Econometric Theory, **25**, 2009, 1415–1432. Printed in the United States of America.
doi:10.1017/S0266466609090719

NOTES AND PROBLEMS

NONSTANDARD QUANTILE-REGRESSION INFERENCE

S.C. GOH AND K. KNIGHT
University of Toronto

It is well known that conventional Wald-type inference in the context of quantile regression is complicated by the need to construct estimates of the conditional densities of the response variables at the quantile of interest. This note explores the possibility of circumventing the need to construct conditional density estimates in this context with scale statistics that are explicitly inconsistent for the underlying conditional densities. This method of studentization leads conventional test statistics to have limiting distributions that are nonstandard but have the convenient feature of depending explicitly on the user's choice of smoothing parameter. These limiting distributions depend on the distribution of the conditioning variables but can be straightforwardly approximated by resampling.

1. MOTIVATION

It is well known that the asymptotic variance of quantile estimates depends on the density of the underlying data-generating process evaluated at the quantile of interest. In particular, regression α-quantiles, where $\alpha \in (0, 1)$ denotes the quantile of interest, are known to have asymptotic covariances that depend on the conditional density of the response variables evaluated at their conditional α-quantiles given covariates. As such, Wald-type inference for models of conditional quantiles is complicated by the need to estimate conditional densities of the observations at quantiles of interest. Consistent estimates of these densities lead straightforwardly to inference procedures based on limiting normal or χ^2 distributions. Although asymptotically valid, these inference procedures are known to be sensitive in finite samples to the choice of smoothing parameter used to implement the embedded density estimates.[1]

Goh's research was partially supported by the Connaught Fund of the University of Toronto. Knight's research was supported by the Natural Sciences and Engineering Research Council of Canada. Both authors are grateful to Paolo Paruolo and to two referees for their helpful comments. Address correspondence to S.C. Goh, Department of Economics, University of Toronto, Max Gluskin House, 150 St. George Street, Toronto, Ontario, Canada M5S 3G7; e-mail: goh@economics.utoronto.ca.

This note explores the possibility of studentizing regression quantiles with scale statistics that are explicitly inconsistent for estimating the underlying conditional densities. As might be expected, studentization in this manner leads conventional test statistics to have nonstandard limiting distributions. Although complex, these limiting distributions are shown to depend explicitly on the smoothing parameter used in the studentization. This differs from the situation that emerges from conventional first-order asymptotic theory, where the finite-sample variability of estimates of estimator precision is ignored. Numerical evidence is provided indicating that quantile-regression Wald tests using the nonstandard asymptotic approximation are potentially much more accurate in small samples than Wald tests based on the usual first-order asymptotics.[2] More generally, the nonstandard asymptotic approximation that emerges in this note may be advantageous in terms of its potential application to guiding the choice of smoothing parameter used to studentize coefficient estimates in this context.

A word on notation: The symbols $\Rightarrow$, $\stackrel{f.d.}{\rightarrow}$, and $\stackrel{d}{\rightarrow}$ denote weak convergence, convergence of finite-dimensional distributions, and convergence in distribution, respectively.

2. MAIN RESULTS

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with possibly nonidentical distributions. Suppose the conditional α-quantile function is linear in a d-vector of conditioning variables $\boldsymbol{x}$, i.e.,

$$F^{-1}_{Y_i|\boldsymbol{x}}(\alpha) = \boldsymbol{x}^\top \boldsymbol{\beta}(\alpha).$$

Given a fixed design sequence $\{\boldsymbol{x}_i\}$, write the conditional distributions of each Y_i as

$$P[Y_i < y|\boldsymbol{x}_i] = F_{Y_i}(y|\boldsymbol{x}_i) \equiv F_i(y)$$

and set

$$\xi_i(\alpha) \equiv F_i^{-1}(\alpha).$$

Consider the family of *regression α-quantiles* given by

$$\left\{\hat{\boldsymbol{\beta}}_n(\alpha) : \alpha \in (0,1)\right\},$$

where

$$\hat{\boldsymbol{\beta}}_n(\alpha) \equiv \arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_\alpha(Y_i - \boldsymbol{x}_i^\top \boldsymbol{b}) \qquad \mathbf{(1)}$$

and

$$\rho_\alpha(s) \equiv s[\alpha - 1(s < 0)].$$

In what follows, the following conditions will be assumed to hold.

Assumption 1. The conditional distribution functions $\{F_i\}$ are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$ with continuous density functions $f_i(\cdot) \equiv f(\cdot|\boldsymbol{x}_i)$, where for each $i = 1, 2, \ldots,$ f_i is uniformly bounded away from zero and infinity at $\xi_i(\alpha)$.

In what follows, consider the function $f(\cdot|\boldsymbol{x})$ in the statement of Assumption 1, where in this case $\boldsymbol{x}$ is any d-vector that lies in a measurable space on which the nonlattice probability measure μ referenced in the second part of Assumption 2, which follows, is defined. Define

$$\kappa(\boldsymbol{x}) \equiv f\left(\boldsymbol{x}^\top \boldsymbol{\beta}(\alpha)\middle|\, \boldsymbol{x}\right). \tag{2}$$

Assumption 2. For each $\boldsymbol{b} \in \mathbb{R}^d$ and for $\kappa(\cdot)$ as given in (2), the following conditions hold for a fixed design sequence $\{\boldsymbol{x}_i\}$:

(a)

$$\max_{1\le i\le n}\left| f_i\left(\xi_i(\alpha) + \frac{1}{\sqrt{n}}\boldsymbol{x}_i^\top \boldsymbol{b}\right) - \kappa(\boldsymbol{x}_i)\right| \to 0$$

as $n \to \infty$.

(b)

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}&\left[F_i\left(\xi_i(\alpha) + \frac{1}{\sqrt{n}}\boldsymbol{x}_i^\top \boldsymbol{b}\right) - F_i(\xi_i(\alpha))\right]\boldsymbol{x}_i \\ &= \frac{1}{n}\sum_{i=1}^{n}\sqrt{n}\left[F_i\left(\xi_i(\alpha) + \frac{1}{\sqrt{n}}\boldsymbol{x}_i^\top \boldsymbol{b}\right) - F_i(\xi_i(\alpha))\right]\boldsymbol{x}_i \\ &\to \boldsymbol{D}_1(\alpha)\boldsymbol{b}\end{aligned}$$

as $n \to \infty$, where

$$\boldsymbol{D}_1(\alpha) \equiv \int \kappa(\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^\top \mu(d\boldsymbol{x}) < \infty \tag{3}$$

is positive-definite for some nonlattice probability measure μ.

(c) The matrix

$$\boldsymbol{D}_0 \equiv \int \boldsymbol{x}\boldsymbol{x}^\top \mu(d\boldsymbol{x}) < \infty \tag{4}$$

is positive-definite, where μ is the same nonlattice probability measure used in the definition of $\boldsymbol{D}_1(\alpha)$.

Remark 1. Parts (b) and (c) of Assumption 2 guarantee the continued convexity as $n \to \infty$ of the localized regression α-quantile objective function

$$\tilde{Z}_n(\boldsymbol{u}) \equiv \sum_{i=1}^{n}\left(\rho_\alpha\left(\epsilon_i - \frac{1}{\sqrt{n}}\boldsymbol{x}_i^\top \boldsymbol{u}\right) - \rho_\alpha(\epsilon_i)\right),$$

where $\epsilon_i \equiv Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}(\alpha)$, which in turn guarantees the uniqueness in large samples of the solution to the regression α-quantile minimization problem referenced in (1).[3]

Assumption 3. The fixed design sequence $\{\boldsymbol{x}_i\}$ referred to in Assumption 2 satisfies

$$\max_{1\le i\le n} \frac{1}{\sqrt{n}} \|\boldsymbol{x}_i\| \to 0$$

as $n \to \infty$.

Under Assumptions 1–3, it is straightforward to show[4] that the regression α-quantile is asymptotically normal, with

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right) \xrightarrow{d} N\left(\mathbf{0}, \alpha(1-\alpha)\boldsymbol{D}_1^{-1}(\alpha)\boldsymbol{D}_0\boldsymbol{D}_1^{-1}(\alpha)\right).$$

In what follows, we explore the possibility of studentizing the limiting distribution of $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right)$ using "regression-quantile spacings" of the form $\hat{\boldsymbol{\beta}}_n(\alpha + (m/n)) - \hat{\boldsymbol{\beta}}_n(\alpha)$ for some *fixed* $m > 0$. By way of comparison, note that the conditional response density $f_i(\xi_i(\alpha))$ can be consistently estimated using the difference quotient

$$\hat{f}_{i,m,n}(\xi_i(\alpha)) \equiv \frac{m}{n\boldsymbol{x}_i^\top \left[\hat{\boldsymbol{\beta}}_n\left(\alpha + \frac{m}{n}\right) - \hat{\boldsymbol{\beta}}_n(\alpha)\right]}, \tag{5}$$

where $(m/n) \to 0$ but $m \to \infty$ as $n \to \infty$. This estimate can then be used to construct a natural plug-in estimator of the matrix $\boldsymbol{D}_1(\alpha)$ appearing in the asymptotic covariance of $\hat{\boldsymbol{\beta}}_n(\alpha)$.[5]

To analyze the large-sample behavior of regression-quantile spacings, define the "localized" criterion function

$$Z_n(\boldsymbol{u}) \equiv n\sum_{i=1}^{n}\left[\rho_{\alpha+(m/n)}\left(Y_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_n(\alpha) - \frac{1}{n}\boldsymbol{x}_i^\top\boldsymbol{u}\right)\right.$$
$$\left.-\rho_{\alpha+(m/n)}\left(Y_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_n(\alpha)\right)\right]. \tag{6}$$

Note that $Z_n(\cdot)$ is minimized over $\mathbb{R}^d$ at

$$\boldsymbol{\Delta}_{n,m} \equiv n\left[\hat{\boldsymbol{\beta}}_n\left(\alpha + \frac{m}{n}\right) - \hat{\boldsymbol{\beta}}_n(\alpha)\right].$$

It will be shown that the limiting behavior of $\{\boldsymbol{\Delta}_{n,m}\}$ is recoverable from the limiting behavior of the sequence $\{Z_n(\cdot)\}$ of localized objective functions. In particular, because $\{Z_n(\cdot)\}$ is convex, the convergence of the finite-dimensional distributions of $\{Z_n\}$ to some limiting element Z implies that $\boldsymbol{\Delta}_{n,m} \xrightarrow{d} \boldsymbol{\Delta}_m$, where $\boldsymbol{\Delta}_m \equiv \arg\min Z$.

Now define regression α-quantile residuals

$$\hat{\epsilon}_i(\alpha) \equiv Y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha)$$

for each $i \in \{1, \ldots, n\}$. Under mild assumptions,[6] exactly d residuals will be identically equal to zero. In particular, the value of $\hat{\boldsymbol{\beta}}_n(\alpha)$ is determined only by those observations with indices i such that $\hat{\epsilon}_i(\alpha) = 0$. In this connection, define

$$\mathcal{H}_n(\alpha) \equiv \{i : \hat{\epsilon}_i(\alpha) = 0\}$$

and set

$$\psi_\alpha(s) \equiv \alpha - 1\,(s < 0)\,.$$

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

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

where for each $i \in \mathcal{H}_n(\alpha)$, $\tau_i(\alpha)$ is a random variable taking values in $[-\alpha, 1-\alpha]$.[7] It is shown subsequently that the asymptotic behavior of $Z_n(\cdot)$ depends crucially on whether the observations $(y_i, \boldsymbol{x}_i^\top)^\top$ entering the sum have indices $i \in \mathcal{H}_n(\alpha)$ or $i \notin \mathcal{H}_n(\alpha)$.

We first examine the limiting behavior of $\{(\boldsymbol{x}_i^\top, \tau_i(\alpha))^\top : i \in \mathcal{H}_n(\alpha)\}$. In this connection, we make one further assumption.

Assumption 4. For any Borel set B in $\mathbb{R}^d$ with finite diameter and any subset H of d elements from $\{1, \ldots, n\}$, there is a measure ν such that

$$n^{\frac{d}{2}} P\left[\sum_{i \notin H} \psi_\alpha\left(\epsilon_i - \frac{1}{\sqrt{n}}\boldsymbol{x}_i^\top \boldsymbol{w}\right)\boldsymbol{x}_i \in B\right] = \frac{\nu(B)}{(2\pi)^{d/2}\,|\boldsymbol{D}_0|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{w}^\top \boldsymbol{D}_1(\alpha)\boldsymbol{D}_0^{-1}\boldsymbol{D}_1(\alpha)\boldsymbol{w}\right) + o(1),$$

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

Assumption 4 is a "mixed" local-limit theorem. In particular, it is the local limit version of the central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha\left(\epsilon_i - \frac{1}{\sqrt{n}}\boldsymbol{x}_i^\top \boldsymbol{w}\right)\boldsymbol{x}_i \xrightarrow{d} N(-\boldsymbol{D}_1(\alpha)\boldsymbol{w}, \boldsymbol{D}_0)$$

that follows from Assumptions 1–3. The typical scenario for which Assumption 4 is intended to apply is the case where the nonintercept component of the probability measure μ embedded in the definitions of the matrices $\boldsymbol{D}_1(\alpha)$ and $\boldsymbol{D}_0$ in Assumption 2 does not have a lattice component and where $\alpha \in (0, 1)$ is rational. In this case the measure ν in the statement of Assumption 4 would be the product of Lebesgue measure on $\mathbb{R}^{d-1}$ and some multiple of counting measure.[8]

LEMMA 1. *Under Assumptions 1–4,*

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

where $\{\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_d\}$ *have a joint distribution given by*

$$h(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d) \equiv \frac{\left|\left[\boldsymbol{x}^1 \ldots \boldsymbol{x}^d\right]\right|^2}{|\boldsymbol{D}_1(\alpha)|} \prod_{j=1}^{d} \left[\kappa\left(\boldsymbol{x}^j\right)\mu(d\boldsymbol{x}_j)\right]$$

$$= \frac{\left|\left[\kappa^{1/2}(\boldsymbol{x}^1)\mu\left(\boldsymbol{x}^1\right) \ldots \kappa^{1/2}(\boldsymbol{x}^d)\mu\left(\boldsymbol{x}^d\right)\right]\right|^2}{|\boldsymbol{D}_1(\alpha)|} \tag{7}$$

on the ordered set $\mathcal{O} \equiv \{\boldsymbol{x}_1 \leq \boldsymbol{x}_2 \leq \cdots \leq \boldsymbol{x}_d\}$ *and where the conditional distribution of* $(\mathcal{T}_1(\alpha), \ldots, \mathcal{T}_d(\alpha))^\top$ *given* $\{\boldsymbol{\mathcal{X}}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{\mathcal{X}}_d = \boldsymbol{x}_d\}$ *is given by*

$$P\left[(\mathcal{T}_1(\alpha), \ldots, \mathcal{T}_d(\alpha))^\top \in B \,\middle|\, \boldsymbol{\mathcal{X}}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{\mathcal{X}}_d = \boldsymbol{x}_d\right] = \frac{\nu\left(\left[\boldsymbol{x}_1 \ldots \boldsymbol{x}_d\right] B\right)}{\left|\left[\boldsymbol{x}_1 \ldots \boldsymbol{x}_d\right]\right|}$$

for any $B \subset (-\alpha, 1-\alpha)^d$. *Here the notation* $\left[\boldsymbol{u}^1 \ldots \boldsymbol{u}^d\right]$ *denotes a* $d \times d$ *matrix with* j*th column given by* $\boldsymbol{u}^j$, *and* μ *and* ν *denote the measures specified in the statements of Assumptions 2 and 4, respectively.*

Proof. The proof appears in Appendix A.

Remark 2. Note that in the case of independent and identically distributed (i.i.d.) observations the $\kappa^{1/2}(\boldsymbol{x}^j)$ cancel out of (7). In the more general setting of conditional heteroskedasticity treated here, observations with larger values of $f_i(\xi_i(\alpha))$ will (unsurprisingly) receive a higher density weight.

We note that Lemma 1 indicates that the design points $\{\boldsymbol{x}_i : i \in \mathcal{H}_n(\alpha)\}$ are asymptotically drawn from a biased version of the d-fold product of the limiting design measure μ. In particular, more dispersed samples—where dispersion is measured in terms of the determinant of the design matrix with each observation receiving the density weight $\kappa(\boldsymbol{x}^j)$ ($j = 1, \ldots, d$)—are favored in the limiting experiment. In general, the limiting distribution of the points $\{\boldsymbol{x}_i : i \in \mathcal{H}_n(\alpha)\}$ depends on the behavior of $f_i(u)$ in a neighborhood of the point $u = \xi_i(\alpha)$.[9]

Now consider the limiting behavior of observations with indices $i \notin \mathcal{H}_n(\alpha)$. Define the point process

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

We show that $\{M_n\}$ converges in distribution to a Poisson process.

LEMMA 2. *Under Assumptions 1–3,* $\{M_n\}$ *converges in distribution—with respect to the vague topology—to a Poisson process* M *with mean measure*

$$m\,(d\epsilon, d\boldsymbol{x}) \equiv \lambda(d\epsilon)\mu(d\boldsymbol{x})\kappa(\boldsymbol{x}),$$

where λ *denotes Lebesgue measure.*

Proof. The proof appears in Appendix A.

Note that the points of the limiting Poisson process in Lemma 2 can be represented as $\left\{(\Gamma_k, \boldsymbol{X}_k^\top)^\top : k \neq 0\right\}$, where $\{\boldsymbol{X}_k\}$ is an i.i.d. sequence with distribution μ and $\Gamma_k = E_{1,k} + \cdots + E_{k,k}$, $\Gamma_{-k} = -\left(E_{1,-k} + \cdots + E_{-k,-k}\right)$, where the $\{E_{i,k}\}$ are independent exponentials with mean $1/(\kappa(\boldsymbol{X}_k))$.

Lemmas 1 and 2 together are used to derive the limiting distribution of the regression-quantile spacing.

THEOREM 1. *Suppose that Assumptions 1–4 hold. Let* $\boldsymbol{\gamma} \equiv \int \boldsymbol{x}\mu(d\boldsymbol{x})$. *Then* $Z_n \overset{f.d.}{\to} Z$, *where*

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

As such, the regression-quantile spacing $\boldsymbol{\Delta}_{n,m} \equiv n\left[\hat{\boldsymbol{\beta}}_n(\alpha + (m/n)) - \hat{\boldsymbol{\beta}}_n(\alpha)\right]$ *satisfies*

$$\boldsymbol{\Delta}_{n,m} \overset{d}{\to} \boldsymbol{\Delta}_m,$$

where $\boldsymbol{\Delta}_m \equiv \arg\min Z$.

Proof. The proof appears in Appendix A.

Note that for sufficiently small values of m, the distribution of $\arg\min(Z)$ will have positive probability mass at the origin. In particular, it is immediate from the conclusion of Theorem 1 that

$$\limsup_{n\to\infty} P\left[\hat{\boldsymbol{\beta}}_n\left(\alpha + \frac{m}{n}\right) = \hat{\boldsymbol{\beta}}_n(\alpha)\right] \leq P\left[\arg\min(Z) = \mathbf{0}\right].$$

A consideration of the relevant portion of the objective function $Z_n(\cdot)$, however, yields a closed-form expression for $\lim_{n\to\infty} P\left[\hat{\boldsymbol{\beta}}_n(\alpha + (m/n)) = \hat{\boldsymbol{\beta}}_n(\alpha)\right]$.

In this connection, note that $\hat{\boldsymbol{\beta}}_n(\alpha + (m/n)) = \hat{\boldsymbol{\beta}}_n(\alpha)$ if and only if $\mathbf{0}$ is in fact contained in the set of minimizers of $Z_n(\mathbf{0})$. If, as in the proof of Lemma 2, $\boldsymbol{\Omega}_n$ is taken to denote the $d \times d$ matrix with columns $\boldsymbol{x}_i$ for $i \in \mathcal{H}_n(\alpha)$ and $\boldsymbol{\tau}_n(\alpha)$ to be the d-vector with elements $\{\tau_i(\alpha) : i \in \mathcal{H}_n(\alpha)\}$. Then

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

However,

$$\mathbf{\Omega}_n^{-1}\left(\frac{1}{n}\sum_{i\notin\mathcal{H}_n(\alpha)}\boldsymbol{x}_i\right)+\boldsymbol{\tau}_n(\alpha)\xrightarrow{d} m\mathbf{\Omega}^{-1}\boldsymbol{\gamma}+\boldsymbol{\mathcal{T}}(\alpha), \tag{8}$$

where $\boldsymbol{\mathcal{T}}(\alpha)\equiv(\mathcal{T}_1(\alpha),\ldots,\mathcal{T}_d(\alpha))^\top$ and $\mathbf{\Omega}$ is a $d\times d$ matrix with columns $\boldsymbol{\mathcal{X}}_1,\ldots,\boldsymbol{\mathcal{X}}_d$. Because the limiting random variable in (8) is continuously distributed, we obtain

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

A similar argument produces

$$P\left[\arg\min(Z)=\mathbf{0}\right]=P\left[m\mathbf{\Omega}^{-1}\boldsymbol{\gamma}+\boldsymbol{\mathcal{T}}(\alpha)\in[-\alpha,1-\alpha]^d\right], \tag{9}$$

a probability that does not depend on $\alpha\in(0,1)$, because

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

where $\boldsymbol{\iota}_d$ is a d-vector of ones. In particular, the elements of the random vector $\boldsymbol{\mathcal{T}}(\alpha)+\alpha\boldsymbol{\iota}_d$ are uniformly distributed on [0, 1].[10]

3. DISCUSSION

It should be clear that the conclusion of Theorem 1 can be exploited to derive an alternative "fixed-m" limiting distribution for any pivotal test statistic involving both the regression α-quantile and an associated spacing statistic

$$n\left[\hat{\boldsymbol{\beta}}_n\left(\alpha+\frac{m}{n}\right)-\hat{\boldsymbol{\beta}}_n(\alpha)\right].$$

Equally clear is that the practicality of such a limiting distribution will be complicated by the dependence of the limiting distribution of the spacing statistic on both $\boldsymbol{\gamma}\equiv\int\boldsymbol{x}\mu(d\boldsymbol{x})$ and the random variables $\{\boldsymbol{\mathcal{X}}_1,\ldots,\boldsymbol{\mathcal{X}}_d\}$. In this connection, it is noted that a familiar resampling strategy can be exploited to obtain consistent estimates of the limiting distribution of

$$n\left[\hat{\boldsymbol{\beta}}_n\left(\alpha+\frac{m}{n}\right)-\hat{\boldsymbol{\beta}}_n(\alpha)\right].$$

In particular, consider the behavior of the minimizer of the bootstrap counterpart to the criterion function given in (6), namely,

$$Z_n^*(\boldsymbol{u})\equiv n\sum_{i=1}^n\left[\rho_{\alpha+(m/n)}\left(Y_i^*-\boldsymbol{x}_i^{*\top}\hat{\boldsymbol{\beta}}_n^*(\alpha)-\frac{1}{n}\boldsymbol{x}_i^{*\top}\boldsymbol{u}\right)\right.$$
$$\left.-\rho_{\alpha+(m/n)}\left(Y_i^*-\boldsymbol{x}_i^{*\top}\hat{\boldsymbol{\beta}}_n^*(\alpha)\right)\right],$$

where

$$\left\{(Y_i^*, \boldsymbol{x}_i^{*\top})^\top : i = 1, \ldots, n\right\} \equiv \mathcal{X}_n^*$$

is a random sample from the empirical distribution

$$\hat{F}_n(y, \boldsymbol{x}) \equiv \frac{1}{n} \sum_{i=1}^{n} 1\,(Y_i < y, \boldsymbol{x}_i < \boldsymbol{x})$$

and $\hat{\boldsymbol{\beta}}_n^*(\alpha)$ is the regression α-quantile computed from the elements of the bootstrap sample $\mathcal{X}_n^*$. As such, $Z_n^*(\boldsymbol{u})$ is convex and minimized by $n\left[\hat{\boldsymbol{\beta}}_n^*(\alpha + (m/n)) - \hat{\boldsymbol{\beta}}_n^*(\alpha)\right]$. Arguments analogous to those leading up to Theorem 1 show that

$$n\left[\hat{\boldsymbol{\beta}}_n^*\left(\alpha + \frac{m}{n}\right) - \hat{\boldsymbol{\beta}}_n^*(\alpha)\right] \xrightarrow{d} \arg\min Z,$$

where Z is as given in the statement of Theorem 1. This shows that a standard nonparametric bootstrap technique can be used to approximate the nonstandard limiting distribution of test statistics embedding regression-quantile spacings as inconsistent "estimates" of the asymptotic precision of regression quantiles.[11] A numerical example involving Wald-type statistics studentized according to the difference-quotient method of Hendricks and Koenker (1992) is provided in Section 4.2.

4. NONSTANDARD WALD-TYPE INFERENCE FOR QUANTILE REGRESSION

4.1. Preamble

This section is intended to furnish some evidence on the practicality of Wald-type inference involving the nonstandard method of studentization described in this note. In this connection, for $\alpha \in [\epsilon, 1-\epsilon]$, where $\epsilon \in (0, 1/2)$, set

$$\boldsymbol{W}_n(\alpha) \equiv \sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right), \tag{10}$$

$$\hat{\boldsymbol{D}}_{n,0} \equiv \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top, \tag{11}$$

$$\hat{\boldsymbol{D}}_{n,1,m}(\alpha) \equiv \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \hat{f}_{i,m,n}(\xi_i(\alpha)), \tag{12}$$

where $\hat{f}_{i,m,n}(\xi_i(\alpha))$ is the conditional density "estimator" as given in (5) embedding a regression quantile spacing. Quantities (10)–(12) yield the expression for the Wald statistic used by Hendricks and Koenker (1992):

$$\mathcal{W}_{n,m}(\alpha) \equiv \boldsymbol{W}_n(\alpha)^\top \left[\hat{\boldsymbol{D}}_{n,1,m}^{-1}(\alpha) \hat{\boldsymbol{D}}_{n,0} \hat{\boldsymbol{D}}_{n,1,m}^{-1}(\alpha)\right]^{-1} \boldsymbol{W}_n(\alpha).$$

The limiting distribution of $\mathcal{W}_{n,m}(\alpha)$ is readily available as a corollary of Theorem 1. In this connection, recall the definition

$$\boldsymbol{\Delta}_m(\alpha) \equiv \arg\min_{\boldsymbol{u}} Z(\alpha, \boldsymbol{u}),$$

where $Z(\alpha, \boldsymbol{u})$ is the criterion function as given in (6) written explicitly as a function of α and $\boldsymbol{u}$. The following result is a direct consequence of Theorem 1.

COROLLARY 1. *Let $\boldsymbol{W}(\alpha)$ denote a random variable with distribution given by the limiting distribution of $\boldsymbol{W}_n(\alpha)$ in (10), i.e.,*

$$\boldsymbol{W}(\alpha) \sim N\left(\mathbf{0}, \boldsymbol{D}_1^{-1}(\alpha)\boldsymbol{D}_0\boldsymbol{D}_1^{-1}(\alpha)\right),$$

where $\boldsymbol{D}_1(\alpha)$ and $\boldsymbol{D}_0$ are as given in (3) and (4), respectively. Let

$$\boldsymbol{D}_{1,m}(\alpha) \equiv p\lim \hat{\boldsymbol{D}}_{n,1,m}(\alpha),$$

where $\hat{\boldsymbol{D}}_{n,1,m}(\alpha)$ is as given in (12).

Then under the assumptions of Theorem 1,

$$\mathcal{W}_{n,m}(\alpha) \xrightarrow{d} \boldsymbol{W}(\alpha)^{\top}\left[\boldsymbol{D}_{1,m}^{-1}(\alpha)\boldsymbol{D}_0\boldsymbol{D}_{1,m}^{-1}(\alpha)\right]^{-1}\boldsymbol{W}(\alpha).$$

The conclusion of Corollary 1 can be extended to accommodate inferential questions regarding the *regression-quantile process* $\left\{\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right) : 0 < \alpha < 1\right\}$, which is regarded as a stochastic process in the space $(D[0,1])^d$ of $\mathbb{R}^d$-valued right-continuous functions with left-hand limits on $[0,1]$.[12] In particular, the following result is a direct consequence of Theorem 1 and existing results on the weak convergence of regression quantile processes in $(D[0,1])^d$.[13]

COROLLARY 2. *Suppose the assumptions of Theorem 1 hold for all $\alpha \in (0,1)$. Then for any $\epsilon \in (0,1/2)$,*

$$\sup_{\alpha\in[\epsilon,1-\epsilon]} \mathcal{W}_n(\alpha) \xrightarrow{d} \sup_{\alpha\in[\epsilon,1-\epsilon]} \left\{\boldsymbol{W}^{\top}(\alpha)\left[\boldsymbol{D}_{1,m}^{-1}(\alpha)\boldsymbol{D}_0\boldsymbol{D}_{1,m}^{-1}(\alpha)\right]^{-1}\boldsymbol{W}(\alpha)\right\}, \quad \textbf{(13)}$$

where the quantities in (13) refer to the same quantities given previously in the statement of Corollary 1.

4.2. Numerical Evidence

We conducted a number of modest Monte Carlo experiments to evaluate the performance of the nonstandard "fixed-m" limiting distribution of the Wald statistic $\mathcal{W}_{n,m}(\alpha)$ given in Corollary 1 against the conventional χ^2 asymptotic approximation.

In the first set of experiments conducted, simulated data were drawn from a pure bivariate Gaussian location-shift model given by

$$y_i = x_i + u_i, \qquad i = 1, \ldots, n, \tag{14}$$

where $\{u_i\}$ is i.i.d. $N(0,1)$, the design $\{x_i\}$ is i.i.d. $N(5,1)$ and distributed independently of $\{u_i\}$, and $n \in \{100, 400\}$. The number of replications used was fixed at 500.

Noting that the conditional α-quantile function of y_i given x_i may be expressed as

$$F^{-1}_{y_i|x_i}(\alpha) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2, \tag{15}$$

where $\beta_2 = 0$, the first set of simulation experiments is concerned with the size performance of a 5% Wald test based on the statistic

$$\mathcal{W}_{n,m}\left(\frac{1}{2}\right) = \frac{n\hat{\beta}^2_{n,2}\left(\frac{1}{2}\right)}{s_{n,m,22}\left(\frac{1}{2}\right)} \tag{16}$$

for the hypothesis $H_0 : \beta_2 = 0$, where in this case $\hat{\beta}_{n,2}(1/2)$ denotes the third component of the regression 0.5-quantile estimate of β_2 in (15) and $s_{n,m,22}(1/2)$ denotes the $(3,3)$-element of the matrix

$$\hat{\boldsymbol{D}}^{-1}_{n,1,m}\left(\frac{1}{2}\right)\hat{\boldsymbol{D}}_{n,0}\hat{\boldsymbol{D}}^{-1}_{n,1,m}\left(\frac{1}{2}\right),$$

where $\hat{\boldsymbol{D}}_{n,0}$ and $\hat{\boldsymbol{D}}_{n,1,m}(\cdot)$ are as given in (11) and (12), respectively.

In particular, the accuracy of the conventional χ^2_1 asymptotics for $\mathcal{W}_{n,m}(1/2)$ as given in (16) under H_0 is compared with that obtained by the (full-sample) nonparametric bootstrap approximation to the alternative limiting distribution given in Corollary 1. In this connection, we note that 5% critical values from the "fixed-m" asymptotic distribution of Corollary 1 were approximated by the 0.95-quantile of the empirical distribution of a bootstrap sample of size 99 consisting of simulated values of the statistic

$$\mathcal{W}^*_{n,m}\left(\frac{1}{2}\right) \equiv \frac{n\left[\hat{\beta}^*_{n,2}\left(\frac{1}{2}\right) - \hat{\beta}_{n,2}\left(\frac{1}{2}\right)\right]^2}{s^*_{n,m,22}\left(\frac{1}{2}\right)}, \tag{17}$$

where in this case the starred quantities in (17) refer to the corresponding quantities in (16), but computed from a sample generated by random draws with replacement from the empirical distribution of the original observations $\left\{(y_i, x_i)^\top : i = 1, \ldots, n\right\}$.

The Monte Carlo sizes of the two distributional approximations to the large-

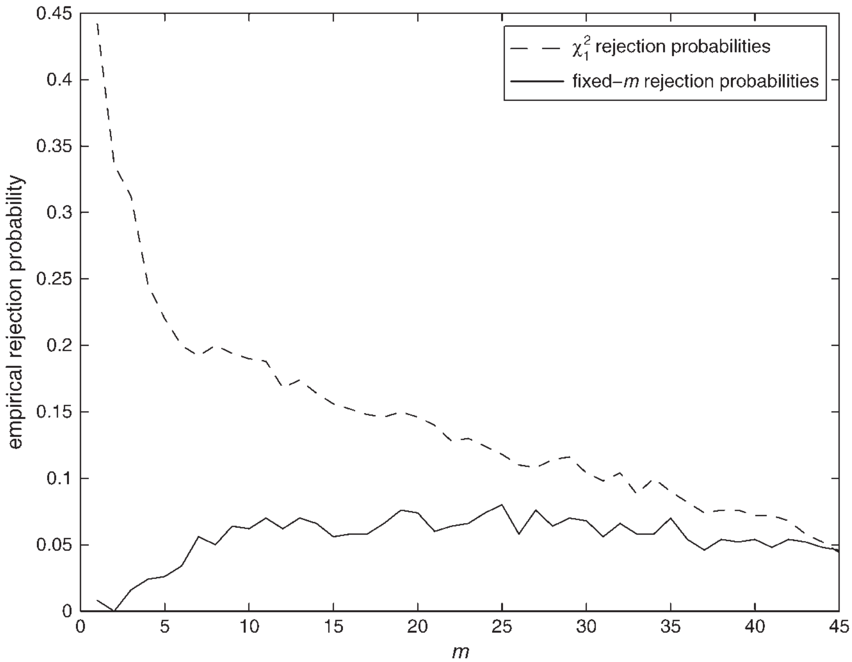

FIGURE 1. Size performance of a 5% Wald test as a function of the smoothing parameter m, $n = 100$.

sample behavior of the Wald statistic under H_0 are displayed in Figures 1 and 2, which plot empirical rejection probabilities as a function of the smoothing parameter m. Figures 1 and 2 report the same experiment repeated at two different sample sizes: $n = 100$ in Figure 1, and $n = 400$ in Figure 2. The same pattern is evident with both sample sizes. In particular, the potential of the conventional asymptotic distribution to lead to substantial overrejection in small samples is evident, whereas the alternative distributional approximation of Corollary 1 seems to be much more effective at controlling the discrepancy between the nominal and empirical sizes of the test. An analytical comparison of the accuracy of the nonstandard distributional approximation given in Corollary 1 with the standard χ^2 asymptotic theory is a topic for further investigation.

The second set of simulation experiments also involves the consideration of a test of the hypothesis $H_0 : \beta_2 = 0$ in the context of the conditional quantile model given in (15). In this case, however, we consider the power performance of the conventional and "fixed-m" Wald testing procedures considered in the first set of experiments against a sequence of alternatives given by

$$H_{1n}(c) : \beta_{2n} = \frac{c}{\sqrt{n}}, \tag{18}$$

where $c > 0$ is a constant. In particular, c is specified to take values in a grid of

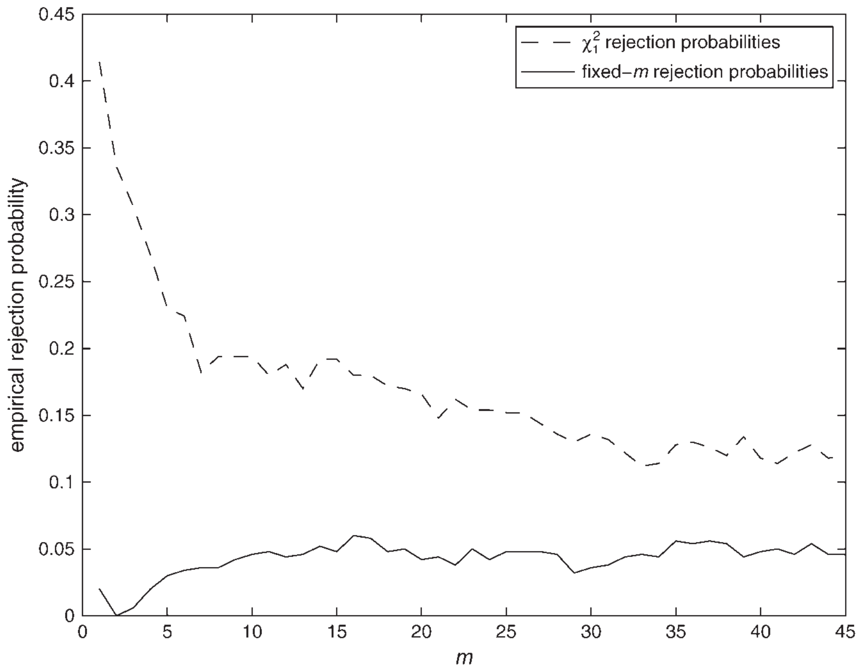

FIGURE 2. Size performance of a 5% Wald test as a function of the smoothing parameter m, $n = 400$.

25 equally spaced points in the interval $[0, 5]$.

In this connection, 25 successive sequences of simulated data were drawn from the "locally location-shift" model given by

$$y_i = x_i + \frac{c}{\sqrt{n}} x_i^2 + u_i, \qquad i = 1, \ldots, n,$$

where as was the case in (14), $\{u_i\}$ is i.i.d. $N(0,1)$ whereas $\{x_i\}$ is i.i.d. $N(5,1)$ and distributed independently of $\{u_i\}$. In this set of experiments, the sample size was fixed at $n = 100$, and 500 replications and a nominal size of 5% were used as before.

Figure 3 summarizes the result of this second set of simulations, which plots empirical rejection probabilities against c, where $c > 0$ indexes each sequence of local alternatives given in (18). As was the case in the first set of simulations, the focus is on empirical rejection probabilities derived from the χ_1^2 and "fixed-m" limiting distributions of the Wald statistic given in (16). The comparison involves power performance associated with critical values derived from the bootstrap approximation to the fixed-m limiting distribution given in Corollary 1 and those obtained using the conventional χ_1^2 asymptotic distribution. Empirical rejection probabilities for fixed-m critical values corresponding

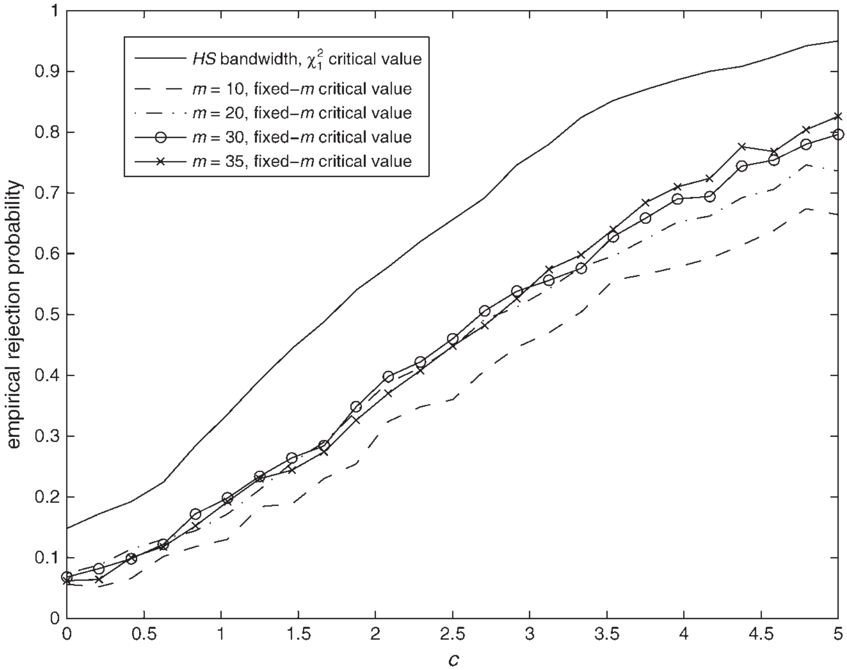

FIGURE 3. Local power performance of a 5% Wald test, $n = 100$.

to $m \in \{10, 20, 30, 35\}$ were investigated, whereas the setting $m = 20$ was used with the standard χ_1^2 critical value. In this connection, it is noted that a value of $m = 20$ is the approximate value of m corresponding to a sample size of $n = 100$ implied by the asymptotically optimal bandwidth sequence proposed by Hall and Sheather (1988) for Wald-type inference regarding a univariate population median and is representative of common empirical practice.[14] The display in Figure 3 repeats the pattern of Figures 1 and 2 in that at $c = 0$, the standard χ_1^2-Wald test implemented using the Hall and Sheather (1988) bandwidth is oversized, whereas the four fixed-m Wald tests have accurate sizes. On the other hand, the nonstandard procedures are conservative relative to the conventional Wald test over the range of local alternatives considered.

NOTES

1. See, among others, numerical evidence presented in De Angelis, Hall, and Young (1993), Buchinsky (1995), and Horowitz (1998).
2. In particular, see Figure 1 in Section 4.
3. E.g., Koenker (2005, p. 122).
4. E.g., Koenker (2005, Thm. 4.1).
5. See, e.g., the estimator of $\boldsymbol{D}_1(\alpha)$ used by Hendricks and Koenker (1992).

6. See, e.g., Koenker (2005, Sect. 2.2.1).

7. The $\tau_i(\alpha)$ have been termed *regression rank scores* by Gutenbrunner and Jurečková (1992).

8. See, e.g., Mukhin (1985, 1991) for more information on conditions under which local limit theorems hold. From the point of view of empirical applications, restricting α to be rational does not appear to be terribly restrictive. In any case, the rationals are dense on [0, 1].

9. Further details are available in Knight and Goh (2008, App. A.1).

10. It is similarly straightforward to derive the limiting distribution of the endpoints of the interval containing the values of m such that $\hat{\boldsymbol{\beta}}_n(\alpha + (m/n)) = \hat{\boldsymbol{\beta}}_n(\alpha)$. This leads naturally to expressions for the limiting distributions of the length of this interval and of the number of "regression-quantile breakpoints"—a problem previously considered by Portnoy (1991). See Knight and Goh (2008, pp. 7–8) for further details.

11. In light of the results of Sakov and Bickel (2000) for the special case of the i.i.d. location model, use of resampling techniques involving resamples of size strictly smaller than the original sample may in fact be preferable.

12. It is noted that consideration of the regression-quantile process is natural in assessments of lack of fit in a quantile-regression setting and also in analyses of the effect of changes in the design variables on the conditional response distribution. See Koenker and Machado (1999) and Koenker and Xiao (2002) for general discussions of this problem.

13. Gutenbrunner and Jurečková (1992, Thm. 1).

14. For tests of sharp hypotheses regarding the value of the α-quantile of a population, Hall and Sheather (1988) consider Edgeworth expansions for the asymptotic distribution of the corresponding sequence of studentized univariate sample quantiles and propose the rule-of-thumb bandwidth $h_{HS} \equiv n^{-(1/3)} z_\tau^{2/3} \left[\frac{3s(\alpha)}{s''(\alpha)}\right]^{1/3}$, where z_τ is the $(1-(\tau/2))$-quantile of an $N(0,1)$-random variate and $\tau \in (0,1)$ denotes the desired size of the test. Here $s(\cdot)$ denotes the quantile density function of an $N(0,1)$-random variate, i.e., for $\alpha \in (0,1)$, $s(\alpha) \equiv 1/\left(\phi(\Phi^{-1}(\alpha))\right)$. In the case of the set of simulations considered here with $\alpha = 0.5$, $\tau = 0.05$, and $n = 100$, $h_{HS} = 0.2093$, which yields an implicit value of $m = h_{HS} \times 100^{3-1} \approx 20$.

REFERENCES

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

Buchinsky, M. (1995) Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study. *Journal of Econometrics* 68, 303–338.

De Angelis, D., P. Hall, & G.A. Young (1993) Analytical and bootstrap approximations to estimator distributions in L^1 regressions. *Journal of the American Statistical Association* 88, 1310–1316.

Gutenbrunner, C. & J. Jurečková (1992) Regression rank scores and regression quantiles. *Annals of Statistics* 20, 305–330.

Hall, P. & S.J. Sheather (1988) On the distribution of a Studentized quantile. *Journal of the Royal Statistical Society, Series B* 50, 381–391.

Hendricks, W. & R. Koenker (1992) Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association* 87, 58–68.

Horowitz, J.L. (1998) Bootstrap methods for median regression models. *Econometrica* 66, 1327–1351.

Knight, K. & C. Goh (2008) Asymptotics of Quantile Regression Basic Solutions with Applications. Mimeo, Department of Statistics, University of Toronto. An earlier version of this paper is available at http://www.chuangoh.org/rqs.pdf.

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

Koenker, R. & J.A.F. Machado (1999) Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* 94, 1296–1310.

Koenker, R. & Z. Xiao (2002) Inference on the quantile regression process. *Econometrica* 70, 1583–

1612.
Mukhin, A.B. (1985) Local limit theorems for distributions of sums of independent random vectors. *Theory of Probability and Its Applications* 29, 369–375.
Mukhin, A.B. (1991) Local limit theorems for lattice random variables. *Theory of Probability and Its Applications* 36, 698–713.
Portnoy, S. (1991) Asymptotic behavior of the number of regression quantile breakpoints. *SIAM Journal on Scientific and Statistical Computing* 12, 867–883.
Sakov, A. & P.J. Bickel (2000) An Edgeworth expansion for the m out of n bootstrapped median. *Statistics and Probability Letters* 49, 217–223.

APPENDIX A: Proofs

A.1. Proof of Lemma 1. The argument that follows is simply an elaboration of the asymptotic normality argument given in Bassett and Koenker (1978). In particular, let $\boldsymbol{W}(\alpha)$ denote the limit of $\boldsymbol{W}_n(\alpha) \equiv \sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right)$. Consider the joint density of $(\boldsymbol{\mathcal{X}}_1^\top, \ldots, \boldsymbol{\mathcal{X}}_d^\top, \mathcal{T}_1(\alpha), \ldots, \mathcal{T}_d(\alpha), \boldsymbol{W}^\top(\alpha))^\top$ with respect to the dominating measure $\prod_{j=1}^d \mu(d\boldsymbol{x}^j) \times \nu(d\boldsymbol{\tau}) \cdot \lambda(d\boldsymbol{w}) \times 1\left\{\boldsymbol{\tau} \in (-\alpha, 1-\alpha)^d\right\}$, where λ denotes Lebesgue measure. Let B_1, B_2, and B_3 be Borel sets in $\overbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}^{d}$, $(-\alpha, 1-\alpha)^d$, and $\mathbb{R}^d$, respectively. Let H be a subset of d elements from $\{1, \ldots, n\}$, $\boldsymbol{\Omega}_H$ the $d \times d$ matrix with columns $\{\boldsymbol{x}_i : i \in H\}$, and

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

Then

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

From Assumption 4 we have

$$n^{d/2} P\left[\boldsymbol{V}_n(\boldsymbol{w}, \boldsymbol{\Omega}_H) \in B_2\right]$$
$$= \nu(B_2) \cdot \frac{|\boldsymbol{\Omega}_H| \prod_{j=1}^d \kappa(\boldsymbol{x}^j)}{(2\pi\alpha(1-\alpha))^{d/2}|\boldsymbol{D}_0|^{1/2}} \exp\left[-\frac{\boldsymbol{w}^\top \boldsymbol{D}_1(\alpha)\boldsymbol{D}_0^{-1}\boldsymbol{D}_1(\alpha)\boldsymbol{w}}{2\alpha(1-\alpha)}\right] + o(1).$$

This implies that

$$P\left[\{\boldsymbol{x}_i : i \in \mathcal{H}_n(\alpha)\} \in B_1, \{\tau_i(\alpha) : i \in \mathcal{H}_n(\alpha)\} \in B_2, \boldsymbol{W}_n(\alpha) \in B_3\right]$$
$$= \nu(B_2) \cdot \sum_{\boldsymbol{\Omega}_H \in B_1} n^{-d} \int_{B_3} \frac{|\boldsymbol{\Omega}_H|^2 \prod_{j=1}^d \kappa(\boldsymbol{x}^j)}{[2\pi\alpha(1-\alpha)]^{d/2}|\boldsymbol{D}_0|^{1/2}} \exp\left[-\frac{\boldsymbol{w}^\top \boldsymbol{D}_1(\alpha)\boldsymbol{D}_0^{-1}\boldsymbol{D}_1(\alpha)\boldsymbol{w}}{2\alpha(1-\alpha)}\right]$$
$$\cdot \lambda(d\boldsymbol{w}) + o(1).$$

Now note that for any Borel set $B_1 \subset \overbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}^{d}$,

$$n^{-d} \sum_{\boldsymbol{\Omega}_H \in B_1} |\boldsymbol{\Omega}_H|^2 \xrightarrow{p} \frac{1}{d!} \int_{B_1} \left| \left[\boldsymbol{x}^1 \ldots \boldsymbol{x}^d \right] \right|^2 \prod_{j=1}^{d} \mu(d\boldsymbol{x}^j).$$

It follows that the joint density of $(\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_d^\top, \mathcal{T}_1(\alpha), \ldots, \mathcal{T}_d(\alpha), \boldsymbol{W}^\top(\alpha))^\top$ is given by

$$\begin{aligned} h(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^d, \boldsymbol{\tau}, \boldsymbol{w}) \\ &= \frac{\left| \left[\boldsymbol{x}^1 \ldots \boldsymbol{x}^d \right] \right|^2 \prod_{j=1}^d \kappa(\boldsymbol{x}^j)}{d!\,(2\pi\alpha(1-\alpha))^{d/2} |\boldsymbol{D}_0|^{1/2}} \exp\left[-\frac{\boldsymbol{w}^\top \boldsymbol{D}_1(\alpha) \boldsymbol{D}_0^{-1} \boldsymbol{D}_1(\alpha) \boldsymbol{w}}{2\alpha(1-\alpha)} \right] \prod_{j=1}^{d} \mu(d\boldsymbol{x}^j) \\ &\quad \cdot \nu(d\boldsymbol{\tau}) \cdot \lambda(d\boldsymbol{w}) \cdot 1\left[\boldsymbol{\tau} \in (-\alpha, 1-\alpha)^d \right]. \end{aligned}$$

The desired conclusion follows. ■

A.2. Proof of Lemma 2. For $\epsilon_i(\alpha) \equiv Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}(\alpha)$, define the point process

$$\bar{M}_n(A \times B) \equiv \sum_{i=1}^{n} 1\left[n\epsilon_i(\alpha) \in A,\ \boldsymbol{x} \in B \right],$$

which in turn converges in distribution to the Poisson process M.

Note that

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

where $\boldsymbol{\Omega}_n$ is a $d \times d$ matrix with columns denoted by $\boldsymbol{x}_i$, $i \in \mathcal{H}_n(\alpha)$, and $\boldsymbol{\xi}_n$ is a vector with components $n\epsilon_i(\alpha)$ for $i \in \mathcal{H}_n(\alpha)$. The conclusion follows by arguing conditionally given $\boldsymbol{\Omega}_n^{-1}\boldsymbol{\xi}_n$ and observing that $\boldsymbol{\Omega}_n^{-1}\boldsymbol{\xi}_n = O_p\left(n^{1/2}\right)$. ■

A.3. Proof of Theorem 1. Applying the identity

$$\rho_\alpha(u - v) - \rho_\alpha(u) = -v\psi_\alpha(u) + \int_0^v (1(u \le s) - 1(u \le 0))\, ds,$$

we have

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

It follows from Lemmas 1 and 2 that, for each $\boldsymbol{u}$,

$$\begin{aligned} Z_n(\boldsymbol{u}) &= -\frac{m}{n} \sum_{i \notin \mathcal{H}_n(\alpha)} \boldsymbol{x}_i^\top \boldsymbol{u} - \sum_{i \in \mathcal{H}_n(\alpha)} \tau_i(\alpha) \boldsymbol{x}_i^\top \boldsymbol{u} + \sum_{i \in \mathcal{H}_n(\alpha)} \rho_{\alpha+(m/n)}(-\boldsymbol{x}_i^\top \boldsymbol{u}) \\ &\quad + \sum_{i=1}^{n} \int_0^{\boldsymbol{x}_i^\top \boldsymbol{u}} \left[1\left(n\hat{\epsilon}_i(\alpha) \le s \right) - 1\left(n\hat{\epsilon}_i(\alpha) < 0 \right) \right] ds \end{aligned}$$

$$
\begin{aligned}
&\xrightarrow{d} -m \int \boldsymbol{u}^\top \boldsymbol{x} \mu(d\boldsymbol{x}) - \sum_{j=1}^{d} \mathcal{T}_j \boldsymbol{u}^\top \boldsymbol{\mathcal{X}}_j + \sum_{j=1}^{d} \rho_{1-\alpha}(\boldsymbol{u}^\top \boldsymbol{\mathcal{X}}_j) \\
&\quad + \sum_{k \neq 0} \int_0^{\boldsymbol{X}_k^\top \boldsymbol{u}} \left[1(\Gamma_k \leq s) - 1(\Gamma_k < 0) \right] ds.
\end{aligned}
$$

As mentioned in Section 2, the convergence in distribution of $n\left[\hat{\boldsymbol{\beta}}_n(\alpha + (m/n)) - \hat{\boldsymbol{\beta}}_n(\alpha)\right]$ follows from the convexity of the objective function $Z_n(\cdot)$. ■