The Annals of Statistics
2012, Vol. 40, No. 3, 1714–1736
DOI: 10.1214/12-AOS1021

NEARLY ROOT-*n* APPROXIMATION FOR REGRESSION QUANTILE PROCESSES

BY STEPHEN PORTNOY[1]

University of Illinois at Urbana-Champaign

Traditionally, assessing the accuracy of inference based on regression quantiles has relied on the Bahadur representation. This provides an error of order $n^{-1/4}$ in normal approximations, and suggests that inference based on regression quantiles may not be as reliable as that based on other (smoother) approaches, whose errors are generally of order $n^{-1/2}$ (or better in special symmetric cases). Fortunately, extensive simulations and empirical applications show that inference for regression quantiles shares the smaller error rates of other procedures. In fact, the "Hungarian" construction of Komlós, Major and Tusnády [*Z. Wahrsch. Verw. Gebiete* **32** (1975) 111–131, *Z. Wahrsch. Verw. Gebiete* **34** (1976) 33–58] provides an alternative expansion for the one-sample quantile process with nearly the root-n error rate (specifically, to within a factor of $\log n$). Such an expansion is developed here to provide a theoretical foundation for more accurate approximations for inference in regression quantile models. One specific application of independent interest is a result establishing that for conditional inference, the error rate for coverage probabilities using the Hall and Sheather [*J. R. Stat. Soc. Ser. B Stat. Methodol.* **50** (1988) 381–391] method of sparsity estimation matches their one-sample rate.

1. Introduction.

Consider the classical regression quantile model: given independent observations $\{(x_i Y_i) : i = 1, \dots, n\}$, with $x_i \in R^p$ fixed (for fixed p), the conditional quantile of the response Y_i given x_i is

$$Q_{Y_i}(\tau|x_i) = x_i'\beta(\tau).$$

Let $\hat{\beta}(\tau)$ be the Koenker–Bassett regression quantile estimator of $\beta(\tau)$. Koenker (2005) provides definitions and basic properties, and describes the traditional approach to asymptotics for $\hat{\beta}(\tau)$ using a Bahadur representation:

$$B_n(\tau) \equiv n^{1/2}(\hat{\beta}(\tau) - \beta(\tau)) = D(x)W(\tau) + R_n,$$

where $W(t)$ is a Brownian Bridge and R_n is an error term.

Unfortunately, R_n is of order $n^{-1/4}$ [see, e.g., Jurečková and Sen (1996) and Knight (2002)]. This might suggest that asymptotic results are accurate only to

Received August 2011; revised May 2012.

[1]Supported in part by NSF Grant DMS-10-07396.

MSC2010 subject classifications. Primary 62E20, 62J99; secondary 60F17.

Key words and phrases. Regression quantiles, asymptotic approximation, Hungarian construction.

this order. However, both simulations in regression cases and one-dimensional results [Komlós, Major and Tusnády (1975, 1976)] justify a belief that regression quantile methods should share (nearly) the $O(n^{-1/2})$ accuracy of smooth statistical procedures (uniformly in τ). In fact, as shown in Knight (2002), $n^{1/4}R_n$ has a limit with zero mean and that is independent of $W(\tau)$. Thus, in any smooth inferential procedure (say, confidence interval lengths or coverages), this error term should enter only through $ER_n^2 = \mathcal{O}(n^{-1/2})$. Nonetheless, this expansion would still leave an error of $o(n^{-1/4})$ (coming from the error beyond the R_n term in the Bahadur representation), and so would still fail to reflect root-n behavior. Furthermore, previous results only provide such a second-order expansion for fixed τ.

It must be noted that the slower $\mathcal{O}(n^{-1/4})$ error rate arises from the discreteness introduced by indicator functions appearing in the gradient conditions. In fact, expansions can be carried out when the design is assumed to be random; see De Angelis, Hall and Young (1993) and Horowitz (1998), where the focus is on analysis of the (x, Y) bootstrap. Specifically, the assumption of a smooth distribution for the design vectors together with a separate treatment of the lattice contribution of the intercept does permit appropriate expansions. Unfortunately, the randomness in X means that all inference must be in terms of the average asymptotic distribution (averaged over X), and so fails to apply to the generally more desirable conditional forms of inference. Specifically, unconditional methods may be quite poor in the heteroscedastic and nonsymmetric cases for which regression quantile analysis is especially appropriate. The main goal of this paper is to reclaim increased accuracy for conditional inference beyond that provided by the traditional Bahadur representation.

Specifically, the aim is to provide a theoretical justification for an error bound of nearly root-n order uniformly in τ. Define

$$\hat{\delta}_n(\tau) = \sqrt{n}\big(\hat{\beta}(\tau) - \beta(\tau)\big).$$

We first develop a normal approximation for the density of $\hat{\delta}$ with the following form:

$$f_{\hat{\delta}}(\delta) = \varphi_\Sigma(\delta)\big(1 + \mathcal{O}\big(L_n n^{-1/2}\big)\big)$$

for $\|\delta\| \le D\sqrt{\log n}$, where $L_n = (\log n)^{3/2}$. We then extend this result to the densities of a pair of regression quantiles in order to obtain a "Hungarian" construction [Komlós, Major and Tusnády (1975, 1976)] that approximates the process $B_n(\tau)$ by a Gaussian process to order $\mathcal{O}(L_n^* n^{-1/2})$, where $L_n^* = (\log n)^{5/2}$ (uniformly for $\varepsilon \le \tau \le 1 - \varepsilon$).

Section 2 provides some applications of the results here to conditional inference methods in regression quantile models. Specifically, an expansion is developed for coverage probabilities of confidence intervals based on the [Hall and Sheather (1988)] difference quotient estimator of the sparsity function. The coverage error rate is shown to achieve the rate $\mathcal{O}(n^{-2/3} \log n)$ for conditional inference, which is nearly the known "optimal" rate obtained for a single sample and for unconditional inference. Section 3 lists the conditions and main results, and offers some remarks. Section 4 provides a description of the basic ingredients of the proof (since this proof is rather long and complicated). Section 5 proves the density approximation for a fixed τ (with multiplicative error). Section 6 extends the result to pairs of regression quantiles (Theorem 1), and Section 7 provides the "Hungarian" construction (Theorem 2) with what appears to be a somewhat innovative induction along dyadic rationals.

2. Implications for applications. As the impetus for this work was the need to provide some theoretical foundation for empirical results on the accuracy of regression quantile inference, some remarks on implications are in order.

REMARK 1. Clearly, whenever published work assesses the accuracy of an inferential method using the error term from the Bahadur representation, the present results will immediately provide an improvement from $\mathcal{O}(n^{-1/4})$ to the nearly root-n rate here. One area of such results is methods based directly on regression quantiles and not requiring estimation of the sparsity function $[1/f(F^{-1}(\tau))]$. There are several papers giving such results, although at present it appears that their methods have theoretical justification only under location-scale forms of quantile regression models.

Specifically, Zhou and Portnoy (1996) introduced confidence intervals (especially for fitted values) based on using pairs of regression quantiles in a way analogous to confidence intervals for one-sample quantiles. They showed that the method was consistent, but the accuracy depended on the Bahadur error term. Thus, results here now provide accuracy to the nearly root-n rate of Theorem 2.

A second approach directly using the dual quantile process is based on the regression ranks of Gutenbrunner et al. (1993). Again, the error terms in the theoretical results there can be improved using Theorem 1 here, though the development is not so direct.

For a third application, Neocleous and Portnoy (2008) showed that the regression quantile process interpolated along a grid of mesh strictly larger than $n^{-1/2}$ is asymptotically equivalent to the full regression quantile process to first order, but (because of additional smoothness) will yield monotonic quantile functions with probability tending to 1. However, their development used the Bahadur representation, which indicated that a mesh of order $n^{-1/3}$ balanced the bias and accuracy and bounded the difference between $\hat{\beta}(\tau)$ and its linear interpolate by nearly $\mathcal{O}(n^{-1/6})$. With some work, use of the results here would permit a mesh slightly larger than the nearly root-n rate here to obtain an approximation of nearly root-n order.

REMARK 2. Inference under completely general regression quantile models appears to require either estimation of the sparsity function or use of resampling methods. The most general methods in the quantreg package [Koenker (2012)] use the "difference quotient" method with the [Hall and Sheather (1988)] bandwidth of order $n^{-1/3}$, which is known to be optimal for coverage probabilities in the one-sample problem. As noted above, expansions using the randomness of the regressors can be developed to provide analogous results for unconditional inference. The results here (with some elaboration) can be used to show that the Hall–Sheather estimates provide (nearly) the same rates of accuracy for coverage probabilities under the conditional form of the regression quantile model.

To be specific, consider the problem of confidence interval estimation for a fixed linear combination of regression parameters: $a'\beta(\tau)$. The asymptotic variance is the well-known sandwich formula

$$s_a^2(\delta) = \tau(1-\tau)a'(X'DX)^{-1}(X'X)(X'DX)^{-1}a, \qquad D \equiv \operatorname{diag}(x_i'\delta), \tag{2.1}$$

where δ is the sparsity, $\delta = \beta'(\tau)$ (with β' being the gradient), and where X is the design matrix.

Following Hall and Sheather (1988), the sparsity may be approximated by the difference quotient $\tilde{\delta} = (\beta(\tau+h) - \beta(\tau-h))/(2h)$. Standard approximation theory (using the Taylor series) shows that

$$\delta = \tilde{\delta} + \mathcal{O}(h^2).$$

The sparsity may be estimated by

$$\hat{\delta} \equiv \Delta(h)/(2h) \equiv (\hat{\beta}(\tau+h) - \hat{\beta}(\tau-h))/(2h), \tag{2.2}$$

and the sparsity (2.1) may be estimated by inserting $\hat{\delta}$ in D.

Then, as shown in the Appendix, the confidence interval

$$a'\beta(\tau) \in a'\hat{\beta}(\tau) \pm z_\alpha s_a(\hat{\delta}) \tag{2.3}$$

has coverage probability $1 - 2\alpha + \mathcal{O}((\log n)n^{-2/3})$, which is within a factor of $\log n$ of the optimal Hall–Sheather rate in a single sample. Furthermore, this rate is achieved at the (optimal) h-value $h_n^* = c\sqrt{\log n}n^{-1/3}$, which is the optimal Hall–Sheather bandwidth except for the $\sqrt{\log n}$ term.

Since the optimal bandwidth depends on R_n^*, the optimal constant for the h_n^* cannot be determined, as it can when X is allowed to be random [and for which the $\mathcal{O}(1/(nh_n))$ term is explicit]. This appears to be an inherent shortcoming for using inference conditional on the design.

Note also that it is possible to obtain better error rates for the coverage probability by using higher order differences. Specifically, using the notation of (2.2),

$$\tfrac{4}{3}\Delta(h) - \tfrac{1}{6}\Delta(2h) = \beta'(\tau) + \mathcal{O}(h^4).$$

As a consequence, the optimal bandwidth for this estimator is of order $n^{-1/5}$, and the coverage probability is accurate to order $n^{-4/5}$ (except for logarithmic factors).

REMARK 3. A third approach to inference applies resampling methods. As noted in the Introduction, while the (x, Y) bootstrap is available for unconditional inference, the practicing statistician will generally prefer to use inference conditional on the design. There are some resampling approaches that can obtain such inference. One method is that of Parzen, Wei and Ying (1994), which simulates the binomial variables appearing in the gradient condition. Another is the "Markov Chain Marginal Bootstrap" of He and Hu (2002) [see also Kocherginsky, He and Mu (2005)]. However, this method also involves sampling from the gradient condition. The discreteness in the gradient condition would seem to require the error term from the Bahadur representation, and thus leads to poorer inferential approximation: the error would be no better than order $n^{-1/2}$ even if it were the square of the Bahadur error term. While some evidence for decent performance of these methods comes from (rather limited) simulations, it is often noticed that these methods perform perhaps somewhat more poorly than the other methods in the `quantreg` package of Koenker (2012). Clearly, a more complete analysis of inference for regression quantiles based on the more accurate stochastic expansions here would be useful.

3. Conditions, fundamental theorems and remarks. Under the regression quantile model of Section 1, the following conditions will be imposed:

Let $\dot{x}_i$ denote the coordinates of x_i except for the intercept (i.e., the last $p - 1$ coordinates, if there is an intercept). Let $\dot{\phi}_i(t)$ denote the conditional characteristic function of the random variable $\dot{x}_i(I(Y_i \le x_i'\beta(\tau) + \delta/\sqrt{n}) - \tau)$, given x_i. Let $f_i(y)$ and $F_i(y)$ denote the conditional density and c.d.f. of Y_i given x_i.

CONDITION X1. For any $\varepsilon > 0$, there is $\eta \in (0, 1)$ such that

$$\inf_{\|t\|>\varepsilon} \prod \dot{\phi}_i(t) \le \eta^n \tag{3.1}$$

uniformly in $\varepsilon \le \tau \le 1 - \varepsilon$.

CONDITION X2. $\|x_i\|$ are uniformly bounded, and there are positive definite $p \times p$ matrices $G = G(\tau)$ and H such that for any $\varepsilon > 0$ (as $n \to \infty$)

$$G_n(\tau) \equiv \frac{1}{n} \sum_{i=1}^{n} f_i(x_i'\beta(\tau)) x_i' x_i = G(\tau)(1 + \mathcal{O}(n^{-1/2})), \tag{3.2}$$

$$H_n \equiv \frac{1}{n} \sum_{i=1}^{n} x_i' x_i = H(1 + \mathcal{O}(n^{-1/2})) \tag{3.3}$$

uniformly in $\varepsilon \le \tau \le 1 - \varepsilon$.

CONDITION F. The derivative of $\log(f_i(y))$ is uniformly bounded on the interval $\{y : \varepsilon \le F_i(y) \le 1 - \varepsilon\}$.

Two fundamental results will be developed here. The first result provides a density approximation with multiplicative error of nearly root-n rate. A result for fixed τ is given in Theorem 5, but the result needed here is a bivariate approximation for the joint density of one regression quantile and the difference between this one and a second regression quantile (properly normalized for the difference in τ-values).

Let $\varepsilon \le \tau_1 \le 1 - \varepsilon$ for some $\varepsilon > 0$, and let $\tau_2 = \tau_1 + a_n$ with $a_n > cn^{-b}$ for some $b < 1$. Here, one may want to take b near 1 [see remark (1) below], though the basic result will often be useful for $b = \frac{1}{2}$, or even smaller. Define

$$B_n = B_n(\tau_1) \equiv n^{1/2}\big(\hat{\beta}(\tau_1) - \beta(\tau_1)\big), \tag{3.4}$$

$$R_n = R_n(\tau_1, \tau_2) \equiv (na_n)^{1/2}\big[\big(\hat{\beta}(\tau_1) - \beta(\tau_1)\big) - \big(\hat{\beta}(\tau_2) - \beta(\tau_2)\big)\big]. \tag{3.5}$$

THEOREM 1. *Under Conditions X1, X2 and F, there is a constant D such that for $|B_n| \le D(\log n)^{1/2}$ and $|R_n| \le D(\log n)^{1/2}$, the joint density of R_n and B_n at δ and s, respectively, satisfies*

$$f_{R_n,B_n}(\delta, s) = \varphi_{\Gamma_n}(\delta, s)\big(1 + \mathcal{O}\big(\big(na_n(\log n)^3\big)^{-1/2}\big)\big),$$

where φ_{Γ_n} is a normal density with covariance matrix Γ_n having the form given in (7.3).

The second result provides the desired "Hungarian" construction:

THEOREM 2. *Assume Conditions X1, X2 and F. Fix $a_n = n^{-b}$ with $b < 1$, and let $\{\tau_j\}$ be dyadic rationals with denominator less than n^b. Define $B_n^*(\tau)$ to be the piecewise linear interpolant of $\{B_n(\tau_j)\}$ [as defined in (3.4)]. Then for any $\varepsilon > 0$, there is a (zero-mean) Gaussian process, $\{Z_n(\tau_j)\}$, defined along the dyadic rationals $\{\tau_j\}$ and with the same covariance structure as $B_n^*(\tau)$ (along $\{\tau_j\}$) such that its piecewise linear interpolant $Z_n^*(\tau)$ satisfies*

$$\sup_{\varepsilon \le \tau \le 1-\varepsilon} \big|B_n^*(\tau) - Z_n^*(\tau)\big| = \mathcal{O}\bigg(\frac{(\log n)^{5/2}}{\sqrt{n}}\bigg)$$

almost surely.

Some remarks on the conditions and ramifications are in order:

(1) The usual construction approximates $B_n(\tau)$ by a "Brownian Bridge" process. Theorem 2 really only provides an approximation for the discrete processes at a sufficiently sparse grid of dyadic rationals. That the piecewise linear interpolants converge to the usual Brownian Bridge follows as in Neocleous and Portnoy (2008). The critical impediment to getting a Brownian Bridge approximation to $B_n(\tau)$ with the error in Theorem 2 is the square root behavior of the modulus of continuity. This prevents approximating the piecewise linear interpolant within

an interval of length greater than (roughly) order $1/n$ if a root-n error is desired. In order to approximate the density of the difference in $B_n(\tau)$ over an interval between dyadic rationals, the length of the interval must be at least of order n^{-b} (for $b < 1$). Clearly, it will be possible to approximate the piecewise linear interpolant by a Brownian Bridge with error $\sqrt{n^{-b}} = n^{-b/2}$, and thus to get arbitrarily close to the value of $\frac{1}{2}$ for the exponent of n. For most purposes, it might be better to state the final result as

$$\sup_{\varepsilon \le \tau \le 1-\varepsilon} \|B_n(\tau) - Z(\tau)\| = \mathcal{O}(n^{-a})$$

for any $a < 1/2$ (where Z is the appropriate Brownian Bridge); but the stronger error bound of Theorem 2 does provide a much closer analog of the result for the one-sample (one-dimensional) quantile process.

(2) The one-sample result requires only the first power of $\log n$, which is known to give the best rate for a general result. The extra addition of $3/2$ in the exponent is clearly needed for the density approximation, but this may be only a technical assumption. Nonetheless, I conjecture that some extra amount is needed in the exponent.

(3) Conditions X1 and X2 can be shown to hold with probability tending to one under smoothness and boundedness assumptions of the distribution of x. Nonetheless, the condition that $\|x\|$ be bounded seems rather strong in the case of random x. It seems clear that this can be weakened, though probably at the cost of getting a poorer approximation. For example, $\|x\|$ having exponentially small tails might increase the bound in Theorem 2 by an additional factor of $\log n$, and algebraic tails are likely worse. However, details of such results remain to be developed.

(4) Similarly, it should be possible to let ε, which defines the compact subinterval of τ-values, tend to zero. Clearly, letting ε_n be of order $1/n$ would lead to extreme value theory and very different approximations. For slower rates of convergence of ε_n, Bahadur expansions have been developed [e.g., see Gutenbrunner et al. (1993)] and extension to the approximation result in Theorem 2 should be possible. Again, however, this would most likely be at the cost of a larger error term.

(5) The assumption that the conditional density of the response (given x) be continuous is required even for the usual first order asymptotics. However, one might hope to avoid Condition F, which requires a bounded derivative at all points. For example, the double exponential distribution does not satisfy this condition. It is likely that the proofs here can be extended to the case where the derivative does not exist on a finite set (or even on a set of measure zero), but dropping differentiability entirely would require a rather different approach. Furthermore, the apparent need for bounded derivatives in providing uniformity over τ in Bahadur expansions suggests the possibility that some differentiability is required.

(6) Theorem 1 provides a bivariate normal density approximation with error rate (nearly) $n^{-1/2}$ when τ_1 and τ_2 are fixed. When $a_n \equiv \tau_2 - \tau_1 \to 0$, of course, the error rate is larger. Note, however, that the slower convergence rate when $a_n \to 0$ does not reduce the order of the error in the final construction since the difference $D_n = \hat{\beta}(\tau_2) - \hat{\beta}(\tau_1)$ is of order $(na_n)^{-1/2}$.

4. Ingredients and outline of proof. The development of the fundamental results (Theorems 1 and 2) will be presented in three phases. The first phase provides the density approximation for a fixed τ, since some of the more complicated features are more transparent in this case. The second phase extends this result to the bivariate approximation of Theorem 1. The final phase provides the "Hungarian" construction of Theorem 2. To clarify the development, the basic ingredients and some preliminary results will be presented first.

INGREDIENT 1. Begin with the finite sample density for a regression quantile [Koenker (2005), Koenker and Bassett (1978)]: assume Y_i has a density, $f_i(y)$, and let τ be fixed. Note that $\hat{\beta}(\tau)$ is defined by having p zero residuals (if the design is in general position). Specifically, there is a subset, h, of p integers such that $\hat{\beta}(\tau) = X_h^{-1} Y_h$, where X_h has rows x_i' for $i \in h$ and Y_h has coordinates Y_i for $i \in h$. Let $\mathcal{H}$ denote the set of all such p-element subsets. Define

$$\hat{\delta} = \sqrt{n}\big(\hat{\beta}(\tau) - \beta(\tau)\big).$$

As described in Koenker (2005), the density of $\hat{\delta}$ evaluated at the argument $\delta = \sqrt{n}(b - \beta(\tau))$ is given by

$$f_{\hat{\delta}}(\delta) = n^{-p/2} \sum_{h \in \mathcal{H}} \det(X_h) P\{S_n \in A_h\} \prod_{i \in h} f_i\big(x_i'\beta(\tau) + n^{-1/2}\delta\big). \tag{4.1}$$

Here, the event in the probability above is the event that the gradient condition holds for a fixed subset, $h : S_n \in A_h$, where $A_h = X_h R$, with R the rectangle that is the product of intervals $(\tau - 1, \tau)$ [see Theorem 2.1 of Koenker (2005)], and where

$$S_n = S_n(h, \beta, \delta) \equiv \sum_{i \notin h} x_i \big(I\big(Y_i \le x_i'\beta + n^{-1/2}\delta\big) - \tau\big). \tag{4.2}$$

INGREDIENT 2. Since $n^{-1/2} S_n$ is approximately normal, and A_h is bounded, the probability in (4.1) is approximately a normal density evaluated at δ. To get a multiplicative bound, we may apply a "Cramér" expansion (or a saddlepoint approximation). If S_n had a smooth distribution (i.e., satisfied Cramér's condition), then standard results would apply. Unfortunately, S_n is discrete. The first coordinate of S_n is nearly binomial, and so a multiplicative bound can be obtained by applying a known saddlepoint formula for lattice variables [see Daniels (1987)]. Equivalently, approximate by an exact binomial and (more directly, but with some rather tedious computation) expand the logarithm of the Gamma function in Stirling's formula. Using either approach, one can show the following result:

THEOREM 3. *Let* $W \sim \text{binomial}(n, p)$, J *be any interval of length* $\mathcal{O}(\sqrt{n})$ *containing* $EW = np$, *and let* $w = \mathcal{O}(\sqrt{n \log(n)})$. *Then*

$$P\{W \in J + w\} = P\{Z \in J + w\}\big(1 + \mathcal{O}\big(n^{-1/2}\sqrt{\log(n)}\big)\big), \tag{4.3}$$

where $Z \sim \mathcal{N}(np, np(1-p))$.

A proof based on multinomial expansions is given for the bivariate generalization in Theorem 1. Note that this result includes an extra factor of $\sqrt{\log(n)}$. This will allow the bounds to hold except with probability bounded by an arbitrarily large negative power of n. This is clear for the limiting normal case (by standard asymptotic expansions of the normal c.d.f.). To obtain such bounds for the distribution of S_n will require some form of Bernstein's inequality. Such inequalities date to Bernstein's original publication in 1924 [see Bernstein (1964)], but a version due to Hoeffding (1963) may be easier to apply.

INGREDIENT 3. Using Theorem 3, it can be shown (see Section 4) that the probability in (4.1) may be approximated as

$$P\{\tilde{S}_n \in A_h\}\big(1 + \mathcal{O}\big(L_n/\sqrt{n}\big)\big),$$

where the first coordinate of $\tilde{S}_n$ is a sum of n i.i.d. $\mathcal{N}(0, \tau(1-\tau))$ random variables, the last $(p-1)$ coordinates are those of S_n, and $L_n = (\log n)^{3/2}$. Since we seek a normal approximation for this probability with multiplicative error, at this point one might hope that a known (multidimensional) "Cramér" expansion or saddlepoint approximation would allow $\tilde{S}_n$ to be replaced by a normal vector (thus providing the desired result). However, this will require that the summands be smooth, or (at least) satisfy a form of Cramér's condition. Let $\dot{x}_i$ denote the last $(p-1)$ coordinates of x_i. One approach would be to assume $\dot{x}_i$ has a smooth distribution satisfying the classical form of Cramér's condition. However, to maintain a conditional form of the analysis, it suffices to impose a condition on $\dot{x}_i$, which is designed to mimic the effect of a smooth distribution and will hold with probability tending to one if $\dot{x}_i$ has such a smooth distribution. Condition X1 specifies just such an assumption.

Note that the characteristic functions of the summands of $\tilde{S}_n$, say, $\{\dot{\phi}_i(t)\}$, will also satisfy Condition X1 [equation (3.1)] and so should allow application of known results on normal approximations. Unfortunately, I have been unable to find a published result providing this and so Section 5 will present an independent proof.

Clearly, some additional conditions will be required. Specifically, we will need conditions that the empirical moments of $\{x_i\}$ converge appropriately, as specified in Condition X2.

Finally, the approach using characteristic functions is greatly simplified when the sums, $\tilde{S}_n$, have densities. Again, to avoid using smoothness of the distribution of $\{\dot{x}_i\}$ (and thus to maintain a conditional approach), introduce a random perturbation V_n which is small and has a bounded smooth density (the bound may depend on n). Section 4 will then prove the following:

THEOREM 4. *Assume Conditions X1 and X2 and the regression quantile model of Section 1. Let δ be the argument of the density of $n^{-1/2}(\hat{\beta} - \beta)$, and suppose*

$$\|\delta\| \le d\sqrt{n}$$

for some constant d. Then a constant d_0 can be chosen so that

$$P\{S_n + V_n \in A_h\} = P\left\{Z_n + \frac{V_n}{\sqrt{n}} \in \frac{A_h}{\sqrt{n}}\right\}\left(1 + \mathcal{O}\left(\frac{\log^{3/2}(n)}{\sqrt{n}}\right)\right) + \mathcal{O}(n^{-d_0}),$$

where Z_n has mean $-G_n^{-1}\delta$ and covariance $\tau(1-\tau)H_n$, d_0 can be arbitrarily large, and V_n is a small perturbation [*see* (5.1)].

Following the proof of this theorem, it will be shown that the effect of V_n can be ignored, if V_n is bounded by n^{-d_1}, where d_1 may depend on d (but not on d_0).

INGREDIENT 4. Expanding the densities in (4.1) is trivial if the densities are sufficiently smooth. The assumption of a bounded first derivative in Condition F appears to be required to analyze second order terms (beyond the first order normal approximation).

INGREDIENT 5. Finally, summing terms involving $\det(X_h)$ in (4.1) over the $\binom{n}{p}$ summands will require Vinograd's theorem and related results from matrix theory concerning adjoint matrices [see Gantmacher (1960)].

The remaining ingredients provide the desired "Hungarian" construction.

INGREDIENT 6. Extend the density approximation to the joint density for $\hat{\beta}(\tau_1)$ and $\hat{\beta}(\tau_2)$ (when standardized). A major complication is that one needs $a_n \equiv |\tau_2 - \tau_1| \to 0$, making the covariance matrix tend to singularity. Thus, we focus on the joint density for standardized versions of $\hat{\beta}(\tau_1)$ and $D_n \equiv \hat{\beta}(\tau_2) - \hat{\beta}(\tau_1)$. Clearly, this requires modification of the proof for the univariate case to treat the fact that D_n converges at a rate depending on a_n. The result is given in Theorem 1.

INGREDIENT 7. Extend the density result to obtain an approximation for the quantile transform for the conditional distribution of differences D_n (between successive dyadic rationals). This will provide (independent) normal approximations to the differences whose sums will have the same covariance structure as the regression quantile process (at least along a sufficiently sparse grid of dyadic rationals).

INGREDIENT 8. Finally, the Hungarian construction is applied inductively along the sparse grid of dyadic rationals. This inductive step requires some innovative development, mainly because the regression quantile process is not directly expressible in terms of sums of random variables (as are the empiric one-sample distribution function and quantile function).

5. Proof of Theorem 4. Let $\dot{S}_n$ be the last $p-1$ coordinates of S_n and $A^{(1)}(\dot{S}_n, h)$ be the interval $\{a : (a, \dot{S}_n) \in A_h\}$. Then,

$$\begin{aligned} P\{S_n \in A_h\} &= P\left\{\sum_{i \notin h}\big(I\big(Y_i \le x_i'\beta + \delta/\sqrt{n}\big) - \tau\big) \in A^{(1)}(\dot{S}_n, h)\right\} \\ &= P\left\{\sum_{i \notin h}\big(I\big(Y_i \le x_i'\beta\big) - \tau\big) \in A^{(1)}(\dot{S}_n, h)\right. \\ &\qquad \left. - \sum_{i \notin h}\big(I\big(Y_i \le x_i'\beta + \delta/\sqrt{n}\big) - I\big(Y_i \le x_i'\beta\big)\big)\right\} \\ &= \sum_{k \in A^*} f_{\text{binomial}}(k; \tau), \end{aligned}$$

where A^* is the set $A^{(1)}$ shifted as indicated above. Note that by Hoeffding's inequality [Hoeffding (1963)], for any fixed d, the shift satisfies

$$\left|\sum_{i \notin h}\big(I\big(Y_i \le x_i'\beta + \delta/\sqrt{n}\big) - I\big(Y_i \le x_i'\beta\big)\big)\right| \le d\sqrt{n}\sqrt{\log(n)}$$

except with probability bounded by $2n^{-2d^2}$. Thus, we may apply Theorem 3 [equation (4.3)] with w equal to the shift above to obtain the following bound (to within an additional additive error of $2n^{-2d^2}$):

$$P\{S_n \in A_h\} = P\big\{nZ\sqrt{\tau(1-\tau)} \in A^{(1)}(\dot{S}_n, h)\big\}\big(1 + \mathcal{O}\big(a_n/\sqrt{n}\big)\big),$$

where $Z \sim \mathcal{N}(0, 1)$ and a_n is a bound on $\dot{S}_n$, which may be taken to be of the form $B\sqrt{\log n}$ (by Hoeffding's inequality). Finally, we obtain

$$P\{S_n \in A_h\} = P\{\tilde{S}_n \in A_h\}\big(1 + \mathcal{O}\big(a_n/\sqrt{n}\big)\big) + 2n^{-2d^2},$$

where the first coordinate of $\tilde{S}_n$ is a sum of n i.i.d. $\mathcal{N}(0, \tau(1-\tau))$ random variables and the last $p-1$ coordinates are those of S_n.

To treat the probability involving $\tilde{S}_n$, standard approaches using characteristic functions can be employed. In theory, exponential tilting (or saddlepoint methods) should provide better approximations, but since we require only the order of the leading error term, we can proceed more directly. As in Einmahl (1989), the first step is to add an independent perturbation so that the sum has an integrable density: specifically, for fixed $h \in \mathcal{H}$ let V_n be a random variable (independent of all observations) with a smooth bounded density and for which (for each $h \in \mathcal{H}$)

$$\|V_n\| \le n^{-d_1}, \tag{5.1}$$

where d_1 will be chosen later. Define

$$S_n^* = \tilde{S}_n + V_n.$$

We now allow A_h to be any (arbitrary) set, say, A. Thus, S_n^* has a density and we can write [with $c_\pi = (2\pi)^{-p}$]

$$P\{S_n^*/\sqrt{n} \in A\} = c_\pi \int \mathrm{Vol}(A)\phi_{\mathrm{Unif}(A)}(t)\phi_{\tilde{S}_n}(t/\sqrt{n})\phi_{V_n}(t/\sqrt{n})\,dt,$$

where ϕ_U denotes the characteristic function of the random variable U.

Break domain of integration into 3 sets: $\|t\| \le d_2\sqrt{\log(n)}$, $d_2\sqrt{\log(n)} \le \|t\| \le \varepsilon\sqrt{n}$, and $\|t\| \ge \varepsilon\sqrt{n}$.

On $\|t\| \le d\sqrt{\log(n)}$, expand $\log \phi_{\tilde{S}_n/\sqrt{n}}(t)$. For this, compute

$$\begin{aligned}
\mu_i &\equiv E x_i\bigl(\tau - I\bigl(y_i \le x_i'\beta + x_i'\delta/\sqrt{n}\bigr)\bigr)\\
&= -f_i\bigl(F_i^{-1}(\tau)\bigr)x_i x_i'\delta/\sqrt{n} + \mathcal{O}\bigl(\|x_i\|^3\|\delta\|^2/n\bigr),\\
\Sigma_i &\equiv \mathrm{Cov}\bigl[x_i\bigl(\tau - I\bigl(y_i \le x_i'\beta + x_i'\delta/\sqrt{n}\bigr)\bigr)\bigr]\\
&= x_i x_i'\tau(1-\tau) + \mathcal{O}\bigl(\|x_i\|^3\|\delta\|^2/n\bigr).
\end{aligned}$$

Hence, using the boundedness of $\|x_i\|$, $\|\delta\|$ and $\|t\|$ (on this first interval),

$$\begin{aligned}
\phi_{\tilde{S}_n}(t/\sqrt{n}) &= \exp\biggl\{-\iota\sum_{i\notin h}\mu_i/\sqrt{n}t'\delta - \frac{1}{2}\sum_{i\notin h}t'\Sigma_i t/n + \mathcal{O}\biggl(\frac{\|\delta\|^2 + \|t\|^3}{\sqrt{n}}\biggr)\biggr\}\\
&= \exp\biggl\{-\iota G_n t'\delta - \frac{1}{2}t' H_n t + \mathcal{O}\bigl((\log n)^{3/2}/\sqrt{n}\bigr)\biggr\},
\end{aligned}$$

where G_n and H_n are defined in Condition X2 [see (3.2) and (3.3)].

For the other two intervals on the t-axis, the integrands will be bounded by an additive error times

$$\int \phi_{V_n}(t/\sqrt{n})\,dt = \mathcal{O}\bigl(n^{-p(d_1+1/2)}\bigr)$$

since $\|V_n\| \le n^{-d_1}$.

On $\|t\| \le \varepsilon\sqrt{n}$, the summands are bounded and so their characteristic functions satisfy $\phi_i(s) \le (1 - b\|t\|^2)$ for some constant c. Thus, on $d_2\sqrt{\log(n)} \le \|t\| \le \varepsilon\sqrt{n}$,

$$|\phi_{\tilde{S}_n}(t/\sqrt{n})| \le (1 - bd_2^2 \log(n)/n)^{n-p} \le c_1 n^{-bd_2^2}$$

for some constant c_1. Therefore, integrating times $\phi_{V_n}(t/\sqrt{n})$ provides an additive bound of order n^{-d^*}, where $d^* = bd_2^2 - p(d_1 + 1/2)$ and (for any d_0) d_2 can be chosen sufficiently large so that $d^* > d_0$.

Finally, on $\|t\| \ge \varepsilon\sqrt{n}$, Condition X1 [see (3.1)] gives an additive bound of η^n directly and, again (as on the previous interval), an additive error bounded by n^{-d_0} can be obtained.

Therefore, it now follows that we can choose d_0 (depending on d, d_1, d_2 and d^*) so that

$$P\left\{S_n + \frac{V_n}{\sqrt{n}} \in A\right\} = c_\pi \int \mathrm{Vol}(A)\phi_{\mathrm{Unif}(A)}(t)\phi_{\mathcal{N}(-G\delta,\tau(1-\tau)H)}(t)\phi_{V_n}\left(\frac{t}{\sqrt{n}}\right)dt$$
$$\times \left(1 + \mathcal{O}\left((\log^3(n)/n)^{1/2}\right)\right) + \mathcal{O}(n^{-d_0}),$$

from which Theorem 4 follows.

Finally, we show that the contribution of V_n can be ignored:

$$\begin{aligned}|P\{\tilde{S}_n \in A_h\} - P\{S_n^* \in A_h\}| &= |P\{\tilde{S}_n \in A_h\} - P\{\tilde{S}_n + V_n \in A_h + V_n\}| \\ &\le P\{\tilde{S}_n + V_n \in A_h \triangle (A_h + V_n)\},\end{aligned}$$

where $\triangle$ denotes the symmetric difference of the sets. Since V_n is bounded and $A_h = X_h R$, this symmetric difference is contained in a set, D, which is the union of $2p$ (boundary) parallelepipeds each of the form $X_h R_j$, where R_j is a rectangle one of whose coordinates has width $2n^{-d_1}$ and all other coordinates have length 1. Thus, applying Theorem 4 (as proved for the set $A = D$),

$$\begin{aligned}|P\{\tilde{S}_n \in A_h\} - P\{S_n^* \in A_h\}| &\le P\{\tilde{S}_n + V_n \in D\} \\ &\le c\,\mathrm{Vol}(D) + \mathcal{O}(n^{-d_0}) \\ &\le c' n^{-d_1},\end{aligned}$$

where c and c' are constants, and d_1 may be chosen arbitrarily large.

6. Normal approximation with nearly root-n multiplicative error.

THEOREM 5. *Assume Conditions X1, X2, F and the regression quantile model of Section 1. Let δ be the argument of the density of $\hat{\delta}_n \equiv n^{-1/2}(\hat{\beta}(\tau) - \beta(\tau))$ and suppose*

$$\|\delta\| \le d\sqrt{\log(n)}$$

for some constant d. Then, uniformly in $\varepsilon \le \tau \le 1-\varepsilon$ (for $\varepsilon > 0$),

$$f_{\hat{\delta}_n}(\delta) = \varphi_\Sigma(\delta)\big(1 + \mathcal{O}\big((\log^3(n)/n)^{1/2}\big)\big),$$

where φ_Σ denotes the normal density with covariance $\Sigma_n = \tau(1-\tau)G_n^{-1}H_nG_n^{-1}$ with G_n and H_n given by (3.2) and (3.3).

PROOF. Recall the basic formula for the density (4.1):

$$f_{\hat{\delta}}(\delta) = n^{-p/2}\sum_{h\in\mathcal{H}} \det(X_h)P\{S_n \in A_h\}\prod_{i\in h} f_i\big(x_i'\beta + n^{-1/2}\delta\big).$$

By Theorem 4, ignoring the multiplicative and additive error terms given in this result and setting $c_\pi' = (2\pi)^{-p/2}$,

$$\begin{aligned}
P\{S_n \in A_h\} &= P\big\{Z_n \in A_h/\sqrt{n}\big\} \\
&= c_\pi'|H_n|^{-1/2}\int_{A_h/\sqrt{n}} \exp\bigg\{-\frac{1}{2}\big(z - G_n^{-1}\delta\big)'\frac{H_n^{-1}}{\tau(1-\tau)}\big(z - G_n^{-1}\delta\big)\bigg\}\,dz \\
&= c_\pi'|H_n|^{-1/2}\exp\bigg\{-\frac{1}{2}\delta'\Sigma_n^{-1}\delta\bigg\}\int_{A_h/\sqrt{n}} dz\big(1 + \mathcal{O}\big(n^{-1/2}\big)\big) \\
&= c_\pi' n^{-p/2}|X_h||H_n|^{-1/2}\exp\bigg\{-\frac{1}{2}\delta'\Sigma_n^{-1}\delta\bigg\}\big(1 + \mathcal{O}\big(n^{-1/2}\big)\big)
\end{aligned}$$

since z is bounded by a constant times $n^{-1/2}$ on $A_h/\sqrt{n}$ and the last integral equals $\mathrm{Vol}(A_h) = n^{-p/2}|X_h|$.

By Ingredient 4, the product is

$$\prod_{i\in h} f_i\big(x_i'\beta\big)\big(1 + \mathcal{O}\big(\|\delta\| n^{-1/2}\big)\big).$$

This gives the main term of the approximation as

$$\sum_{h\in\mathcal{H}} n^{-p}|X_h|^2\prod_{i\in h} f_i\big(x_i'\beta\big)|H_n|^{-1/2}\exp\bigg\{-\frac{1}{2}\delta'\Sigma_n^{-1}\delta\bigg\}.$$

The penultimate step is to apply results from matrix theory on adjoint matrices [specifically, the Cauchy–Binet theorem and the "trace" theorem; see, e.g., Gantmacher (1960), pages 9 and 87]: the sum above is just the trace of the pth adjoint of $(X'D_fX)$, which equals $\det(X'D_fX)$.

The various determinants combine (with the factor n^{-p}) to give $\det(\Sigma_n)^{-1/2}$, which provides the asymptotic normal density we want.

Finally, we need to combine the multiplicative and additive errors into a single multiplicative error. So consider $\|\delta\| \le d\sqrt{\log(n)}$ (for some constant d). Then, the asymptotic normal density is bounded below by n^{-cd} for some constant c.

Thus, since the constant d_0 (which depends on d_1, d_2, d^* and η) can be chosen so that the additive errors are smaller than $\mathcal{O}(n^{-cd-1/2})$, the error is entirely subsumed in the multiplicative factor: $(1+\mathcal{O}((\log^3(n)/n)^{1/2}))$. $\square$

7. The Hungarian construction.

We first prove Theorem 1, which provides the bivariate normal approximation.

PROOF OF THEOREM 1. The proof follows the development in Theorem 5. The first step treats the first (intercept) coordinate. Since the binomial expansions were omitted in the proof of Theorem 3, details for the trinomial expansion needed for the bivariate case here will be presented.

The binomial sum in the first coordinate of (4.2) will be split into the sum of observations in the intervals $[x_i'\hat{\beta}(0), x_i'\hat{\beta}(\tau_1))$, $[x_i'\hat{\beta}(\tau_1), x_i'\hat{\beta}(\tau_1+a_n))$ and $[x_i'\hat{\beta}(\tau_1+a_n), x_i'\hat{\beta}(1))$. The expected number of observations in each interval is within p of n times the length of the corresponding interval. Thus, ignoring an error of order $1/n$, we expand a trinomial with n observations and $p_1=\tau_1$ and $p_2=a_n$. Let (N_1, N_2, N_3) be the (trinomially distributed) number of observation in the respective intervals and consider $P^*\equiv P\{N_1=k_1, N_2=k_2, N_3=n-k_1-k_2\}$. We may take

$$k_1 = \mathcal{O}\bigl((n\log n)^{1/2}\bigr),\qquad k_2=\mathcal{O}\bigl(a_n(\log n)^{1/2}\bigr), \tag{7.1}$$

since these bounds are exceeded with probability bounded by n^{-d} for any (sufficiently large) d. So $P^*\equiv A\times B$, where

$$A=\frac{n!}{(np_1+k_1)!(np_2+k_2)!(n(1-p_1-p_2)-k_1-k_2)!},$$
$$B=p_1^{np_1+k+1}p_2^{np_2+k_2}(1-p_1-p_2)^{n(1-p_1-p_2)-k_1-k_2}.$$

Expanding (using Sterling's formula and some computation),

$$\begin{aligned}A=\frac{1}{2\pi}\exp\biggl\{&2+\Bigl(n+\frac12\Bigr)\log\Bigl(n+\frac1n\Bigr)\\
&-\Bigl(np_1+k_1+\frac12\Bigr)\log\Bigl(np_1+\frac{k_1+1}{np_1}\Bigr)\\
&-\Bigl(np_2+k_2+\frac12\Bigr)\log\Bigl(np_2+\frac{k_2+1}{np_2}\Bigr)\\
&-\Bigl(n(1-p_1-p_2)-k_1-k_2+\frac12\Bigr)\\
&\quad\times\log\Bigl(n(1-p_1-p_2)-\frac{k_1+k_2-1}{n(1-p_1-p_2)}\Bigr)+\mathcal{O}\Bigl(\frac{1}{np_2}\Bigr)\biggr\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\exp\biggl\{\frac{1}{2}\log n - np_1\log p_1 - \left(k_1+\frac{1}{2}\right)\log(np_1) \\
&\qquad - np_2\log p_2 - \left(k_2+\frac{1}{2}\right)\log(np_2) \\
&\qquad - n(1-p_1-p_2)\log(1-p_1-p_2) - \left(k_1+k_2+\frac{1}{2}\right) \\
&\qquad \times \log\bigl(n(1-p_1-p_2)\bigr) - \frac{k_1^2}{np_1} - \frac{k_2^2}{np_2} \\
&\qquad - \frac{(k_1+k_2)^2}{n(1-p_1-p_2)} + \mathcal{O}\left(\frac{k_2^3}{(np_2)^2}\right)\biggr\} \\
&= \frac{1}{2\pi}\exp\biggl\{-\log n - \left(np_1+k_1+\frac{1}{2}\right)\log p_1 - \left(np_2+k_2+\frac{1}{2}\right)\log p_2 \\
&\qquad - \left(n(1-p_1-p_2)-k_1-k_2+\frac{1}{2}\right)\log(1-p_1-p_2) \\
&\qquad - \frac{k_1^2}{np_1} - \frac{k_2^2}{np_2} - \frac{(k_1+k_2)^2}{n(1-p_1-p_2)} + \mathcal{O}\left(\frac{(log n)^{3/2}}{na_n^2}\right)\biggr\},
\end{aligned}
$$

$$
\begin{aligned}
B &= \exp\bigl\{(np_1+k_1)\log p_1 + (np_2+k_2)\log p_2 \\
&\qquad + \bigl(n(1-p_1-p_2)-k_1-k_2\bigr)\log(1-p_1-p_2)\bigr\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
A\times B &= \exp\biggl\{-\frac{1}{2}p_1 - \frac{1}{2}p_2 - \frac{1}{2}(1-p_1-p_2) \\
&\qquad - \frac{k_1^2}{np_1} - \frac{k_2^2}{np_2} - \frac{(k_1+k_2)^2}{n(1-p_1-p_2)} + \mathcal{O}\left(\frac{(log n)^{3/2}}{na_n^2}\right)\biggr\}.
\end{aligned}
$$

Some further simplification shows that $A \times B$ gives the usual normal approximation to the trinomial with a multiplicative error of $(1 + o(n^{-1/2}))$ [when k_1 and k_2 satisfy (7.1)].

The next step of the proof follows that of Theorem 4 (see Ingredient 3). Since the proof is based on expanding characteristic functions (which do not involve the inverse of the covariance matrices), all uniform error bounds continue to hold. This extends the result of Theorem 4 to the bivariate case:

$$
\begin{aligned}
&P\{S_n(\tau_1)\in A_{h_1}, S_n(\tau_2)\in A_{h_2}\} \\
&\quad = P\{Z_1\in A_{h_1}/\sqrt{n}, Z_2\in A_{h_2}/\sqrt{n}\} \\
&\quad = P\{Z_1\in A_{h_1}/\sqrt{n}\}\times P\{(Z_2-Z_1)/\sqrt{n}\in (A_{h_2}-Z_2)/\sqrt{n}|Z_1\}
\end{aligned}
\tag{7.2}
$$

for appropriate normally distributed (Z_1, Z_2) (depending on n). This last equation is needed to extend the argument of Theorem 5, which involves integrating normal densities. The joint covariance matrix for $(S_n(\tau_1), S_n(\tau_2))$ is nearly singular (for $\tau_2 - \tau_1$ small) and complicates the bounds for the integral of the densities. The first factor above can be treated exactly as in the proof of Theorem 5, while the conditional densities involved in the second factor can be handled by simple rescaling. This provides the desired generalization of Theorem 5.

Thus, the next step is to develop the parameters of the normal distribution for $(B_n(\tau_1), R_n)$ [see (3.4), (3.5)] in a usable form. The covariance matrix for $(B_n(\tau_1), B_n(\tau_2))$ has blocks of the form

$$\operatorname{Cov}\bigl(B_n(\tau_1), B_n(\tau_2)\bigr) = \begin{pmatrix} \tau_1(1-\tau_1)\Lambda_{11} & \tau_1(1-\tau_2)\Lambda_{12} \\ \tau_1(1-\tau_2)\Lambda_{21} & \tau_2(1-\tau_2)\Lambda_{22} \end{pmatrix},$$

where $\Lambda_{ij} = G_n^{-1}(\tau_i) H_n G_n^{-1}(\tau_j)$ with G_n and H_n given in Condition X2 [see (3.2) and (3.3)].

Expanding $G_n(\tau)$ about $\tau = \tau_1$ (using the differentiability of the densities from Condition F),

$$\Lambda_{ij} = \Lambda_{11} + (\tau_2 - \tau_1)\Delta_{ij} + o\bigl(|\tau_2 - \tau_1|\bigr),$$

where Δ_{ij} are derivatives of G_n at τ_1 (note that $\Delta_{11} = 0$). Straightforward matrix computation now yields the joint covariance for $(B_n(\tau_1), R_n)$:

$$\operatorname{Cov}\bigl(B_n(\tau_1), R_n\bigr) = \begin{pmatrix} \tau_1(1-\tau_1)\Lambda_{11} & (\tau_2-\tau_1)\Delta_{12}^* \\ (\tau_2-\tau_1)\Delta_{21}^* & (\tau_2-\tau_1)\Delta_{22}^* \end{pmatrix} + o\bigl(|\tau_2 - \tau_1|\bigr), \tag{7.3}$$

where Δ_{ij}^* are uniformly bounded matrices.

Thus, the conditional distribution of $R_n = \sqrt{(\tau_2 - \tau_1)}(B_n(\tau_2) - B_n(\tau_1))$ given $B_n(\tau_1)$ has moments

$$E\bigl[R_n | B_n(\tau_1)\bigr] = (\tau_2 - \tau_1)\Lambda_{11}^{-1}\Delta_{12}/\bigl(\tau_1(1-\tau_1)\bigr), \tag{7.4}$$

$$\operatorname{Cov}\bigl[R_n | B_n(\tau_1)\bigr] = (\tau_2 - \tau_1)\biggl[\Delta_{22}^* - \frac{\tau_2 - \tau_1}{\tau_1(1-\tau_1)}\Delta_{21}^*\Lambda_{11}^{-1}\Delta_{12}^*\biggr] \tag{7.5}$$

and analogous equations also hold for $\{Z_2 - Z_1 | Z_1\}$.

Finally, recalling that $\tau_2 - \tau_1 = a_n$, the second term in (7.2) can be written

$$P\biggl\{\frac{Z_2 - Z_1}{\sqrt{n}} \in \frac{A_{h_2} - Z_1}{\sqrt{n}}\Big| Z_1\biggr\} = P\biggl\{\frac{Z_2 - Z_1}{\sqrt{n(\tau_2 - \tau_1)}} \in \frac{A_{h_2} - Z_1}{\sqrt{n a_n}}\Big| Z_1\biggr\}.$$

Thus, since the conditional covariance matrix is uniformly bounded except for the $a_n = (\tau_2 - \tau_1)$ factor, the argument of Theorem 5 also applies directly to this conditional probability. □

Finally, the above results are used to apply the quantile transform for increments between dyadic rationals inductively in order to obtain the desired "Hungarian" construction. The proof of Theorem 2 is as follows:

PROOF OF THEOREM 2. (i) Following the approach in Einmahl (1989), the first step is to provide the result of Theorem 1 for conditional densities one coordinate at a time. Using the notation of Theorem 1, let $\tau_1 = k/2^\ell$ and $\tau_2 = (k+1)/2^\ell$ be successive dyadic rationals (between ε and $1-\varepsilon$) with denominator 2^ℓ. So $a_n = 2^{-\ell}$. Let R_m be the mth coordinate of $R_n(\tau_1, \tau_2)$ [see (3.5)], let $\dot{R}_m$ be the vector of coordinates before the mth one, and let $S = B_n(\tau_1)$. Then the conditional density of $R_m|(\dot{R}_m, S)$ satisfies

$$f_{R_m|(\dot{R}_m,S)}(r_1|r_2,s) = \varphi_{\mu,\Sigma}(r_1|r_2,s)\left(1+\mathcal{O}\left(\frac{(\log n)^{3/2}}{\sqrt{n}}\right)\right) \tag{7.6}$$

for $\|r_1\| < D\sqrt{\log n}$, $\|r_2\| < D\sqrt{\log n}$, and $\|s\| < D\sqrt{\log n}$, and where μ and σ are easily derived from (7.4) and (7.5). Note that μ has the form

$$\mu = \sqrt{a_n}\alpha' S, \tag{7.7}$$

where $\|\alpha\|$ can be bounded (independent of n) and Σ can be bounded away from zero and infinity (independent of n).

This follows since the conditional densities are ratios of marginal densities of the form $f_Y(y) = \int f_{X,Y}\,dx$ (with $f_{X,Y}$ satisfying Theorem 1). The integral over $\|x\| \le D\sqrt{\log n}$ has the multiplicative error bound directly. The remainder of the integral is bounded by n^{-d}, which is smaller than the normal integral over $\|x\| \le D\sqrt{\log n}$ (see the end of the proof of Theorem 5).

(ii) The second step is to develop a bound on the (conditional) quantile transform in order to approximate an asymptotic normal random variable by a normal one. The basic idea appears in Einmahl (1989). Clearly, from (7.6),

$$\int_0^r f_{R_m|(\dot{R}_m,S)}(u|r_2,s)\,du = \int_0^r \varphi_{\mu,\sigma}(u|r_2,s)\,du\left(1+\mathcal{O}\left(\frac{(\log n)^{3/2}}{\sqrt{n}}\right)\right)$$

for $\|u\| < D\sqrt{\log n}$, $\|r_2\| < D\sqrt{\log n}$, and $\|s\| < D\sqrt{\log n}$. By Condition F, the conditional densities (of the response given x) are bounded above zero on $\varepsilon \le \tau \le 1-\varepsilon$. Hence, the inverse of the above versions of the c.d.f.'s also satisfy this multiplicative error bound, at least for the variables bounded by $D\sqrt{\log n}$. Thus, the quantile transform can be applied to show that there is a normal random variable, Z^*, such that $(R_m - Z^*) = \mathcal{O}((\log n)^{3/2}/\sqrt{n})$ so long as R_m and the quantile transform of R_m are bounded by $D\sqrt{\log n}$. Using the conditional mean and variance [see (7.7)], and the fact that the random variables exceed $D\sqrt{\log n}$ with probability bounded by n^{-d} (where d can be made large by choosing D large enough), there is a random variable Z_m that can be chosen independently so that

$$R_m = a_n\alpha' S + Z_m + \mathcal{O}\left(\frac{(\log n)^{3/2}}{\sqrt{n}}\right) \tag{7.8}$$

except with probability bounded by n^{-d}.

(iii) Finally, the "Hungarian" construction will be developed inductively. Let $\tau(k,\ell) = k/2^\ell$ and consider induction on ℓ. First consider the case where $\tau \geq \frac{1}{2}$; the argument for $\tau < \frac{1}{2}$ is entirely analogous.

Define $\varepsilon_n^* = c(\log n)^{3/2}/\sqrt{n}$, where c bounds the big-O term in any equation of the form (7.8). Let A be a bound [uniform over $\tau \in (\varepsilon, 1-\varepsilon)$] on α in (7.8). The induction hypothesis is as follows: there are normal random vectors $Z_n(k,\ell)$ such that

$$\left\| B_n\left(\frac{k}{2^\ell}\right) - Z_n(k,\ell) \right\| \leq \varepsilon(\ell) \tag{7.9}$$

except with probability $2\ell n^{-d}$, where for each ℓ, $Z_n(\cdot,\ell)$ has the same covariance structure as $B_n(\cdot/2^\ell)$, and where

$$\varepsilon(\ell) = \ell \varepsilon_n^* \prod_{j=1}^{\ell} (1 + A2^{-j/2}). \tag{7.10}$$

Note: since the earlier bounds apply only for intervals whose lengths exceed n^{-a} (for some positive a), ℓ must be taken to be smaller than $a \log_2(n) = \mathcal{O}(\log n)$. Thus, the bound in (7.10) becomes $\mathcal{O}((\log n)^{5/2}/\sqrt{n})$, as stated in Theorem 1.

To prove the induction result, note first that Theorem 1 (or Theorem 5) provides the normal approximation for $B_n(\frac{1}{2})$ for $\ell = 1$. The induction step is proved as follows: following Einmahl (1989), take two consecutive dyadic rationals $\tau(k,\ell)$ and $\tau(k-1,\ell)$ with k odd. So

$$\tau(k-1,\ell) = [k/2]/2^{\ell-1} = \tau\big([k/2], \ell-1\big).$$

Condition each coordinate of $B_n(\tau(k,\ell))$ on previous coordinates and on $B_n(\tau([k/2],\ell-1))$. Let $b_n(\tau(k,\ell)) = b_n(k/2^\ell)$ be one such coordinate.

Now, as above, define $R(k,\ell)$ by

$$b_n\big(\tau(k,\ell)\big) = b_n\big(\tau\big([k/2],\ell-1\big)\big) + R(k,\ell).$$

From (7.8), there is a normal random variable $Z_n(k,\ell)$ such that

$$\big|R(k,\ell) - \sqrt{2^{-\ell}}\alpha' B_n\big(\tau\big([k/2],\ell-1\big)\big) - Z_n(k,\ell)\big| \leq \varepsilon_n^*.$$

By the induction hypothesis for $(\ell-1)$, $B_n(\tau([k/2],\ell-1)$ is approximable by normal random variables to within $\varepsilon(\ell-1)$ (except with probability n^{-d}). Thus, a coordinate $b_n(\tau([k/2],\ell-1)$ is also approximable with this error, and the error in approximating $a_n\alpha' B_n(\tau([k/2],\ell-1)$ is bounded by $\varepsilon(\ell-1)$ times $A\sqrt{a_n} = A2^{-\ell/2}$. Finally, since $Z_n(k,\ell)$ is independent of these normal variables, the errors can be added to obtain

$$\big(1 + A2^{-\ell/2}\big)\varepsilon(\ell - 1) + \varepsilon_n^*.$$

Therefore, except with probability less than $2(\ell - 1)n^{-d} + 2n^{-d} = 2\ell n^{-d}$, the induction hypothesis (7.9) holds with error

$$\begin{aligned}(\ell - 1)\varepsilon_n^* \prod_{j=1}^{\ell-1}\big(1 + 2^{-j/2}\big) \times \big(1 + 2^{-\ell/2}\big) + \varepsilon_n^* \\ \le \ell \prod_{j=1}^{\ell}\big(1 + 2^{-j/2}\big)\varepsilon_n^* = \varepsilon(\ell),\end{aligned}$$

and the induction is proven.

The theorem now follows since the piecewise linear interpolants satisfy the same error bound [see Neocleous and Portnoy (2008)]. □

APPENDIX

RESULT 1. *Under the conditions for the theorems here, the coverage probability for the confidence interval* (2.3) *is* $1 - 2\alpha + \mathcal{O}((\log n)n^{-2/3})$, *which is achieved at* $h_n = c\sqrt{\log n}\, n^{-1/3}$ (*where* c *is a constant*).

SKETCH OF PROOF. Recall the notation of Remark 2 in Section 2. Using Theorem 1 and the quantile transform as described in the first steps of Theorem 2 (and not needing the dyadic expansion argument), it can be shown that there is a bivariate normal pair (W, Z) such that

$$\begin{aligned}\sqrt{n}\big(\hat\beta(\tau) - \beta(\tau)\big) &= W + R_n, \qquad R_n = \mathcal{O}_p\big(n^{-1/2}(\log n)^{3/2}\big), \\ \sqrt{n}\big(\hat\Delta(h_n) - \Delta(h_n)\big) &= Z + R_n^*, \qquad R_n^* = \mathcal{O}_p\big(n^{-1/2}(\log n)^{3/2}\big).\end{aligned} \tag{A.1}$$

Note that from the proofs of Theorems 1 and 2, the $\mathcal{O}_p$ terms above are actually $\mathcal{O}$ terms except with probability n^{-d} where d is an arbitrary fixed constant. The "almost sure" results above take $d > 1$, but $d = 1$ will suffice for the bounds on the coverage probability here.

Incorporating the approximation error in (A.1),

$$\sqrt{n}(\hat\delta - \delta) = Z/h_n + R_n^*/h_n + \mathcal{O}\big(n^{1/2}h_n^2\big).$$

Now consider expanding $s_a(\delta)$. First, note that under the design conditions here, s_a will be of exact order $n^{-1/2}$; specifically, if X is replaced by $\sqrt{n}\tilde X$, all terms involving $\tilde X'\tilde X$ will remain bounded, and we may focus on $\sqrt{n}s_a(\delta)$. Note also that for $h_n = \mathcal{O}(n^{-1/3})$, the terms in the expansion of $(\hat\delta - \delta)$ tend to zero [specifically,

$1/(\sqrt{n}h_n) = \mathcal{O}(n^{-1/6})$]. So the sparsity, $s_a(\delta)$, may be expanded in a Taylor series as follows:

$$\begin{aligned}\sqrt{n}s_a(\hat{\delta}) &= \sqrt{n}s_a(\delta) + b_1'(\hat{\delta}-\delta) + b_2(\hat{\delta}-\delta) + b_3(\hat{\delta}-\delta) + \mathcal{O}(n^{-2/3})\\ &\equiv \sqrt{n}s_a(\delta) + K,\end{aligned}$$

where b_1 is a (gradient) vector that can be defined in terms of $\tilde{X}$ and $\beta(\tau)$ (and its derivatives), b_2 is a quadratic function (of its vector argument) and b_3 is a cubic function. Note that under the design conditions, all the coefficients in b_1, b_2 and b_3 are bounded, and so it is not hard to show that all the terms in K tend to zero as long as $h_n\sqrt{n} \to \infty$. Specifically, if h_n is of order $n^{-1/3}$, then all the terms in K tend to zero. Also, R_n^* is within a $\log n$ factor of $\mathcal{O}(n^{-1/2})$ and h_n^2 is even smaller. Finally, Z is a difference of two quantiles separated by $2h$, and so $b_1'Z$ has variance proportional to h. Thus, $E(b_1'Z/(\sqrt{n}h_n))^2 = \mathcal{O}(1/(nh_n))$. Thus, not only does $b_1'Z/(\sqrt{n}h_n) \to^p 0$, but powers of this term greater than 2 will also be $\mathcal{O}_p(n^{-1})$.

It follows that the coverage probability may be computed using only two terms of the Taylor series expansion for the normal c.d.f.:

$$\begin{aligned}&P\{\sqrt{n}a'(\hat{\beta}(\tau)-\beta(\tau)) \le z_\alpha\sqrt{n}s_a(\hat{\delta})\}\\ &\quad = P\{a'(W+R_n) \le z_\alpha\sqrt{n}s_a(\hat{\delta}) + K\}\\ &\quad = E\Phi_{a'W|Z}(z_\alpha\sqrt{n}s_a(\delta) + K - a'R_n)\\ &\quad = E\{\Phi_{a'W|Z}(\sqrt{n}s_a(\delta)) + \phi_{a'W|Z}(\sqrt{n}s_a(\delta))(K - a'R_n)\\ &\qquad\quad + \tfrac{1}{2}\phi'_{a'W|Z}(\sqrt{n}s_a(\delta))(K-a'R_n)^2 + \mathcal{O}((\log n)^3/n)\}\\ &\quad \equiv 1 - \alpha + T_1 + T_2 + \mathcal{O}((\log n)^3/n).\end{aligned}$$

Note that the (normal) conditional distribution of W given Z is straightforward to compute (using the usual asymptotic covariance matrix for quantiles): the conditional mean is a small constant (of the order of h_n) times Z, and the conditional variance is bounded.

Expanding the lower probability in the same way and subtracting provides some cancelation. The contribution of R_n will cancel in the T_1 differences, and is negligible in subsequent terms since $R_n^2 = \mathcal{O}((\log n)^3/n)$. Similarly, the $R_n^*/(\sqrt{n}h_n)$ term will appear only in the T_1 difference where it contributes a term that is $(\log n)^{3/2}$ times a term of order $1/(nh_n)$, and will also be negligible in subsequent terms. Also, the h_n^2 term will only appear in T_1, as higher powers will be negligible. The only remaining terms involve $Z/(\sqrt{n}h_n))$. For the first power (appearing in T_1), $EZ = 0$. For the squared Z-terms in T_2, since $\mathrm{Var}(b_1'Z)$ is proportional to h_n, $E(b_1'Z)^2/(nh_n^2) = c_1/(nh_n)$, and all other terms involving Z have smaller order.

Therefore, one can obtain the following error for the coverage probability: for some constants c_1 and c_2, the error is

$$\frac{b_1' R_n^*}{\sqrt{n} h_n} + \frac{c_1}{n h_n} + c_2 h_n^2$$

(plus terms of smaller order). Since R_n^* is of order nearly $n^{-1/2}$, the first terms have nearly the same order. Using $b_1' R_n^* = c(\log n)/(\sqrt{n} h_n)$, it is straightforward to find the optimal h_n to be a constant times $\sqrt{\log n} n^{-1/3}$, which bounds the error in the coverage probability by $\mathcal{O}(\log n n^{-2/3})$. $\square$

REFERENCES

BERNSTEIN, S. N. (1964). On a modification of Chebyshev's inequality and of the error formula of Laplace. In *Sobranie Sochineniĭ* **4** 71–79. Nauka, Moscow [original publication: *Ann. Sci. Inst. Sav. Ukraine, Sect. Math.* **1** (1924)].

DANIELS, H. E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55** 37–48. MR0962940

DE ANGELIS, D., HALL, P. and YOUNG, G. A. (1993). Analytical and bootstrap approximations to estimator distributions in L^1 regression. *J. Amer. Statist. Assoc.* **88** 1310–1316. MR1245364

EINMAHL, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68. MR0996984

GANTMACHER, F. R. (1960). *Matrix Theory*. Amer. Math. Soc., Providence, RI.

GUTENBRUNNER, C., JUREČKOVÁ, J., KOENKER, R. and PORTNOY, S. (1993). Tests of linear hypotheses based on regression rank scores. *J. Nonparametr. Stat.* **2** 307–331. MR1256383

HALL, P. and SHEATHER, S. J. (1988). On the distribution of a Studentized quantile. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **50** 381–391. MR0970974

HE, X. and HU, F. (2002). Markov chain marginal bootstrap. *J. Amer. Statist. Assoc.* **97** 783–795. MR1941409

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

HOROWITZ, J. L. (1998). Bootstrap methods for median regression models. *Econometrica* **66** 1327–1351. MR1654307

JUREČKOVÁ, J. and SEN, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*. Wiley, New York. MR1387346

KNIGHT, K. (2002). Comparing conditional quantile estimators: First and second order considerations. Technical report, Univ. Toronto.

KOCHERGINSKY, M., HE, X. and MU, Y. (2005). Practical confidence intervals for regression quantiles. *J. Comput. Graph. Statist.* **14** 41–55. MR2137889

KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657

KOENKER, R. (2012). `quantreg`: Quantile regression. R-package, Version 4.79. Available at cran.r-project.org.

KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883

NEOCLEOUS, T. and PORTNOY, S. (2008). On monotonicity of regression quantile functions. *Statist. Probab. Lett*. **78** 1226–1229. MR2441467

PARZEN, M. I., WEI, L. J. and YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81** 341–350. MR1294895

ZHOU, K. Q. and PORTNOY, S. L. (1996). Direct use of regression quantiles to construct confidence sets in linear models. *Ann. Statist*. **24** 287–306. MR1389891

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN
725 S. WRIGHT
CHAMPAIGN, ILLINOIS 61820
USA
E-MAIL: sportnoy@illinois.edu