + n^{-1}m) = \hat{\boldsymbol{\beta}}_n(\alpha)\right\} \\
\Psi_n(\alpha) &\equiv \sup\left\{m > 0 : \hat{\boldsymbol{\beta}}_n(\alpha + n^{-1}m) = \hat{\boldsymbol{\beta}}_n(\alpha)\right\}.
\end{aligned}$$

From the arguments given above, it follows that $\{\Lambda_n(\alpha), \Psi_n(\alpha)\} \xrightarrow{d} \{\Lambda, \Psi\}$, where for $s, t > 0$,

$$P\{\Lambda \le -s, \Psi \ge t\} = P\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} \in \left\{[0, t^{-1}]^d - t^{-1}\boldsymbol{A}'\right\} \bigcap \left\{[-s^{-1}, 0]^d + s^{-1}\boldsymbol{A}'\right\}\right],$$

where $\boldsymbol{A}' \equiv \boldsymbol{A} + \alpha\boldsymbol{\iota}$ is a vector of independent uniform random variables on $[0, 1]$.

In the case where $d = 2$, it is possible to obtain reasonably simple closed-form representations of the limiting distributions considered above. Write $\boldsymbol{\mathcal{X}}_j \equiv (1, X_j)^T$ for $j = 1, 2$ and without loss of generality, assume that $\boldsymbol{\gamma} \equiv (1, 0)^T$, if necessary centring X_1 and X_2 to have mean zero). Then

$$\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma} = \frac{1}{X_2 - X_1}\begin{pmatrix} X_2 \\ -X_1 \end{pmatrix}.$$

Setting $V \equiv \frac{X_2}{X_2 - X_1}$, we get the following:

1. For $0 < V < 1$,

$$\begin{aligned} \Psi &= \min\left(\frac{1-\mathcal{T}_1}{V}, \frac{1-\mathcal{T}_2}{1-V}\right) \\ \Lambda &= \max\left(-\frac{\mathcal{T}_1}{V}, -\frac{\mathcal{T}_2}{1-V}\right); \end{aligned}$$

2. for $V > 1$,

$$\begin{aligned} \Psi &= \min\left(\frac{1-\mathcal{T}_1}{V}, -\frac{\mathcal{T}_2}{1-V}\right) \\ \Lambda &= \max\left(-\frac{\mathcal{T}_1}{V}, \frac{1-\mathcal{T}_2}{1-V}\right); \end{aligned}$$

3. for $-1 < V < 0$,

$$\begin{aligned} \Psi &= \min\left(-\frac{\mathcal{T}_1}{V}, -\frac{\mathcal{T}_2}{1-V}\right) \\ \Lambda &= \max\left(\frac{1-\mathcal{T}_1}{V}, \frac{1-\mathcal{T}_2}{1-V}\right); \end{aligned}$$

4. for $V < -1$,

$$\begin{aligned} \Psi &= \min\left(-\frac{\mathcal{T}_1}{V}, \frac{1-\mathcal{T}_2}{1-V}\right) \\ \Lambda &= \max\left(\frac{1-\mathcal{T}_1}{V}, -\frac{\mathcal{T}_2}{1-V}\right). \end{aligned}$$

A The General Case

A.1 Non-regular error distributions

Define F_ϵ to be the distribution function of the errors $\epsilon_i \equiv Y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}(\alpha)$. The assumption that the density $f_\epsilon(0) > 0$ implies that $F_\alpha(t)$ is approximately linear in a neighbourhood of zero. In particular, $F_\alpha(t) \approx f_\alpha(0)t$ for small t. This assumption can be generalized to replace the linear function by an increasing function, which we now demonstrate.

Assume that F_α and f_α satisfy

$$\sqrt{n}\left\{F_\alpha(a_n^{-1}t) - \alpha\right\} \to \eta(t),$$

and that

$$\frac{\sqrt{n}}{a_n} f_\alpha(a_n^{-1}t) \to \eta'(t),$$

where η is a strictly increasing function that is almost everywhere differentiable; η and η' will satisfy the scaling conditions

$$\begin{aligned} \eta(\lambda t) &= \lambda^a \eta(t) \\ \eta'(\lambda t) &= \lambda^{a-1}\eta'(t) \end{aligned}$$

for any $\lambda > 0$ where $a > 0$. If $a < 1$, then $f_\alpha(t)$ increases to infinity as $t \to 0$, while $f_\alpha(0) = 0$ if $a > 1$. We will also assume that the empirical distribution of $\{\boldsymbol{x}_i\}$ converges weakly to a measure μ and that

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{F_\alpha(a_n^{-1}\boldsymbol{x}_i^T\boldsymbol{w}) - \alpha\right\}\boldsymbol{x}_i &= \frac{1}{n}\sum_{i=1}^{n}\sqrt{n}\left\{F_\alpha(a_n^{-1}\boldsymbol{x}_i^T\boldsymbol{w}) - \alpha\right\}\boldsymbol{x}_i \\ &\to \int \eta(\boldsymbol{x}^T\boldsymbol{w})\boldsymbol{x}\mu(d\boldsymbol{x}) \\ &= \boldsymbol{\zeta}(\boldsymbol{w}) \end{aligned}$$

for each $\boldsymbol{w}$. Note that $\boldsymbol{\zeta}(\boldsymbol{w})$ satisfies the same scaling condition as η, i.e., $\boldsymbol{\zeta}(\lambda\boldsymbol{w}) = \lambda^a\boldsymbol{\zeta}(\boldsymbol{w})$.

Under these conditions, we have $a_n\left\{\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right\} = O_p(1)$; see Knight (1998) for details. The limiting distribution of $\{(\boldsymbol{x}_i^T, \tau_i) : i \in \mathcal{H}_n\}$ can be determined as was done in the proof of Lemma 1. In particular,

$$\begin{aligned} & P\left[\{\boldsymbol{x}_i : i \in \mathcal{H}_n\} \in B_1, \{\tau_i : i \in \mathcal{H}_n\} \in B_2, \boldsymbol{W}_n \in B_3\right] \\ = \;& a_n^{-p}\sum_{\boldsymbol{\Omega}_H \in B_1}\int_{B_3}|\boldsymbol{\Omega}_H|\prod_{i\in H} f_\epsilon(a_n^{-1}\boldsymbol{x}_i^T\boldsymbol{w})P\left\{\boldsymbol{V}_n(\boldsymbol{w},\boldsymbol{\Omega}_H) \in B_2\right\}\lambda(d\boldsymbol{w}), \end{aligned}$$

where now

$$\boldsymbol{V}_n(\boldsymbol{w}, \boldsymbol{\Omega}_H) = \sum_{i \notin H} \psi_\alpha(\epsilon_i - a_n^{-1}\boldsymbol{x}_i^T\boldsymbol{w})\boldsymbol{\Omega}_H^{-1}\boldsymbol{x}_i.$$

Now

$$n^{-\frac{1}{2}}\boldsymbol{V}_n(\boldsymbol{w}, \boldsymbol{\Omega}_H) \xrightarrow{d} N\left(-\boldsymbol{\Omega}_H^{-1}\boldsymbol{\zeta}(\boldsymbol{w}), \alpha(1-\alpha)\Omega_H^{-1}\boldsymbol{C}\boldsymbol{\Omega}_H^{-1}\right),$$

and so applying Mukhin (1985, Theorem 1), we have for $B_2 \subset (-\alpha, 1-\alpha)^d$,

$$\begin{aligned} & n^{\frac{d}{2}} P\left\{\boldsymbol{V}_n(\boldsymbol{w}, \boldsymbol{\Omega}_H) \in B_2\right\} \\ = \quad & \lambda(B_2)\frac{|\boldsymbol{\Omega}_H|}{\{2\pi\alpha(1-\alpha)\}^{\frac{d}{2}}|\boldsymbol{C}|^{\frac{1}{2}}}\exp\left(-\frac{\boldsymbol{\zeta}^T(\boldsymbol{w})\boldsymbol{C}^{-1}\boldsymbol{\zeta}(\boldsymbol{w})}{2\alpha(1-\alpha)}\right) + o(1), \end{aligned}$$

which yields as before the result

$$\{\tau_i : i \in \mathcal{H}_n\} \xrightarrow{d} (\mathcal{T}_1, \ldots, \mathcal{T}_d),$$

where $\mathcal{T}_1, \ldots, \mathcal{T}_d$ are independent uniforms on $[-\alpha, 1-\alpha]$. Furthermore, we also obtain

$$\{\boldsymbol{x}_i : i \in \mathcal{H}_n\} \xrightarrow{d} (\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_d),$$

which are independent of $\mathcal{T}_1, \ldots, \mathcal{T}_d$ and have joint density (with respect to measure $\mu(d\boldsymbol{x}^1)\cdots\mu(d\boldsymbol{x}^d)$) given by

$$h(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d) = \frac{|(\boldsymbol{x}^1 \cdots \boldsymbol{x}^d)|}{d!}\phi(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d),$$

where

$$\begin{aligned} \phi(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d) \quad &= \quad \frac{1}{\{2\pi\alpha(1-\alpha)\}^{\frac{d}{2}}|\boldsymbol{C}|^{\frac{1}{2}}}\int\left\{\prod_{i=1}^d \eta'(\boldsymbol{x}^{iT}\boldsymbol{w})\right\}\exp\left(-\frac{\boldsymbol{\zeta}^T(\boldsymbol{w})\boldsymbol{C}^{-1}\boldsymbol{\zeta}(\boldsymbol{w})}{2\alpha(1-\alpha)}\right)\lambda(d\boldsymbol{w}) \\ &= \quad \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{C}|^{\frac{1}{2}}}\int\left\{\prod_{i=1}^d \eta'(\boldsymbol{x}^{iT}\boldsymbol{w})\right\}\exp\left(-\frac{1}{2}\boldsymbol{\zeta}^T(\boldsymbol{w})\boldsymbol{C}^{-1}\boldsymbol{\zeta}(\boldsymbol{w})\right)\lambda(d\boldsymbol{w}). \end{aligned}$$

The weighting function $\phi(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d)$ essentially smooths the function $\prod_i \eta'(\boldsymbol{x}^{iT}\boldsymbol{w})$ over projections $\boldsymbol{w}$; its behaviour depends on the value of the scaling parameter a. When $a > 1$, $\phi(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d)$ tends to increase as the linear independence amongst $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d$, while for $a < 1$, $\phi(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d)$ tends to increase as $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d$ become more linearly dependent.

A.2 Heteroskedastic errors

Here we suppose that the errors $\epsilon_i = Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}$ are not iid but have densities $f_1, \ldots, f_n$ such that

$$f_i(0) = \kappa(\boldsymbol{x}_i),$$

for some function κ. We will assume that for each $\boldsymbol{w}$,

$$\max_{1\le i\le n} \left| f_i(n^{-\frac{1}{2}}\boldsymbol{x}_i^T\boldsymbol{w}) - \kappa(\boldsymbol{x}_i)\right| \to 0,$$

and that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{F_i(n^{-\frac{1}{2}}\boldsymbol{x}_i^T\boldsymbol{w}) - F_i(0)\right\}\boldsymbol{x}_i &= \frac{1}{n}\sum_{i=1}^{n}\sqrt{n}\left\{F_i(n^{-\frac{1}{2}}\boldsymbol{x}_i^T\boldsymbol{w}) - F_i(0)\right\}\boldsymbol{x}_i\\ &\to \boldsymbol{D}_\kappa\boldsymbol{w},\end{aligned}$$

where

$$\boldsymbol{D}_\kappa = \int \kappa(\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^T\mu(d\boldsymbol{x}) < \infty.$$

As before, we have

$$\begin{aligned}&P\left[\{\boldsymbol{x}_i : i\in\mathcal{H}_n\}\in B_1, \{\tau_i : i\in\mathcal{H}_n\}\in B_2, \boldsymbol{W}_n\in B_3\right]\\ =\ & n^{-\frac{d}{2}}\sum_{\boldsymbol{\Omega}_H\in B_1}\int_{B_3}|\boldsymbol{\Omega}_H|\prod_{i\in H} f_i\left(n^{-\frac{1}{2}}\boldsymbol{x}_i^T\boldsymbol{w}\right)P\left\{\boldsymbol{V}_n(\boldsymbol{w},\boldsymbol{\Omega}_H)\in B_2\right\}\lambda(d\boldsymbol{w}),\end{aligned}$$

where again

$$\boldsymbol{V}_n(\boldsymbol{w},\boldsymbol{\Omega}_H) = \sum_{i\notin H}\psi_\alpha(\epsilon_i - n^{-\frac{1}{2}}\boldsymbol{x}_i^T\boldsymbol{w})\boldsymbol{\Omega}_H^{-1}\boldsymbol{x}_i.$$

Again applying Mukhin (1985, Theorem 1), we get

$$\begin{aligned}&n^{\frac{d}{2}}P\left\{\boldsymbol{V}_n(\boldsymbol{w},\boldsymbol{\Omega}_H)\in B_2\right\}\\ =\ &\lambda(B_2)\frac{|\boldsymbol{\Omega}_H|}{\{2\pi\alpha(1-\alpha)\}^{\frac{d}{2}}|\boldsymbol{C}|^{\frac{1}{2}}}\exp\left(-\frac{\boldsymbol{w}^T\boldsymbol{D}_\kappa\boldsymbol{C}^{-1}\boldsymbol{D}_\kappa\boldsymbol{w}}{2\alpha(1-\alpha)}\right) + o(1).\end{aligned}$$

Thus proceeding as before, it follows that

$$\{\tau_i : i\in\mathcal{H}_n\}\xrightarrow{d}\{\mathcal{T}_1,\ldots,\mathcal{T}_d\},$$

for $\mathcal{T}_1,\ldots,\mathcal{T}_d$ independent uniform random variables on $[-\alpha, 1-\alpha]$, while

$$\{\boldsymbol{x}_i : i\in\mathcal{H}_n\}\xrightarrow{d}(\boldsymbol{\mathcal{X}}_1,\ldots,\boldsymbol{\mathcal{X}}_d),$$

which are again independent of $(\mathcal{T}_1, \ldots, \mathcal{T}_d)$ and have the joint density

$$\begin{aligned} h(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d) &= \frac{\left|(\boldsymbol{x}^1 \cdots \boldsymbol{x}^d\right|^2}{d!|\boldsymbol{D}_\kappa|} \prod_{i=1}^d \kappa(\boldsymbol{x}^i) \\ &= \frac{\left|\left(\kappa^{\frac{1}{2}}(\boldsymbol{x}^1)\boldsymbol{x}^1 \cdots \kappa^{\frac{1}{2}}(\boldsymbol{x}^d)\boldsymbol{x}^d\right)\right|^2}{d!\,|\boldsymbol{D}_\kappa|}. \end{aligned}$$

Thus (this is not surprising!) observations with larger values of $f_i(0)$ will receive higher weight than they would in the case where the sequence $\{\epsilon_i\}$ is iid.

References

Bassett, G., and R. Koenker (1978) 'Asymptotic theory of least absolute error regression.' *Journal of the American Statistical Association* 73, 618–622

David, H. A., and H. N. Nagaraja (2003) *Order Statistics,* third ed. (Hoboken, NJ: Wiley)

Knight, K. (1998) 'Limiting distributions for L_1 regression estimators under general conditions.' *Annals of Statistics* 26, 755–770

Koenker, R. (2005) *Quantile Regression* (New York: Cambridge University Press)

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

Mukhin, M. B. (1985) 'Local limit theorems for distributions of sums of independent random vectors.' *Theory of Probability and its Applications* 29, 369–375

Portnoy, S. (1991) 'Asymptotic behavior of the number of regression quantile breakpoints.' *SIAM Journal on Scientific and Statistical Computing* 12, 867–883

Pyke, R. (1965) 'Spacings.' *Journal of the Royal Statistical Society, Series B* 27, 395–449. With discussion.

Siddiqui, M. M. (1960) 'Distribution of quantiles in samples from a bivariate population.' *Journal of Research of the National Bureau of Standards—B. Mathematics and Mathematical Physics* 64B, 145–150