Keith Knight

Asymptotics of the regression quantile basic solution under misspecification

Applications of Mathematics, Vol. 53 (2008), No. 3, 223–234

Persistent URL: http://dml.cz/dmlcz/140317

ASYMPTOTICS OF THE REGRESSION QUANTILE BASIC SOLUTION UNDER MISSPECIFICATION*

Keith Knight, Toronto

(Invited)

Abstract. We consider the asymptotic distribution of covariate values in the quantile regression basic solution under weak assumptions. A diagnostic procedure for assessing homogeneity of the conditional densities is also proposed.

Keywords: regression quantiles, basic solution, misspecified model

MSC 2010: 62J05, 60F05, 62J20

1. Introduction

Quantile regression is an extension of more classical regression methods that allows estimation of the conditional quantile of a response variable given covariate values. It was developed by Koenker and Bassett [11] and has been widely applied in practice. A comprehensive treatment of quantile regression is given in the monograph by Koenker [10].

Suppose we observe $(\boldsymbol{x}_1, Y_1), \ldots, (\boldsymbol{x}_n, Y_n)$ and assume that the conditional τ quantile of the response Y_i given the predictor $\boldsymbol{x}_i$ is $g_\tau(\boldsymbol{x}_i)$ for an unknown function g_τ. In practice, we often assume a linear form for g_τ so that $g_\tau(\boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}(\tau)$. Henceforth, we will focus on a fixed $\tau \in (0,1)$ and suppress the dependence on τ of $\boldsymbol{\beta}(\tau)$ and its estimators.

The quantile regression estimator $\hat{\boldsymbol{\beta}}_n$ introduced in [11] minimizes

$$\sum_{i=1}^{n} \varrho_\tau(Y_i - \boldsymbol{x}_i^T \boldsymbol{\phi}) \tag{1.1}$$

* The research for this paper was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

where $\varrho_\tau(s) = s\{\tau - I(s < 0)\}$; $\hat{\boldsymbol{\beta}}_n$ can be computed by solving a linear programming problem and the hyperplane $\boldsymbol{x}^T\hat{\boldsymbol{\beta}}_n$ will go through at least $p = \dim(\boldsymbol{x}_i)$ points. Under certain mild conditions on $\{\boldsymbol{x}_i, Y_i\}$, we will have $Y_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_n$ for exactly p points.

The purpose of this paper is to study the asymptotics of the p points $\{\boldsymbol{x}_i\}$ lying in the so-called basic solution under very general conditions, including the case where the conditional quantile function is misspecified, that is, $g_\tau(\boldsymbol{x}) \neq \boldsymbol{x}^T\boldsymbol{\beta}$. We will also propose a diagnostic procedure for "homogeneity" of the conditional density function of the response over $\boldsymbol{x}$.

2. Asymptotics

The quantile regression estimation problem can be expressed as a linear programming problem, and can be solved using linear programming algorithms, for example, the simplex algorithm ([4]) or interior point algorithms ([5], [7]); see [10], [12], and [13] for details on their implementation in quantile regression. Thus the quantile regression solution is potentially subject to degeneracy—when $Y_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_n$ for more than p points—and multiple solutions. However, under appropriate weak conditions on the model, these are not issues.

Under mild regularity conditions on $\{(\boldsymbol{x}_i, Y_i)\}$ which will hold, for example, if $\{\varepsilon_i\}$ are continuous random variables (see [10]), $\hat{\boldsymbol{\beta}}_n$ is determined exactly (with probability 1) by p points $(\boldsymbol{x}_{i_1}, Y_{i_1}), \ldots, (\boldsymbol{x}_{i_p}, Y_{i_p})$ where $\{i_1, \ldots, i_p\}$ is a subset of $\{1, \ldots, n\}$ so that $Y_{i_j} = \boldsymbol{x}_{i_j}^T\hat{\boldsymbol{\beta}}_n$ for $j = 1, \ldots, p$. More precisely, define

$$\mathcal{H}_n = \{i \leqslant n\colon\ Y_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_n\}; \tag{2.1}$$

then $\hat{\boldsymbol{\beta}}_n$ satisfies

$$\sum_{i \notin \mathcal{H}_n} \psi_\tau(Y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_n)\boldsymbol{x}_i = \sum_{i \in \mathcal{H}_n} \alpha_i \boldsymbol{x}_i \tag{2.2}$$

where $\psi_\tau(s) = \tau - I(s < 0)$ and $\alpha_i \in [-\tau, 1-\tau]$ for $i \in \mathcal{H}_n$. The estimator $\hat{\boldsymbol{\beta}}_n$ is a unique minimizer of (1.1) if and only if $\alpha_i \in (-\tau, 1-\tau)$ for $i \in \mathcal{H}_n$. Uniqueness will be ensured if

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i \neq \boldsymbol{0}$$

with $a_i = -\tau$ or $1-\tau$ for $i = 1, \ldots, n$; when n is large, this condition will follow provided that $\{\boldsymbol{x}_i\}$ are not concentrated on a lattice on $\mathbb{R}^p$.

We will assume the following regularity conditions in order to study the asymptotic behaviour of $\{(\boldsymbol{x}_i, \alpha_i)\colon\ i \in \mathcal{H}_n\}$.

(A1) $Y_1, \ldots, Y_n$ are independent random variables with bounded densities $f_1, \ldots, f_n$ such that

$$f_i = f(\cdot; \boldsymbol{x}_i)$$

where for each $\boldsymbol{x}$, $f(\cdot; \boldsymbol{x})$ is continuous.

(A2) For some probability measure μ having a non-lattice component,

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

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

(A3) $\max_i \boldsymbol{x}_i^T \boldsymbol{x}_i = o(n)$ and

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

(A4) For $\varepsilon > 0$ and $|t| \leqslant \varepsilon$, $f(\boldsymbol{x}^T\boldsymbol{\beta}_0 + t; \boldsymbol{x}) \leqslant M_\varepsilon(\boldsymbol{x})$ where

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} M_\varepsilon(\boldsymbol{x}_i) \|\boldsymbol{x}_i\|^2 < \infty$$

and $\boldsymbol{\beta}_0$ is the solution of

$$\iint \psi_\tau(y - \boldsymbol{x}^T\boldsymbol{\beta}_0) \boldsymbol{x} f(y; \boldsymbol{x}) \mu(\mathrm{d}\boldsymbol{x})\, \mathrm{d}y = \mathbf{0}.$$

(A5) For some measure ν and any subset H of p elements from $\{1, \ldots, n\}$, we have

$$\begin{aligned} & n^{p/2} P\Big\{ \sum_{i \notin H} \psi_\tau(Y_i - \boldsymbol{x}_i^T(\boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{w}))\boldsymbol{x}_i \in B \Big\} \\ & \quad = \frac{\nu(B)}{\{2\pi\}^{p/2} |C_\tau|^{1/2}} \exp\Big(-\frac{1}{2} \boldsymbol{w}^T D_\kappa C_\tau^{-1} D_\kappa \boldsymbol{w}\Big) + o(1) \end{aligned}$$

where the $o(1)$ remainder term is uniform in $\boldsymbol{w}$ over compact sets.

Condition (A2) essentially implies that the design behaves asymptotically like an i.i.d. sample from the probability measure μ, even if the design is not random; the asymptotics of the observations in $\mathcal{H}_n$ and its complement depend on the measure μ, albeit in different ways. Note that conditions (A1) and (A2) together imply that the empirical measure Q_n defined by

$$Q_n(A) = \frac{1}{n} \sum_{i=1}^{n} I\{(\boldsymbol{x}_i, Y_i) \in A\}$$

converges weakly (in probability) to a measure Q with $Q(\mathrm{d}\boldsymbol{x}\times\mathrm{d}y) = \mu(\mathrm{d}\boldsymbol{x})f(y;\boldsymbol{x})\,\mathrm{d}y$. Conditions (A2)–(A4) can be modified to accommodate potentially unbounded covariates by introducing a sequence of normalizing matrices $\{\Delta_n\}$ with $\Delta_n^{-1}\boldsymbol{x}_i$ replacing $\boldsymbol{x}_i$; for example, (A2) becomes

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

for sets B with $\mu(\partial B) = 0$. In this case, we are considering the asymptotic behaviour of $\{(\Delta_n^{-1}\boldsymbol{x}_i, \alpha_i)\colon\ i \in \mathcal{H}_n\}$. Condition (A5) is a local limit version of the central limit theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_\tau(Y_i - \boldsymbol{x}_i^T(\boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{w}))\boldsymbol{x}_i \xrightarrow{d} \mathcal{N}(-D_\kappa\boldsymbol{w}, C_\tau),$$

which follows from conditions (A1)–(A4). For more information on conditions under which local limit theorems hold, see [16] and [17]. The typical scenario envisaged is one where the non-intercept component of μ has no lattice components and τ is rational. In this case, ν would be a product of Lebesgue measure on $\mathbb{R}^{p-1}$ and a multiple of counting measure. Practically speaking, restricting τ to be rational is not terribly restrictive—first, the rationals are dense in the interval $[0,1]$ and second, in applications one typically considers rational values of τ.

Under conditions (A1)–(A4), the estimator $\hat{\boldsymbol{\beta}}_n$ converges in probability to $\boldsymbol{\beta}_0$ defined in (A4). Moreover, as in [1], under conditions (A1)–(A4) we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, D_\kappa^{-1}C_\tau D_\kappa^{-1})$$

where

$$D_\kappa = \int \kappa(\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^T\mu(\mathrm{d}\boldsymbol{x}), \tag{2.3}$$

$$\kappa(\boldsymbol{x}) = f(\boldsymbol{x}^T\boldsymbol{\beta}_0;\boldsymbol{x}) \tag{2.4}$$

and

$$C_\tau = \iint \{\tau - I(y < \boldsymbol{x}^T\boldsymbol{\beta}_0)\}^2\boldsymbol{x}\boldsymbol{x}^T f(y;\boldsymbol{x})\,\mathrm{d}y\,\mu(\mathrm{d}\boldsymbol{x}). \tag{2.5}$$

See [1] for more details on misspecified quantile regression models.

Theorem 2.1. *Define $\mathcal{H}_n$ and $\{\alpha_i\}$ as in* (2.1) *and* (2.2). *Then under conditions* (A1)–(A5),

$$\{(\boldsymbol{x}_i, \alpha_i)\colon\ i \in \mathcal{H}_n\} \xrightarrow{d} \{(\mathbf{T}_1, \mathcal{A}_1), \ldots, (\mathbf{T}_p, \mathcal{A}_p)\}$$

where $(\mathbf{T}_1, \ldots, \mathbf{T}_p)$ have a joint distribution given by

$$\frac{|(\boldsymbol{t}_1 \ldots \boldsymbol{t}_p)|^2}{|D_\kappa|} \prod_{j=1}^{p} \{\kappa(\boldsymbol{t}_j)\mu(\mathrm{d}\boldsymbol{t}_j)\}$$

on the ordered set[1] $\mathcal{O} = \{\boldsymbol{t}_1 \leqslant \boldsymbol{t}_2 \leqslant \ldots \leqslant \boldsymbol{t}_p\}$ *where D_κ and $\kappa(\boldsymbol{t})$ are defined in* (2.3) *and* (2.4) *and the conditional distribution of $(\mathcal{A}_1, \ldots, \mathcal{A}_p)$ given $\mathbf{T}_1 = \boldsymbol{t}_1, \ldots, \mathbf{T}_p = \boldsymbol{t}_p$ is*

$$P\{(\mathcal{A}_1, \ldots, \mathcal{A}_p) \in B \mid \mathbf{T}_1 = \boldsymbol{t}_1, \ldots, \mathbf{T}_p = \boldsymbol{t}_p\} = \frac{\nu\{(\boldsymbol{t}_1 \ldots \boldsymbol{t}_p)B\}}{|(\boldsymbol{t}_1 \ldots \boldsymbol{t}_p)|}.$$

P r o o f. The idea behind the proof is similar to the proof of asymptotic normality given in [2]. Defining $\mathbf{W}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$, we will find the limiting joint density of

$$(\{(\boldsymbol{x}_i, \alpha_i)\colon\ i \in \mathcal{H}_n\}, \mathbf{W}_n).$$

Let B_1, B_2, B_3 be subsets of $\mathcal{O}$, $(-\tau, 1-\tau)^p$ and $\mathbb{R}^p$ respectively. Then following [2], we have

$$\begin{aligned} &P[\{\boldsymbol{x}_i\colon\ i \in \mathcal{H}_n\} \in B_1,\ \{\alpha_i\colon\ i \in \mathcal{H}_n\} \in B_2,\ \mathbf{W}_n \in B_3] \\ &= n^{-p/2} \sum_{\Omega_H \in B_1} \int_{B_3} |\Omega_H| \prod_{i \in H} f_i(\boldsymbol{x}_i^T(\boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{w}))P\{\mathbf{V}_n(\boldsymbol{w}, \Omega_H) \in B_2\}\, \lambda(\mathrm{d}\boldsymbol{w}) \end{aligned}$$

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

$$\mathbf{V}_n(\boldsymbol{w}, \Omega_H) = \sum_{i \notin H} \psi_\tau(Y_i - \boldsymbol{x}_i^T(\boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{w}))\Omega_H^{-1}\boldsymbol{x}_i. \tag{2.6}$$

Using the local limit condition (A5) on $\mathbf{V}_n(\boldsymbol{w}, \Omega_H)$ in (2.6), we have $B_2 \subset (-\tau, 1-\tau)^p$,

$$n^{p/2}P\{\mathbf{V}_n(\boldsymbol{w}, \Omega_H) \in B_2\} = \frac{\nu(\Omega_H B_2)}{\{2\pi\}^{p/2}|C_\tau|^{1/2}} \exp\Big(-\frac{1}{2}\boldsymbol{w}^T D_\kappa C_\tau^{-1} D_\kappa \boldsymbol{w}\Big) + o(1)$$

[1] $\boldsymbol{t}_i \leqslant \boldsymbol{t}_j$ if and only if no component of $\boldsymbol{t}_i$ is strictly greater than the corresponding component of $\boldsymbol{t}_j$.

and thus we have

$$P[\{\boldsymbol{x}_i\colon\ i \in \mathcal{H}_n\} \in B_1,\ \{\alpha_i\colon\ i \in \mathcal{H}_n\} \in B_2,\ \mathbf{W}_n \in B_3]$$
$$= \frac{n^{-p}}{(2\pi)^{p/2}} \sum_{\Omega_H \in B_1} \int_{B_3} \frac{|\Omega_H|\nu(\Omega_H B_2)}{|C_\tau|^{1/2}} \prod_{i \in H} \kappa(\boldsymbol{x}_i) \exp\Big(-\frac{1}{2}\boldsymbol{w}^T D_\kappa C_\tau^{-1} D_\kappa \boldsymbol{w}\Big) \lambda(\mathrm{d}\boldsymbol{w})$$
$$+ o(1).$$

The conclusion follows by noting that for any set B_1

$$n^{-p} \sum_{\Omega_H \in B_1} |\Omega_H|^2 \Big\{ \prod_{i \in H} \kappa(\boldsymbol{x}_i) \Big\} \to \int_{B_1} |(\boldsymbol{t}_1 \dots \boldsymbol{t}_p)|^2 \prod_{j=1}^{p} \{\kappa(\boldsymbol{t}_j)\mu(\mathrm{d}\boldsymbol{t}_p)\}$$

and then integrating over $\boldsymbol{w}$. □

Note that the proof of Theorem 2.1 also implies that the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is independent of the distribution of $(\mathbf{T}_1, \dots, \mathbf{T}_p, \mathcal{A}_1, \dots, \mathcal{A}_p)$.

The limiting distribution of $\{\boldsymbol{x}_i\colon\ i \in \mathcal{H}_n\}$ is simply a biased version of the p-fold product of the limiting measure μ. Note that the limiting density (with respect to this p-fold product measure) of $(\mathbf{T}_1, \dots, \mathbf{T}_p)$ can be written as

$$\Phi(\boldsymbol{t}_1, \dots, \boldsymbol{t}_p) = \frac{|(\kappa^{1/2}(\boldsymbol{t}_1)\boldsymbol{t}_1 \dots \kappa^{1/2}(\boldsymbol{t}_p)\boldsymbol{t}_p)|^2}{|D_\kappa|}. \tag{2.7}$$

The determinant in the numerator of (2.7) is a measure of the dispersion of the vectors $\kappa^{1/2}(\boldsymbol{t}_1)\boldsymbol{t}_1, \dots, \kappa^{1/2}(\boldsymbol{t}_p)\boldsymbol{t}_p$.

Example 2.1. Consider the case where $p = 2$ with $\boldsymbol{x}_i = (1, x_i)^T$ and let μ be the limiting measure of $\{x_i\}$. Then the limiting density (with respect to the product measure) of the non-intercept components is

$$\Phi(x_1, x_2) = \frac{\kappa(x_1)\kappa(x_2)(x_1 - x_2)^2}{\sigma_\kappa^2} \quad \text{for } x_1 \leqslant x_2$$

where

$$\sigma_\kappa^2 = E_\mu\{\kappa(X)\}E_\mu\{\kappa(X)X^2\} - E_\mu^2\{\kappa(X)X\}.$$

One quantity of interest is the limiting distribution of the length $R_n = \{|x_i - x_j|\colon\ i, j \in \mathcal{H}_n\}$, the distance between the x values in the basic solution. If μ has a density ψ with respect to the Lebesgue measure then $R_n \xrightarrow{d} R$, where R has density

$$g(r) = \frac{r^2}{\sigma_\kappa^2} \int_{-\infty}^{\infty} \kappa(y)\kappa(y - r)\psi(y)\psi(y - r)\,\mathrm{d}y \quad \text{for } r > 0.$$

If μ is a uniform distribution on $[0,1]$ with $\kappa(x) = x^\gamma$ for $\gamma > -1$ then

$$g(r) = (\gamma+1)(\gamma+3)(\gamma+2)^2 r^2 \int_r^1 y^\gamma (y-r)^\gamma \, \mathrm{d}y \quad \text{for } 0 < r < 1.$$

Fig. 1 shows the densities for $\gamma = -1/2, 0, 1, 5$; R tends to be larger (closer to 1) for negative values of γ and smaller (closer to 0) for positive values of γ.

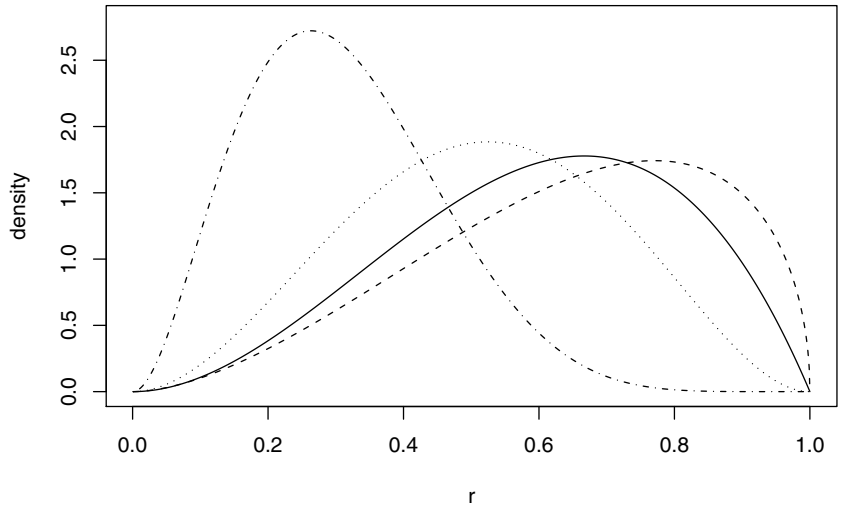

Figure 1. Densities of R in Example 2.1 for $\gamma = -1/2, 0, 1, 5$; as γ increases the mode of the density decreases towards 0 (the dashed line corresponds to $\gamma = -1/2$, the solid line to $\gamma = 0$, the dotted line to $\gamma = 1$).

3. An application

If $\kappa(\boldsymbol{x})$ in (2.4) is constant over $\boldsymbol{x}$ then the limiting distribution of $\{\boldsymbol{x}_i \colon i \in \mathcal{H}_n\}$ depends only on the limiting measure μ. For example, $\kappa(\boldsymbol{x})$ is constant if $Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i$ $(i = 1, \ldots, n)$ where $\{\varepsilon_i\}$ are i.i.d. random variables. More generally, inferential procedures for $\hat{\boldsymbol{\beta}}_n$ are greatly simplified if we are able to assume that $\kappa(\boldsymbol{x})$ is constant.

A simple diagnostic test for the constancy of $\kappa(\boldsymbol{x})$ can be obtained by estimating the distribution of $\{\boldsymbol{x}_i \colon i \in \mathcal{H}_n\}$ using subsampling or, alternatively, bootstrap sampling. Here we will focus on subsampling although the ideas described below will also apply to bootstrap sampling.

Specifically, for some $\alpha \in (0,1)$ we draw subsamples (without replacement) of size $\lfloor \alpha n \rfloor$ (that is, the integer part of αn) from the pairs $\{(\boldsymbol{x}_1, Y_1), \ldots, (\boldsymbol{x}_n, Y_n)\}$.

Define $\mathcal{S}$ to be such a subsample of $\{1,\ldots,n\}$ and define $\hat{\boldsymbol{\beta}}_{\mathcal{S}}$ to minimize

$$\sum_{i\in\mathcal{S}} \varrho_\tau(Y_i - \boldsymbol{x}_i^T\boldsymbol{\phi}).$$

Also define $\mathcal{H}_{\mathcal{S}} = \{i\colon Y_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{\mathcal{S}}\}$. The distribution of $\{\boldsymbol{x}_i\colon i\in\mathcal{H}_{\mathcal{S}}\}$ can be compared to the probability distribution on subsets H of size p from $\{1,\ldots,n\}$ under the assumption that $\kappa(\boldsymbol{x})$ is constant; the proof of Theorem 2.1 suggests that the distribution of $\mathcal{H}_n$ is well-approximated by

$$\mathcal{P}(H) = \frac{|\Omega_H|^2}{\left|\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T\right|} \quad \text{for } H = \{i_1 < i_2 < \ldots < i_p\} \tag{3.1}$$

where Ω_H is the matrix with columns $\{\boldsymbol{x}_i\colon i\in H\}$. In particular, we can compare the distribution of appropriate real-valued (or low dimensional) functions $\phi(\Omega_H)$ under subsampling to the distribution obtained via sampling from $\mathcal{P}$ in (3.1). When $\kappa(\boldsymbol{x})$ is constant, one would expect the two distributions to be similar and, if the function ϕ is chosen appropriately, distinctly different if $\kappa(\boldsymbol{x})$ is not constant. One possible function that is intuitively attractive is the (absolute) determinant of the matrix Ω_H or, more generally, functions of the eigenvalues of $\Omega_H^T\Omega_H$ such as the largest eigenvalue or the trace.

Generating random variates from $\mathcal{P}$ in (3.1) is non-trivial. For small p, it is feasible to sample from $\mathcal{P}$ using rejection sampling since for a given distribution $\mathcal{Q}(H)$ (for example, a uniform distribution on the subsets H), $\sup_H \mathcal{P}(H)/\mathcal{Q}(H)$ can be evaluated analytically. More generally, we can use the upper bound

$$\mathcal{P}(H) = \frac{|\Omega_H|^2}{\left|\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T\right|} \leqslant \left(\frac{1}{p}\sum_{j\in H} h_j\right)^p$$

where $h_1,\ldots,h_n$ are the diagonals of the "hat" matrix:

$$h_j = \boldsymbol{x}_j^T\left(\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}\boldsymbol{x}_j.$$

Since rejection sampling using a uniform distribution for $\mathcal{Q}$ may be quite inefficient, we may wish to sample from $\mathcal{Q}$ that more closely approximates $\mathcal{P}$. One possibility is to sample observations with large h_j with higher probability since these observations are somewhat more likely to be generated by $\mathcal{P}$ than observations with small h_j. For example, we can define

$$\mathcal{Q}(H) \propto \prod_{j\in H} h_j^\delta \tag{3.2}$$

for some $\delta > 0$. Alternatively, we can use a Metropolis-Hastings algorithm ([15], [6]) to sample from $\mathcal{P}$; for example, taking i.i.d. samples from a proposal distribution $\mathcal{Q}$ (for example, such as defined in (3.2)) we can define an independence Metropolis-Hastings algorithm that accepts a proposal H^* generated by $\mathcal{Q}$ given a previous state H_j (and sets $H_{j+1} = H^*$) if

$$U_j < \frac{|\Omega_{H^*}|^2}{|\Omega_{H_j}|^2} \frac{\mathcal{Q}(H_j)}{\mathcal{Q}(H^*)} \tag{3.3}$$

where $\{U_j\}$ is a sequence of i.i.d. uniform random variables on the interval $[0, 1]$ with U_j independent of both H^* and H_j; if the condition (3.3) fails, $H_{j+1} = H_j$.

E x a m p l e 3.1. Values of two predictors $\{(x_{1i}, x_{2i})\}$ for $i = 1, \ldots, 100$ are drawn from a zero mean bivariate normal distribution with covariance matrix

$$C = \begin{pmatrix} 1 & 1 \\ 1 & 1.04 \end{pmatrix}.$$

The two predictors are highly correlated; the sample correlation is 0.970 while the theoretical correlation is 0.962. We will consider three scenarios for the response $\{Y_i\}$:

(a) $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ where $\{\varepsilon_i\}$ are i.i.d. $\mathcal{N}(0, 1)$;

(b) $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ where $\{\varepsilon_i\}$ are independent with $\varepsilon_i \sim \mathcal{N}(0, |x_{1i} - x_{2i}|^2)$;

(c) $Y_i = x_{1i} x_{2i} + \varepsilon_i$ where $\{\varepsilon_i\}$ are i.i.d. $\mathcal{N}(0, 1)$.

For each case, we fit a linear quantile model in the two covariates taking $\tau = 1/2$. The heteroscedasticity given by (b) is not immediately apparent looking at (bivariate) scatterplots of the response versus each of the predictors while the true quantile functions for (c) are non-linear. Figs. 2–4 show the distributions of $\phi(\Omega_H) = |(\boldsymbol{x}_{i_1} \ldots \boldsymbol{x}_{i_3})|$ for half samples (that is, taking subsamples of size 50) compared to its distribution under $\mathcal{P}$. For these data, the distribution of determinants $\phi(\Omega_H)$ based on subsamples from the data generated by model (a) is much closer to the distribution $\mathcal{P}$ than the corresponding distributions based on subsamples from the data generated by (b) and (c). In particular, this method appears to be quite successful in detecting the non-constant $\kappa(\boldsymbol{x})$ for the heteroscedastic data generated by model (b). In particular, the distributions of $\phi(\Omega_H)$ for the data from models (b) and (c) appear to be stochastically smaller than the distribution of $\phi(\Omega_H)$ induced by sampling from $\mathcal{P}$.

Example 3.1 shows that the subsampling method has some potential as a diagnostic for testing the assumption that $\kappa(\boldsymbol{x})$ is constant. A conjecture is that under constancy of $\kappa(\boldsymbol{x})$, the difference (as measured by some appropriate metric for weak convergence) between the distribution under subsampling and the distribution under $\mathcal{P}$ will converge to 0 for subsample sizes $m_n \to \infty$ satisfying $m_n/n \to \alpha \in [0, 1)$.

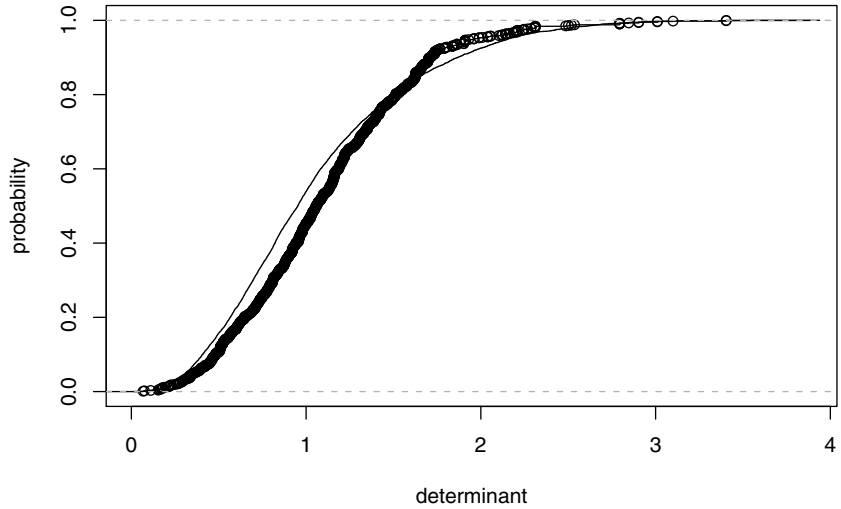

Figure 2. Distribution of $|(\boldsymbol{x}_{i_1} \dots \boldsymbol{x}_{i_p})|$ for $i_1 < \dots < i_p \in \mathcal{H}_S$ compared to the distribution from $\mathcal{P}$ in (3.1) for model (a) in Example 3.1.

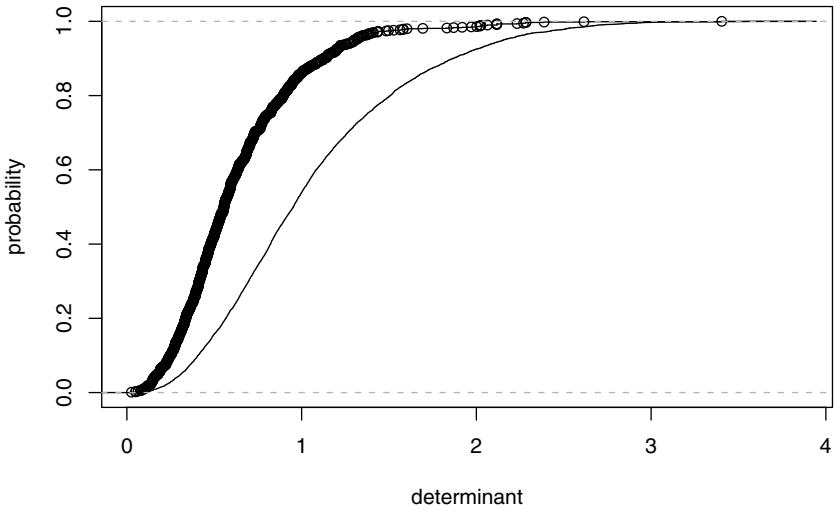

Figure 3. Distribution of $|(\boldsymbol{x}_{i_1} \dots \boldsymbol{x}_{i_p})|$ for $i_1 < \dots < i_p \in \mathcal{H}_S$ compared to the distribution from $\mathcal{P}$ in (3.1) for model (b) in Example 3.1.

In practice, taking the subsampling fraction (whose limit is α) to be small appears to be somewhat better although more research needs to be done here. There are, of course, many other issues that need to be resolved, in particular, "good" choices of $\phi(\Omega_H)$ for distributional comparisons. As with any graphical procedure, some care must be taken in interpreting the plots. However, there are two attractive features of this procedure—first, it is essentially non-parametric in the sense that the reference

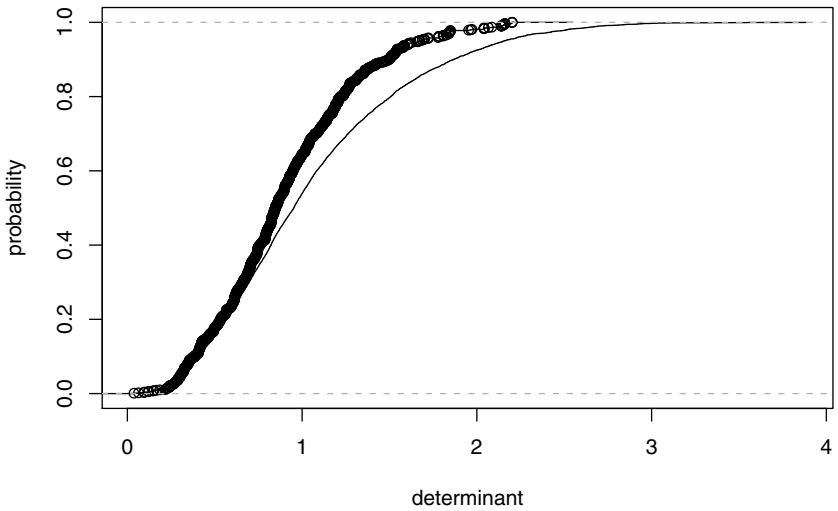

Figure 4. Distribution of $|(\boldsymbol{x}_{i_1} \dots \boldsymbol{x}_{i_p})|$ for $i_1 < \dots < i_p \in \mathcal{H}_S$ compared to distribution from $\mathcal{P}$ in (3.1) for model (c) in Example 3.1.

distribution $\mathcal{P}$ does not depend on any unknown parameters and second, it gives a check of the constancy of $\kappa(\boldsymbol{x})$ for a given quantile τ independent of other quantiles.

Assessing the constancy of $\kappa(\boldsymbol{x})$ for a given quantile τ_0 in a linear quantile regression model is typically by looking at estimates of $\boldsymbol{\beta}(\tau)$ for values of τ in an interval $\mathcal{I}$ that contains τ_0. Assuming (as is almost always the case) that the first element of $\boldsymbol{x}_i$ is 1, we can assess the constancy of $\kappa(\boldsymbol{x})$ for $\tau = \tau_0$ by testing the hypothesis $\beta_j(\tau) = \beta_j$ for $\tau \in \mathcal{I}$ and $j = 2, \dots, p$. Goodness-of-fit tests for the quantile regression process have been proposed by Koenker and Xiao [14] and Chernozhukov and Fernández-Val [3]. Koenker and Xiao [14] use an approach based on the martingale transformation proposed by Khmaladze [8] while [3] uses subsampling to obtain critical values for test statistics.

Alternatively, we could let the width of the interval $\mathcal{I}$ shrink to the point τ_0 as the sample size increases. Recent work by the author and Chuan Goh [9] indicates that the asymptotics of the spacings process $n\{\hat{\boldsymbol{\beta}}_n(\tau_0 + s/n) - \hat{\boldsymbol{\beta}}_n(\tau_0)\}$ depends on the asymptotics of $\{(\boldsymbol{x}_i, \alpha_i)\colon\ i \in \mathcal{H}_n\}$ and of the point process of small but non-zero residuals; the limiting distribution of the point process is a non-homogeneous point process.

Acknowledgement. The author would like to thank Gib Bassett and Roger Koenker for their helpful comments on a draft of this paper.

References

[1] *J. Angrist, V. Chernozhukov, I. Fernández-Val*: Quantile regression under misspecification with an application to the U.S. wage structure. Econometrica *74* (2006), 539–563.

[2] *G. Bassett Jun., R. Koenker*: Asymptotic theory of least absolute error regression. J. Am. Stat. Assoc. *73* (1978), 618–622.

[3] *V. Chernozhukov, I. Fernández-Val*: Subsampling inference on quantile regression processes. Sankhya *67* (2005), 253–276.

[4] *G. B. Dantzig*: Linear Programming and Extensions. Princeton University Press, Princeton, 1951.

[5] *A. V. Fiacco, G. P. McCormick*: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York, 1968.

[6] *W. K. Hastings*: Monte Carlo sampling methods using Markov chains and their applications. Biometrika *57* (1970), 97–109.

[7] *N. Karmarker*: A new polynomial time algorithm for linear programming. Combinatorica *4* (1984), 373–395.

[8] *E. V. Khmaladze*: An innovative approach to goodness-of-fit tests in $\mathbb{R}^m$. Ann. Stat. *16* (1988), 1503–1516.

[9] *K. Knight, C. Goh*: Asymptotics of quantile regression solutions with application. Manuscript in preparation.

[10] *R. Koenker*: Quantile Regression. Cambridge University Press, Cambridge, 2005.

[11] *R. Koenker, G. W. Bassett*: Regression quantiles. Econometrica *46* (1978), 33–50.

[12] *R. Koenker, V. d'Orey*: Computing regression quantiles. Appl. Stat. *36* (1987), 383–393.

[13] *R. Koenker, B. J. Park*: An interior point algorithm for nonlinear quantile regression. J. Econom. *71* (1996), 265–283.

[14] *R. Koenker, Z. Xiao*: Inference on the quantile regression process. Econometrica *70* (2002), 1583–1612.

[15] *N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller*: Equations of state calculations by fast computing machines. J. Chem. Phys. *21* (1953), 1087–1092.

[16] *A. B. Mukhin*: Local limit theorems for distributions of sums of independent random vectors. Theory Probab. Appl. *29* (1985), 369–375.

[17] *A. Mukhin*: Local limit theorems for lattice random variables. Theory Probab. Appl. *36* (1991), 698–713.

Author's address: *K. Knight*, Department of Statistics, University of Toronto, Toronto ON M5S 3G3, Canada, e-mail: `keith@utstat.toronto.edu`.