The Annals of Statistics
1996, Vol. 24, No. 1, 287–306

DIRECT USE OF REGRESSION QUANTILES TO CONSTRUCT CONFIDENCE SETS IN LINEAR MODELS

BY KENNETH Q. ZHOU AND STEPHEN L. PORTNOY[1]

University of Illinois, Urbana–Champaign

Direct use of the empirical quantile function provides a standard distribution-free approach to constructing confidence intervals and confidence bands for population quantiles. We apply this method to construct confidence intervals and confidence bands for regression quantiles and to construct prediction intervals based on sample regression quantiles. Comparison of the direct method with the studentization and the bootstrap methods are discussed. Simulation results show that the direct method has the advantage of robustness against departure from the normality assumption of the error terms.

1. Introduction. Using distribution-free statistics to test hypotheses and to construct confidence intervals has become a popular robust approach for statistical inferences. Here, sample regression quantiles will be used directly to construct confidence intervals and confidence bands for regression quantile functions and to construct prediction intervals for future response variables. Although regression quantiles were introduced by Koenker and Basset (1978) just 15 years ago, the sample quantile is a much older concept. During the past three decades, various approaches were proposed for constructing confidence intervals and confidence bands for quantiles. Since regression quantiles are a natural generalization of sample quantiles from the location model to the linear model, it is worthwhile to give a brief review of those used in the sample quantile case.

We especially consider three approaches: studentization, the bootstrap approach and the direct (or distribution-free) approach. For convenience, we assume $X_1, \dots, X_n$ is a random sample from a population with distribution function F, density f and quantile F^{-1}. Let $X_{(1)}, \dots, X_{(n)}$ be the corresponding order statistics, and let $\hat{F}$ ($\hat{F}^{-1}$) be the empirical distribution (quantile) function. The approximated studentized confidence interval for $F^{-1}(\theta)$ for some $\theta \in (0,1)$ has the standard form $\hat{F}^{-1}(\theta) \pm s z_\alpha \sqrt{\theta(1-\theta)/n}$, where s is an $n^{1/4}$-consistent estimator of the *sparsity function* $1/f(F^{-1}(\theta))$. Note that the notation z_α and s will be used through the end of the paper. Hence, to obtain a confidence interval for $F^{-1}(\theta)$, one has to obtain a consistent estimator of $1/f(F^{-1}(\theta))$, which is often hard. An alternative approach that

Received March 1994; revised May 1995.

[1]Reserach partially supported by NSF Grant SES-91-21776.

AMS 1991 *subject classifications*. 62G15, 62E20, 62J05

Key words and phrases. Conditional quantiles, Bahadur representation, confidence intervals, confidence bands, empirical levels, Brownian bridge.

circumvents the problem of estimating the sparsity function is the bootstrap. However, the bootstrap approach has been considered by several authors to be unsatisfying in constructing confidence intervals for quantiles. See Hall (1992).

The direct method is another approach that does not need to deal with the sparsity function and is especially interesting because of its efficiency of computation. The idea is simple. For some $\theta \in (0, 1)$, we use two order statistics $(X_{(p)}, X_{(q)})$ to be a confidence interval for $F^{-1}(\theta)$, where p and q are two integers chosen in such a way that $P[X_{(p)} \le F^{-1}(\theta) \le X_{(q)}] \ge 1 - 2\alpha$, where computation of p and q can be carried out by using tables of binomial probabilities, by using tables of the incomplete beta functions or by large-sample approximation. Detailed procedures are discussed in Serfling (1980) among many others. In the case of constructing confidence bands for quantiles, Cšorgő and Révész (1984) discussed two approaches: one is studentization, another is distribution free (here we call it the direct approach). The first approach leads to a confidence band $(\hat{F}_n^{-1}(\theta) \pm sc_\alpha(\theta(1-\theta))^{1/2} n^{-1/2}, \varepsilon_n \le \theta \le 1 - \varepsilon_n)$, where c_α is the $(1 - 2\alpha)$th percentile of the supremum of a Brownian bridge and s is defined above. The second approach results in a confidence band $\{\hat{F}_n^{-1}(\theta \pm c_\alpha(\theta(1-\theta))^{1/2} n^{-1/2}), \varepsilon_n \le \theta \le 1 - \varepsilon_n\}$. They show that the two confidence bands are asymptotically equivalent. All of the approaches discussed above can be generalized to the case of regression quantiles. We shall emphasize the asymptotic properties of direct confidence intervals and confidence bands and their equivalence to the studentized versions.

The model we consider is

$$y_i = x_i' \beta + u_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where the p-dimensional design vectors $x_i = (1, \tilde{x}_i)$ are partitioned corresponding to $\beta = (\beta_0, \beta_1)$, in which β_0 is the intercept parameters. Error terms $u_1, \dots, u_n$ are assumed to be i.i.d. according to F with median 0. Now the θth sample regression quantiles are defined as the vector $b \in R^p$ minimizing $\sum_{i=1}^n \rho_\theta(y_i - x_i' b)$, where $\rho_\theta(\mu) = \theta\mu^+ + (1-\theta)\mu^-$. When $x_i = 1$, the (sample) regression quantile is simply the sample quantile. From this definition, it is not difficult to see that the population analog is $\beta(\theta) = \beta + \mathbf{e}_1 F^{-1}(\theta)$. As shown in Bassett and Koenker (1982), this important parameters configuration relates directly to the concept of conditional quantile function of y given x,

$$Q_y(\theta|x) = x'\beta(\theta) = x'\beta + F^{-1}(\theta). \tag{1.2}$$

A wide literature has been published on developing methods based on regression quantiles. Extensive reviews can be found in some recent papers: Koenker and Portnoy (1987), Portnoy and Koenker (1989) and Gutenbrunner, Jurečková, Koenker and Portnoy [(1993), GJKP hereafter]. In the next section, we will introduce confidence intervals for the conditional quantile function $x'\beta(\theta)$ for fixed x and θ and then construct prediction intervals for future response variables $\{Y_{n+i}\}$. Section 3 is devoted to discussion of confidence bands for regression quantiles. We shall discuss our simulation results in Section 4 and offer some conclusions and remarks in Section 5.

2. Confidence intervals for conditional quantiles. Fix θ, $\alpha \in (0,1)$ and let z_α denote the $(1-\alpha)$th standard normal percentile point. The studentized approach leads to a $(1-2\alpha)100\%$ confidence interval for $x'\beta(\theta)$ of the form

$$I_{1n} = \left(x'\hat{\beta}(\theta) - a_n,\ x'\hat{\beta}(\theta) + a_n\right), \tag{2.1}$$

where $a_n = sz_\alpha\sqrt{x'Q^{-1}x\theta(1-\theta)}\,/\sqrt{n}$ and s is the consistent estimate of the sparsity function described in Section 1. The direct approach yields a confidence interval of the form

$$I_{2n} = \left(x'\hat{\beta}(\theta - b_n),\ x'\hat{\beta}(\theta + b_n)\right), \tag{2.2}$$

where $b_n = z_\alpha\sqrt{x'Q^{-1}x\theta(1-\theta)}\,/\sqrt{n}$. The question is: How do we compare I_{2n} with I_{1n}?

Before stating the results in Corollary 2.1, we shall verify the asymptotic properties of $\hat{\beta}(\theta + b_n)$ and $\hat{\beta}(\theta)$ under some regularity conditions on the design matrix X and the distribution function F. The conditions are as follows:

(F1) F is twice differentiable at $F^{-1}(\theta)$ and $f(F^{-1}(\theta)) = F'(F^{-1}(\theta)) > 0$;
(X1) $\max_{ij}|x_{i,j}| = O(n^{1/4})$;
(X2) $\sum_{i=1}^n \|x_j\|^3 = O(n)$;
(X3) $n^{-1}\sum_{i=1}^n x_i x_i' = Q + O(n^{-1/4}\log n)$, where Q is a positive definite $p \times p$ matrix.

By equivariance of regression quantiles, we set $\sum_{i=1}^n x_{ij} = 0$ for $j = 2,\dots,p$ without loss of generality. Also, to simplify the discussion, we introduce additional notations: $\Psi_\theta(x) = \theta - I(x < 0)$ and $\mathbf{e}_1 = (1,0,\dots,0)$. Now we establish a Bahadur representation.

THEOREM 2.1. *Assume that conditions* (F1) *and* (X1)–(X3) *hold. Let* $\mathbf{e}_1$ *and* $\Psi_\theta(\cdot)$ *be defined as above. If* k_n *is a sequence of constants such that* $k_n = \theta + kn^{-1/2} + O((\log n/n)^{3/4})$ *for some* $\theta \in (0,1)$ *and constant* k, *then, for* $n \to \infty$,

$$\begin{aligned}\hat{\beta}(k_n) - \beta(\theta) = {}& \frac{k\mathbf{e}_1 n^{-1/2}}{f(F^{-1}(\theta))} + \frac{n^{-1}Q^{-1}}{f(F^{-1}(\theta))}\sum_{i=1}^n x_i\Psi_\theta\left(u_i - F^{-1}(\theta)\right) \\ & + O\left(\left(\frac{\log n}{n}\right)^{3/4}\right).\end{aligned} \tag{2.3}$$

PROOF. Define $W_n(\mathbf{t},\theta) = (1/n)\sum_{i=1}^n x_i\Psi_\theta(u_i - x_i\mathbf{t} - F^{-1}(\theta))$. Then, by Lemma A1 in the Appendix and the definition of k_n, we have

$$W_n\left(\left(\hat{\beta}(k_n) - \beta(\theta)\right),\theta\right) = k\mathbf{e}_1 n^{-1/2} + O\left(\left(\frac{\log n}{n}\right)^{3/4}\right) \quad \text{a.s.} \tag{2.4}$$

By Lemma A2, we have

$$\sup_{\|t\|\le M_n} \|W_n(\mathbf{t},\theta) - W_n(0,\theta) + f(F^{-1}(\theta))Q\mathbf{t}\| = O_p\big(n^{-3/4}(\log n)^{3/4}\big), \tag{2.5}$$

where M_n is defined in Lemma A2. Now, using (2.5) and the method of Jurečková [(1977), Lemma, 5.2], we have $\beta(k_n) - \beta(\theta) = O_p(n^{-1/2}(\log n)^{1/2})$. Therefore, by substituting $\hat\beta(k_n) - \beta(\theta)$ for $\mathbf{t}$ in (2.5) and then substituting (2.4) into (2.5), we have

$$\begin{aligned}\frac{k\mathbf{e}_1}{\sqrt{n}} + W_n(0,\theta) &= f(F^{-1}(\theta))Q\big(\hat\beta(k_n) - \beta(\theta)\big)\\ &\quad + O_p\big(n^{-3/4}(\log n)^{3/4}\big).\end{aligned} \tag{2.6}$$

□

Theorem 2.1 can be viewed as a generalization to linear models of a Bahadur representation of order statistics in the location model. When $k_n = \theta$ for all n, equation (2.3) reduces to the representation obtained by Ruppert and Carroll (1980) except for the asymptotic order of the remainder terms. Note that condition (F1) is the same as the condition in Theorem 2.5.1 of Serfling (1980). The conditions on the design X are standard in linear regression analysis. Samples from typical distributions like student-t with degrees of freedom larger than 4, gamma, beta, uniform, lognormal and normal will satisfy conditions (X1)–(X3).

COROLLARY 2.1. *Let I_{1n}, I_{2n}, a_n and b_n be defined in* (2.1) *and* (2.2). *Under conditions* (F1) *and* (X1)–(X3), $\forall\ x = (1,\tilde x)$, $\tilde x \in R^{p-1}$, $\forall\ \theta \in (0,1)$ *and* $n \to \infty$, *we have the following*:

(a) $$x'\hat\beta(\theta) \pm a_n = x'\hat\beta(\theta \pm b_n) + O_p\left(\left(\frac{\log n}{n}\right)^{3/4}\right);$$

(b) $$\sqrt{n}\big(x'\hat\beta(\theta\pm b_n) - x'\beta(\theta)\big) \to_{\mathrm{D}} N\left(\pm\frac{z_\alpha\sqrt{x'Q^{-1}x\theta(1-\theta)}}{f(F^{-1}(\theta))}, \frac{x'Q^{-1}x\theta(1-\theta)}{f^2(F^{-1}(\theta))}\right);$$

(c) $$\begin{aligned}P[x'\beta(\theta) \in I_{1n}] &= P[x'\beta(\theta) \in I_{2n}] + O\big(n^{-1/4}(\log n)^{3/4}\big)\\ &= 1 - 2\alpha + O\big(n^{-1/4}(\log n)^{3/4}\big).\end{aligned}$$

PROOF. The proof of parts (a) and (b) is a straightforward application of Theorem 2.1. See Zhou (1995) for details. As for part (c), follow the argument of Hall [(1992), Section 2.7] and apply Theorem 2.1 and the Berry–Esseen theorem for sums of independent but nonidentical random variables [Feller (1966), page 521]. □

Corollary 2.1 indicates that the coverage errors of I_{2n} are of order $O(n^{-1/4}(\log n)^{3/4})$; I_{1n} and I_{2n} are asymptotically equivalent when an $n^{1/4}$-consistent estimator of $1/f$ is used. Note that the computation of b_n is quite simple even when sample size is large. Portnoy (1991) shows that the number of breakpoints of regression quantile solutions is of order only $O(n \log n)$. Efficient algorithms and Fortran subroutines are described in Koenker and D'Orey (1987).

In many situations, a prediction interval for a future response variable is useful. The conventional method based on studentized least squared estimator requires the normality assumption on F. Here, the direct approach is applied to construct prediction intervals without this assumption. Let $\{x_i, y_i\}_1^n$ be the samples corresponding to model (1.1), and let Y_{n+1} be the future response variable. The $(1-2\alpha)100\%$ prediction confidence interval for Y_{n+1} is

$$I_{p_1} = \left[x'_{n+1}\hat{\beta}_n(\theta_1),\, x'_{n+1}\hat{\beta}_n(\theta_2) \right], \tag{2.7}$$

where $0 < \theta_1 < \theta_2 < 1$, $\theta_2 - \theta_1 = 1 - 2\alpha$ and x_{n+1} is the predictor assumed known. The subscript n of $\hat{\beta}(\theta)$ indicates that the estimation is based on past n observations. The following corollary gives the asymptotic order of coverage probability of I_{p_1}.

COROLLARY 2.2. *Let θ_1 and θ_2 be the two points defined above. Assume F is twice differentiable at θ_1 and θ_2 and has positive density at the two points. Also assume conditions* (X1)–(X3). *Then the asymptotic coverage probabilities of I_{p_1} are $1-2\alpha$ with errors of the order $O(n^{-1/2})$.*

PROOF. Apply Theorem 2.1 and Taylor's series expansion theorem. □

Sometimes, a confidence region for $Y_{n+1}, \dots, Y_{n+m}$ may be of interest. By using the Bonferroni principle, we extend the prediction confidence interval for Y_{n+1} to

$$I_{p_2} = \left[\begin{pmatrix} x'_{n+1}\hat{\beta}_n(\theta_1) \\ \vdots \\ x'_{n+m}\hat{\beta}_n(\theta_1) \end{pmatrix}, \begin{pmatrix} x'_{n+1}\hat{\beta}_n(\theta_2) \\ \vdots \\ x'_{n+m}\hat{\beta}_n(\theta_2) \end{pmatrix} \right], \tag{2.8}$$

which is a confidence region for $(Y_{n+1}, \dots, Y_{n+m})$, where $\theta_2 - \theta_1 = (1 - 2\alpha)^{1/m}$. Applying Corollary 2.2, we have

$$P\big((Y_{n+1}, \dots, Y_{n+m}) \in I_{p_2}\big) = 1 - 2\alpha + O(n^{-1/2}), \qquad n \to \infty. \tag{2.9}$$

The above results do not rely on the symmetry of F. However, when F is symmetric and θ_1 is equal to θ_2, we expect the coverage errors of the prediction intervals to be of smaller order than those when F is asymmetric because of the cancellation among the second- or higher-order terms of their

Bahadur representations. In addition, the confidence intervals I_{2n}, I_{p_1} and I_{p_2} defined above can be replaced by other terms: $I'_{2n} = (\hat{\beta}(\theta \pm b_n) + x'\hat{\beta}_1(\theta))$, $I'_{p_1} = (\hat{\beta}_0(\theta_1 - \delta_n) + \tilde{x}'_{n+1}\hat{\beta}_1(\theta_1),\ \hat{\beta}_0(\theta_2 + \delta_n) + \tilde{x}'_{n+1}\hat{\beta}_1(\theta_2))$ and I'_{p_2} which can be expressed in a similar way as I'_{p_1}. The asymptotic equivalence between I_{2n}, I_{p_1}, I_{p_2} and $I'_{2n}, I'_{p_1}, I'_{p_2}$ can be easily justified by showing that $\sqrt{n}\,(\tilde{x}'\hat{\beta}_1(\theta_n) - \tilde{x}\hat{\beta}_1(\theta)) = O(n^{-1/4}(\log n)^{3/4})$ when $\theta_n = \theta + O(n^{-1/2})$. However, this is an immediate consequence of Theorem 2.1. One advantage of this expression is that the confidence intervals are always well defined. Bassett and Koenker (1982) show the conditional quantile function $x'\hat{\beta}(\theta)$ is not necessarily a monotone function of θ except at $x = \bar{x}$.

So far, we have shown that the studentization and the direct confidence intervals are asymptotically equivalent. It is of theoretical interest to go further to obtain the exact order of coverage errors of the confidence intervals. For the studentized confidence interval I_{1n}, estimating $1/f$ is crucial to the coverage error of the resulting confidence intervals. A natural estimator of $1/f$ is $(\hat{F}_n^{-1}(\theta + h) - \hat{F}_n^{-1}(\theta - h))/2h$ whose convergence rate depends on the bandwidth h. It has been shown by Bofinger (1975) that the estimator has an optimal covergence rate when $h = O(n^{-1/5})$. However, Hall and Sheather (1988) show that the optimal convergence rate of an estimator of $1/f$ does not imply an optimal coverage probability of the corresponding confidence interval. By using Edgeworth expansions, they show that in location models when $h = O(n^{-1/3})$, the coverage error of $\hat{I}_{1n}$ is of order $O(n^{-2/3})$ and is only $O(n^{-2/5})$ when the Bofinger estimator is used. The coverage errors of bootstrap confidence intervals have also been investigated by some researchers. Hall (1992) shows that ordinary percentile confidence intervals for quantiles have coverage errors of order $O(n^{-1/2})$. Also Falk and Kaufmann (1991) show that the coverage probabilities of backward bootstrap confidence intervals of quantile are higher than that of ordinary percentile confidence intervals and are exactly the same as that of the direct confidence intervals. Now one tends to conjecture that the above conclusion is also true in the case of conditional quantiles. However, Hall (1992) implies that the unsmoothed bootstrap percentile confidence intervals have coverage error of order no lower than $n^{-1/4}$ and both smoothed bootstrap and studentized confidence intervals for the 0.5th regression quantiles have coverage errors of order no lower than $n^{-2/5}$. Reviewing the above results, we see that the confidence intervals for 0.5th regression quantile functions have convergence order no higher than the 0.5th quantiles. The exact convergence order of the coverage errors of the direct confidence intervals may be obtained by deriving Edgeworth expansions of $P(x'\hat{\beta}(\theta - cn^{1/2}) \le x'\beta(\theta) \le x'\hat{\beta}\,(\theta + cn^{1/2}))$ for fixed x and θ. These are not in the scope of this paper, but we conjecture that, in the case of quantile regression, the coverage errors of the direct confidence intervals have convergence order no lower than both studentized and bootstrap confidence intervals. It is also reasonable to expect that the backward bootstrap approach performs better than the ordinary percentile method for conditional quantiles. We shall compare the performance of these methods in our simulation study in Section 4.

3. Confidence bands for conditional quantiles. First, we are interested in constructing confidence bands for a class of conditional quantile functions $\{x'\beta(\theta),\ x \in \mathscr{X}\}$, where $\mathscr{X}$ is a set of x with θ fixed. A popular way to construct confidence bands for $S^* = \{x'\beta(\theta),\ x \in R^p - \{0\}\}$ is the so-called S-method (or Scheffé's method). The following proposition gives the result.

PROPOSITION 3.1. *Let S^* be the class defined above. Denote the $(1-\alpha)$ 100% level confidence band for the function $x'\beta(\theta)$, $x \in S^*$, as J_{1n}. Then under conditions* (F1) *and* (X1)–(X3),

$$J_{1n} = \left\{ x'\hat{\beta}(\theta) \pm \left[\frac{\chi^2_{p,\alpha}\theta(1-\theta)}{nf^2\big(F^{-1}(\theta)\big)} x'Q^{-1}x \right]^{1/2} \right\}, \tag{3.1}$$

for any $\theta \in (0,1)$ where $\chi^2_{p,\alpha}$ is the $(1-\alpha)$th percentile of the chi-square distribution with p degrees of freedom.

The proposition follows easily from the Cauchy–Schwarz inequality and Theorem 4.2 of Koenker and Basset (1978). Note that the confidence band given by the S-method involves the sparsity function $1/f$. Now, the confidence band given by the direct method is

$$J_{2n} = \left\{ x'\hat{\beta}\left(\theta \pm \left(\chi^2_{p,\alpha} x'Q^{-1}x\theta(1-\theta)/n\right)^{1/2}\right);\ \forall\ x \in S \right\}, \tag{3.2}$$

which is distribution free. Here we need to impose some constraint on the class S. Obviously, by the definition of the k_n in Theorem 2.1, $\sqrt{x'Q^{-1}x/n}$ should be of order $O(n^{-1/2})$. Set $D_K = \{x = (1, \tilde{x}),\ \tilde{x} \in R^{p-1},\ \sqrt{x'Q^{-1}x} \le K\}$, where K is a constant not depending on n. The following theorem presents a uniform representation of $x'\hat{\beta}(\theta \pm n^{-1/2}c\sqrt{x'Q^{-1}x\theta(1-\theta)}\,) - x'\beta(\theta)$ for all $x \in D_K$ and some constant c.

THEOREM 3.1. *Let D_K be the set just defined, and let c be a constant. Under the same conditions as those in Theorem* 2.1, *we have*

$$\begin{aligned} x'\hat{\beta}&\left(\theta \pm \frac{c\sqrt{x'Q^{-1}x\theta(1-\theta)}}{n^{1/2}}\right) - x'\beta(\theta) \\ &= \pm\frac{c\sqrt{x'Q^{-1}x\theta(1-\theta)}}{n^{1/2}f\big(F^{-1}(\theta)\big)} + \frac{x'Q^{-1}}{nf\big(F^{-1}(\theta)\big)} \sum_{i=1}^{n} x_i\Psi_\theta\big(u_i - F^{-1}(\theta)\big) \\ &\quad + O_p\big(n^{-3/4}(\log n)^{3/4}\big) \end{aligned}$$

uniformly for all $x \in D_K$ as $n \to \infty$.

PROOF. Let $\lambda_{\min}$ be the smallest eignevalue of Q. Note that $x'\mathbf{e}_1 = 1$. Then

$$\sup_{x \in D_K} \Bigg| x'\hat{\beta}\Big(\theta + n^{-1/2} c\sqrt{x'Q^{-1}x\theta(1-\theta)}\Big) - x'\beta(\theta) - \frac{c\sqrt{x'Q^{-1}x\theta(1-\theta)}}{n^{1/2} f(F^{-1}(\theta))} - n^{-1} f^{-1}(F^{-1}(\theta)) x'Q^{-1} \sum_{i=1}^{n} x_i' \Psi_\theta\big(u_i - F^{-1}(\theta)\big) \Bigg|$$

$$\leq \sup_{x \in D_K} \|x\| \times \Bigg\| \hat{\beta}\Big(\theta + n^{-1/2} c\sqrt{x'Q^{-1}x\theta(1-\theta)}\Big) - \beta(\theta) - \frac{c\sqrt{x'Q^{-1}x\theta(1-\theta)}}{n^{1/2} f(F^{-1}(\theta))} - n^{-1} f^{-1}(F^{-1}(\theta)) Q^{-1} \sum_{i=1}^{n} x_i' \Psi_\theta\big(u_i - F^{-1}(\theta)\big) \Bigg\|.$$

Note the first factor $\sup_{x \in D_K} \|x\| \leq K$. By Theorem 2.1, the second norm is of order $O(n^{-3/4}(\log n)^{3/4})$. Thus Theorem 3.1 follows. □

The following corollary shows that confidence band J_{2n} in (3.2) with S replaced by D_K has coverage probability at least $(1 - 2\alpha)100\%$ for $\{x'\beta(\theta), x \in D_K\}$.

COROLLARY 3.1. *Assume the conditions of Theorem* 3.1. *For any fixed* $\theta \in (0, 1)$,

$$P\big[x'\beta(\theta) \in J_{2n},\ x \in D_K\big] \geq 1 - \alpha + O\Big(n^{-1/4}(\log n)^{3/4}\Big) \quad \text{as } n \to \infty. \tag{3.3}$$

PROOF. The proof is a straightforward application of Theorem 3.1, the Cauchy–Schwarz inequality and Feller's version of the Berry–Esseen theorem [Feller (1966)]. □

For a finite set of x, the Bonferroni method may give a narrower confidence band, thereby being of more practical interest.

Next we are interested in constructing confidence bands for the set $L_n = \{x'\beta(\theta), \varepsilon_n \leq \theta \leq 1 - \varepsilon_n\}$, where $\varepsilon_n = n^{1/(1+4b)}$ with b defined below in condition (X5). This notation will be used through the end of the paper. Generalizing the approach of Cšorgő and Révész (1984), we obtain

$$J_{4n} = \Big\{x'\hat{\beta}_n(\theta \pm d_n),\ \varepsilon_n \leq \theta \leq 1 - \varepsilon_n\Big\}, \tag{3.4}$$

which is a $(1-\alpha)100\%$ direct confidence band for L_n, where $d_n = v_\alpha\sqrt{x'Q^{-1}x/n}$, $v_\alpha = \inf\{t\colon P(\sup_{0<\theta<1}|B(\theta)| \leq t) \geq 1 - \alpha\}$, where $B(\theta)$ is a

Brownian bridge. The studentized confidence band is

$$J_{3n} = \left\{x'\hat{\beta}_n(\theta) \pm c_n,\ \varepsilon_n \le \theta \le 1 - \varepsilon_n\right\}, \tag{3.5}$$

where $c_n = sv_\alpha\sqrt{x'Q^{-1}x/n}$. As in Section 1, we first investigate the asymptotic properties of the processes $\{\hat{\beta}(\theta \pm d_n),\ \varepsilon_n \le \theta \le 1 - \varepsilon_n\}$ and $\{\hat{\beta}(\theta),\ \varepsilon_n \le \theta \le 1 - \varepsilon_n\}$. The following further conditions on the distribution function F and design X are required as shown in GJKP (these conditions are needed to extend the Bahadur representation into the tails as $\theta \to 0$ and 1):

(F2) $|f'(x)/f(x)| \le c$, for $|x| \ge K \ge 0$ and $c > 0$ where K is some constant;
(F3) $|F^{-1}(\theta)| \le c\theta(1-\theta)^{-\alpha}$ for $0 < \theta \le \theta_0, 1 - \theta_0 \le \theta < 1$, $c > 0$;
(F4) $1/f(F^{-1}(\theta)) \le c(\theta(1-\theta))^{-1-\alpha}$ for $0 < \theta \le \theta_0$ and $1 - \theta_0 \le \theta < 1$, $c > 0$;
(F5) $|f(x)| \to 0$ as $x \to A+$ and $x \to B-$, where $-\infty \le A = \sup\{x: F(x) = 0\}$ and $+\infty \ge B = \inf\{x: F(x) = 1\}$;
(X4) $\sum_{i=1}^{n}\|x_i\|^4 = O(n)$ as $n \to \infty$;
(X5) $\max_{1 \le i \le n}\|x_i\| = O(n^{(2(b-a)-\delta)/(1+4b)})$ for some $b > 0$ and $\delta > 0$ such that $0 < b - a < \varepsilon/2$ (hence $0 < b < 1/4 - \varepsilon/2$).

THEOREM 3.2. *Let* k_n *be the sequence of constants defined in Theorem* 2.1. *Under conditions* (F1)–(F3) *and* (X1)–(X5), *we have the following*:

(a) *The representation in Theorem* 2.1 *holds uniformly for all* $\varepsilon_n \le \theta \le 1 - \varepsilon_n$.
(b) *The representation in Theorem* 3.1 *holds uniformly for both* $\varepsilon_n \le \theta \le 1 - \varepsilon_n$ *and* $x \in D_K$.

PROOF. (a) Let $W_n(\mathbf{t}, \theta)$ be defined as in the proof of Theorem 2.1. By Lemma A1, $W_n(\hat{\beta}(k_n) - \beta(\theta), \theta) = n^{-1/2}k\mathbf{e}_1 + o(n^{-1/2})$ almost surely. Denoting $\Omega = \{\varepsilon_n \le \theta \le 1 - \varepsilon_n,\ \mathbf{t} \le Kn^{-1/2}\}$ and applying the procedures of proof of Lemma 3.1 of GJKP, we have that

$$\sup_{\Omega}\|W_n(\mathbf{t}, \theta) - W_n(0, \theta) - E(W_n(\mathbf{t}, \theta) - W_n(0, \theta))\| = o_p(n^{-1/2}) \tag{3.6}$$

and

$$\sup_{x \in D_K,\ \varepsilon_n \le \theta \le 1 - \varepsilon_n}\left\|\hat{\beta}\left(\theta + n^{-1/2}v_\alpha\sqrt{b'Q^{-1}b}\right) - \beta(\theta)\right\| = O_p(n^{-1/2}).$$

The rest of the proof follows from the proof of Theorem 3.1 of GJKP.
(b) Combining part (a) and Theorem 2.1 leads to Theorem 3.2(b). □

Theorem 3.2(a) gives uniform-type Bahadur representations of $\hat{\beta}(k_n) - \beta(\theta)$. The one for $\hat{\beta}(\theta) - \beta(\theta)$ has been obtained by GJKP and by Koenker and Portnoy (1987). Note the Bahadur representation in GJKP and Koenker and Portnoy (1987) can be viewed as special cases of Theorem 3.2.

COROLLARY 3.2. *Let* J_{3n}, J_{4n}, c_n, d_n, D *and* L_n *be defined by* (3.4), (3.5) *and Theorem* 3.1. *Let* $B_p(\theta)$ *be a p-dimensional Brownian bridge. Under conditions* (F1)–(F5) *and* (X1)–(X5), *we have the following* $\forall x = (1, \tilde{x})$, $\tilde{x} \in R^{p-1}$ *and* $n \to \infty$:

(a) $$\sup_{\varepsilon_n \le \theta \le 1-\varepsilon_n} \sqrt{n} f(F^{-1}(\theta)) \| \hat{\beta}(\theta) - \beta(\theta) \| \to_D \sup_{0<\theta<1} \|B_p(\theta)\|;$$

(b) $$x'\hat{\beta}(\theta) \pm c_n = x'\hat{\beta}(\theta \pm d_n) + o_p(n^{-1/2})$$

uniformly for $\theta \in (\varepsilon_n, 1-\varepsilon_n)$;

(c) $$P[L_n \in J_{3n}] = P[L_n \in J_{4n}] + o(1) = 1 - \alpha + o(1).$$

Here by $L_n \in J_{3n}$ *or* $L_n \in J_{4n}$ *we mean that each element of* L_n *is contained in the corresponding interval of* J_{3n} *or* J_{4n}.

PROOF. The proof of parts (a) and (b) is a straightforward application of Theorem 3.2(a). For part (c), just follow the proof of Corollary 2.1(c). □

Like Corollary 2.1 of Cšorgő and Révész (1984), which gives a confidence band for the quantile process $\{F^{-1}(\theta),\ \varepsilon_n \le \theta \le 1-\varepsilon_n\}$, Corollary 3.2(c) indicates that J_{4n} is a $(1-\alpha)100\%$ confidence band for regression quantile process L_n with coverage probability $1-\alpha$. Now the question is: Is it possible to construct a confidence band which is simultaneous for both $\theta \in (\varepsilon_n, 1-\varepsilon_n)$ and $x \in D_K$? The answer is yes. Define $M_n = \{x'\beta(\theta),\ \theta \in (\varepsilon_n, 1-\varepsilon_n),\ x \in D_K\}$ and $u_\alpha = \inf\{t\colon P(\sup_{0<\theta<1}\|B_p(\theta)\| \le t) \le 1-\alpha\}$. The studentized confidence band for M_n is

$$J_{5n} = \left\{x'\hat{\beta}(\theta) \pm g_n;\ \theta \in (\varepsilon_n, 1-\varepsilon_n),\ x \in D_K\right\}, \tag{3.7}$$

where $g_n = s u_\alpha \sqrt{x' Q^{-1} x / n}$. The direct confidence band for M_n is

$$J_{6n} = \left\{x'\hat{\beta}(\theta \pm h_n);\ \theta \in (\varepsilon_n, 1-\varepsilon_n),\ x \in D_K\right\}, \tag{3.8}$$

where $h_n = u_\alpha \sqrt{x' Q^{-1} x / n}$.

COROLLARY 3.3. *Under the conditions of Theorem* 3.2,

$$P(M_n \in J_{5n}) = P(M_n \in J_{6n}) + o(1) = 1 - \alpha + o(1) \quad \text{as } n \to \infty, \tag{3.9}$$

where the notations $M_n \in J_{5n}$ *and* $M_n \in J_{6n}$ *are defined as in Corollary* 3.2.

PROOF. Apply Theorem 3.2(b) and follow the proof of Corollary 3.2(c). □

As in the case of confidence intervals, we obtain alternative confidence bands J'_{2n}, J'_{4n} and J'_{5n} by replacing the term $x'\hat{\beta}(\theta \pm O(n^{-1/2}))$ in J_{2n}, J_{4n}, J_{5n} and J_{6n} by $\hat{\beta}_0(\theta \pm O(n^{-1/2})) + x'\hat{\beta}_1(\theta)$. The asymptotic equivalence between $J_{2n}, J_{4n}, J_{5n}, J_{6n}$ and $J'_{2n}, J'_{4n}, J'_{5n}, J'_{6n}$ can be easily justified by showing that

$$\sup_{\varepsilon_n \le \theta \le 1-\varepsilon_n,\ x \in D_K} \sqrt{n}\left(\tilde{x}'\hat{\beta}_1(\theta \pm O(n^{-1/2})) - \tilde{x}'\hat{\beta}_1(\theta)\right) = o_p(1), \tag{3.10}$$

which is an immediate consequence of Theorem 3.2.

The exact order of the errors in coverage probabilities of the above confidence bands are also of interest. The coverage error of J_{2n} is the same as I_{2n} defined in Section 2. The coverage errors of J_{3n}, J_{4n}, J_{5n} and J_{6n} are supposed to be higher than J_{1n} and J_{2n} because the latter are not related to ε_n. We shall compare J_{1n} with J_{2n} by simulations in Section 4.3.

4. Simulation results. The objectives of this section are as follows: (1) to compare the performance of the confidence intervals given by the three methods (the direct method based on regression quantiles, studentization based on Bofinger's variance estimator and studentization based on Hall and Sheather's variance estimator); (2) to compare the performance of the confidence intervals generated by three resampling methods; (3) to compare the performance of confidence bands by the direct method and the S-method; and (4) to compare the performance of the prediction confidence bands by the direct method and the conventional approach. All of the simulations are performed by using S-Plus software and Sun Sparc stations available in the Advanced Computing Lab of the Department of Statistics at the University of Illinois, Champaign.

4.1. *The direct, Bofinger and Hall–Sheather confidence intervals.* Only a small part of our extensive simulation studies will be reported here. See Zhou (1995) for further details. To simplify the simulation, we fix the design $X = (1, \tilde{X})$ as a matrix consisting of four columns of values: 1, $N(0,1)$, Uniform$(-1.2, 1.2)$, and Student-t with five degrees of freedom [denoted as $T(5)$]. The columns of $\tilde{X}$ are centered so that each has mean 0. The cdf of Y is chosen from three distributions [$T(1)$, $T(3)$ and $T(8)$], and θ is specified as 0.25, 0.5, or 0.90. So there are nine combinations of F and θ. This is similar to the simulation scheme used by Koenker (1994), who compares the confidence intervals for a slope parameter of linear regression models based on regression rank test statistics with other methods like percentile bootstrap and the resampling approach of Parzen, Wei and Ying (1992). For each of the nine cases, the sample size n is set to 50 except for the case when $F = T(1)$ and $\theta = 0.9$, where $n = 100$. Empirical lengths of confidence intervals corresponding to empirical level 0.9 are computed based on 1000 replications. The simulation results are summarized in Table 1, which shows that if the cdf of Y is $T(3)$ or $T(8)$, the direct confidence intervals are slightly shorter than the confidence intervals by the other two methods. When F is $T(1)$, the direct interval is a little longer than the other two except for $\theta = 0.9$, in which the direct method gives much shorter confidence intervals than the other two

TABLE 1
Empirical lengths of the three types of confidence interval corresponding to 90% *empirical levels; F is student-t with degrees of freedom* 1, 3 *and* 8; *an asterisk means that sample size* 100 *is used*

θ	**df's**	**Bofinger**	**Direct**	**Hall–Sheather**
0.25	1	1.915	1.901	1.903
	3	0.893	0.859	0.877
	8	0.714	0.711	0.727
0.5	1	0.860	0.861	0.853
	3	0.662	0.654	0.665
	8	0.616	0.608	0.615
0.9	1	10.51*	8.080*	9.010*
	3	2.032	2.013	2.059
	8	1.136	1.033	1.107

approaches. Comparison based on the lengths of confidence intervals alone does not show large differences between the three methods.

Now the question is: How well do the empirical levels approximate the nominal or theoretical levels under the three methods? In other words, are the confidence intervals conservative on coverage probabilities? Figure 1 is a plot of the nominal levels versus empirical levels for confidence intervals of the regression median $\bar{x}'\beta(0.5) = \beta_0(0.5)$. The diagonal lines are 45° lines that may be used to measure the correspondences between the nominal levels and empirical levels. For the Bofinger-type studentized confidence intervals, the points are well above the line especially when F is Cauchy; this shows that the Bofinger confidence intervals are too conservative in the sense that they are too wide at the fixed nominal level 0.9. Similar patterns show on the plots for the Hall–Sheather confidence intervals, although the conservativeness of the latter is not so severe as that of Bofinger confidence intervals. The direct confidence intervals are unbiased in the sense that their nominal levels and their empirical levels match well. Figure 2 shows the plots between the empirical levels and the lengths of the confidence intervals of $\beta_0(0.5)$ under the three methods. For the Bofinger approach, the shapes of the plotted points are similar to exponential curves whereas the shapes of the plotted data for the direct approach are almost linear. For the Hall–Sheather approach, the situation is between the two. What Figures 1 and 2 tell us is that the lengths of the Bofinger confidence intervals increase more rapidly than the direct confidence intervals when the empirical levels approach 1. Similar remarks hold for the comparison of the confidence intervals by the Hall–Sheather method and the direct method.

Our simulations also show that, when θ is set to 0.90, the conservativeness phenomenon described above disappears and the empirical coverage probabilities both of Bofinger and Hall–Sheather confidence intervals are substantially below the nominal level (say, 0.9), especially when F is Cauchy, where

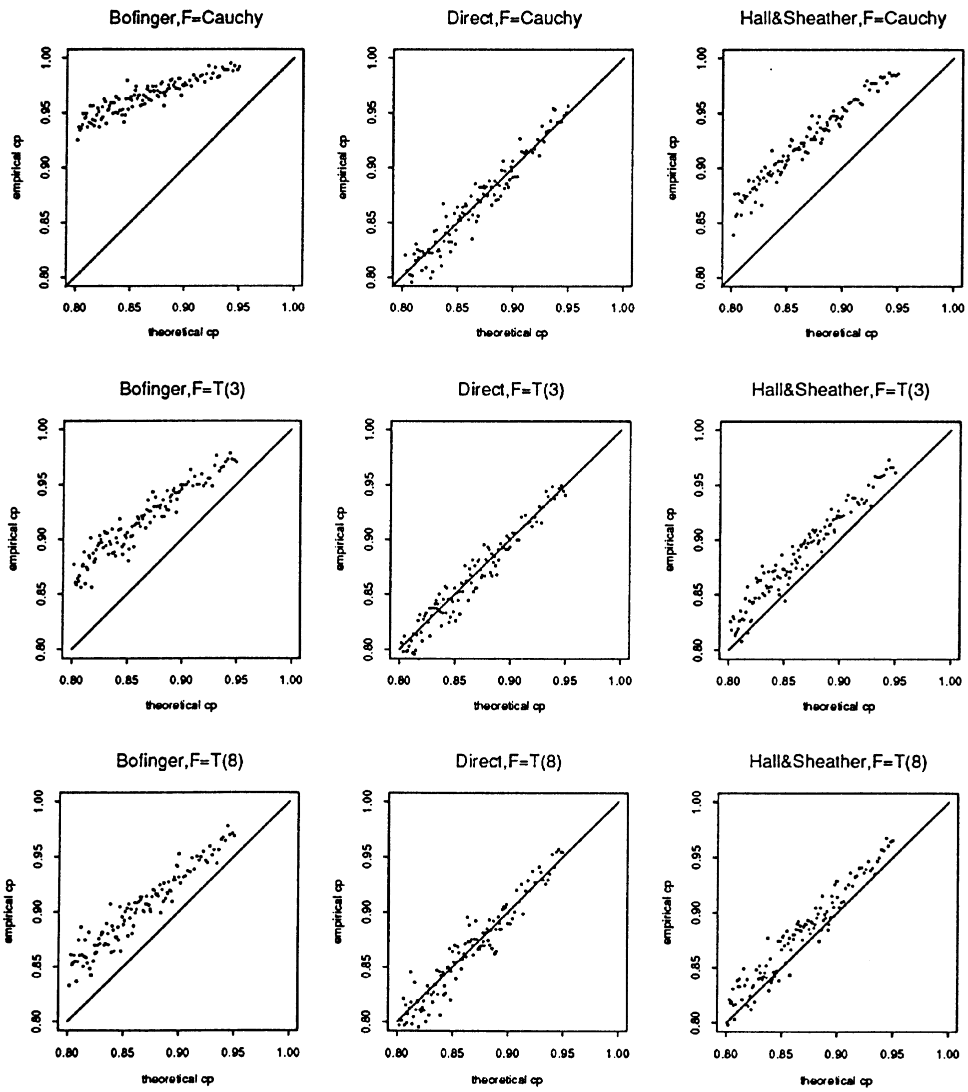

FIG. 1. *Plots of empirical levels versus nominal levels for the three types of confidence interval of* $\beta(0.5)$. *The diagonal lines are* 45°. *Three types of error distribution are considered. Each point is computed based on* 1000 *samples with size* 50.

the direct confidence intervals are a little conservative. Hence, the two studentized confidence intervals are very sensitive to F, θ and α whereas the direct confidence intervals are not and are much more robust to the departure from the normality assumption of F. Hence from the point of view of robustness, the direct method is preferred. The direct method is also more efficient computationally than the other two methods.

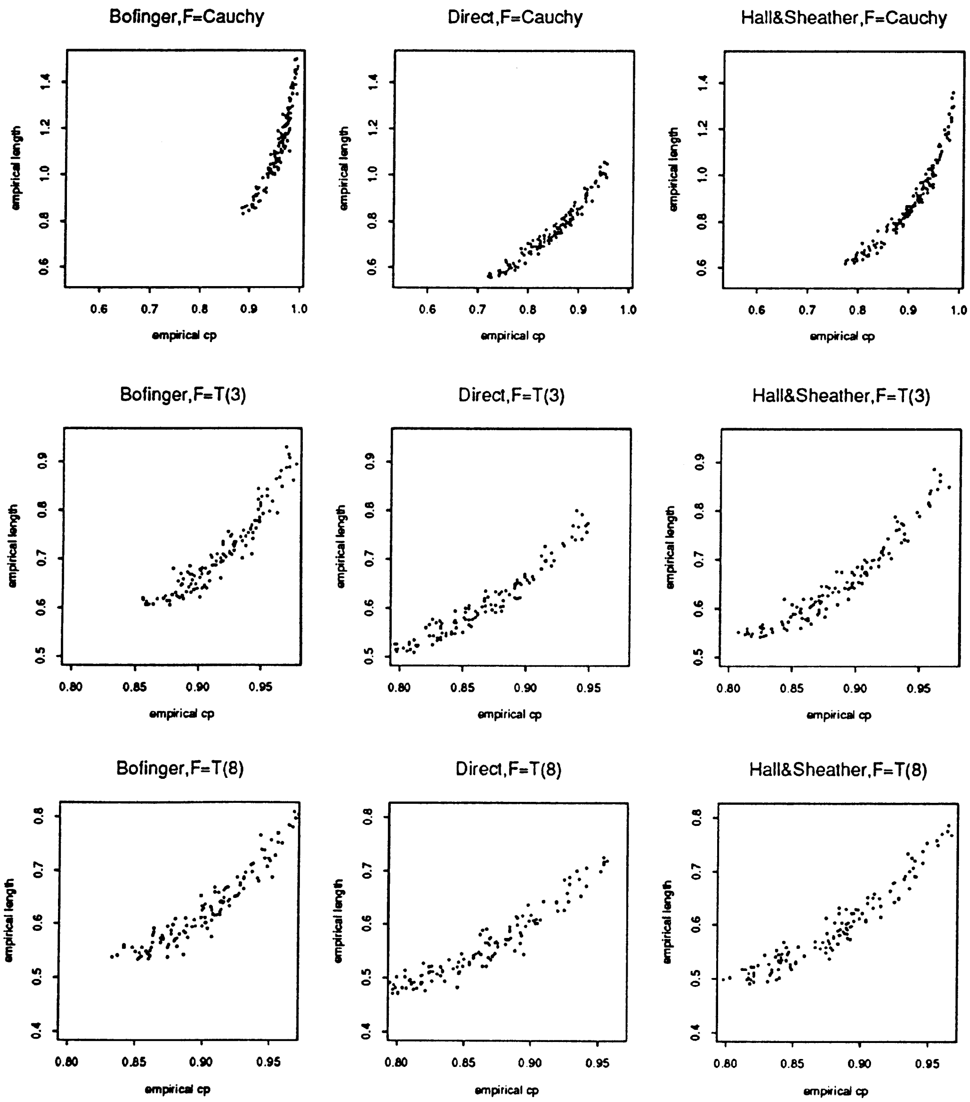

FIG. 2. *Plots of empirical levels versus empirical levels for the three types of confidence intervals of* $\beta(0.5)$. *Three error distribution functions are considered. Each point is computed based on* 1000 *samples with size* 50.

4.2. *Bootstrap confidence intervals.* We perform simulations on the confidence intervals given by the three versions of bootstrap and resampling procedures: ordinary percentile with pairwise resampling (OPPR); backward percentile with pairwise resampling (BPPR); and Parzen–Wei–Ying resampling (PWY). The details of the three resampling procedures will not be presented here. For OPPR and BPPR, we refer to Efron (1979) and Falk and

Kaufmann (1991). As for PWY, Parzen, Wei and Ying (1992) has the detailed description. The design X, F, sample size, number of replications and number of parameters p are the same as above. The resampling size is 200 for each version type of resampling approach. From Table 2, we see that BPPR confidence intervals have the same lengths of OPPR confidence intervals, but their empirical levels are higher than those of OPPR confidence intervals. This result confirms the conjectures in Section 1 that the backward bootstrap confidence interval has smaller coverage errors than the ordinary bootstrap confidence intervals. Table 2 also shows that PWY confidence intervals are relatively longer than the other two bootstrap versions. Now, comparing Tables 1 and 2 based on lengths and coverage probabilities (CP), we see that the three resampling confidence intervals are less appealing than the three confidence intervals discussed before. This conclusion is expected and is consistent with the conjecture mentioned at the end of Section 1.

4.3. *Confidence bands for regression quantile functions*. We shall compare the direct confidence bands for regression quantile functions and the confidence bands by the S-method with θ fixed at 0.5. In order to plot the results, we restrict p to 2. The sample size is set to 50 and the model can be expressed as $y_i = 1.0 + 0.5x_i + e_i$. Let x_{ij} be from $Uniform(-1.5, 1.5)$ and e_i be from either $N(0,1)$ or $T(3)$. Intuitively, the comparison between the two confidence bands is similar to the comparison discussed in the previous part. However, some criteria are needed to measure the performance of a confidence band. From Section 1, we know the confidence band we are going to construct is simultaneous for all x in an interval which is set to $(-2, 2)$ here. We first choose a sequence of x values from $(-2, 2)$. Then we construct a confidence interval for each x point and then compute the corresponding length and coverage level. The average of the lengths and coverage levels are then used to measure the performance of the confidence band. This approach can only be used as an approximation and is by no means perfect. From Table 3, we see that the direct confidence bands, like the direct confidence intervals, are robust to the departure from the normality assumption of F.

4.4. *Simulations of prediction confidence intervals*. Here, we compare the prediction interval from Section 1 to the prediction interval based on studen-

TABLE 2
Empirical coverage probabilities and empirical lengths of the bootstrap confidence intervals with nominal level 0.90 *and* $\theta = 0.5$

	***T*(1)**		***T*(3)**		***T*(8)**	
Bootstrap types	**Length**	**CP**	**Length**	**CP**	**Lenght**	**CP**
OPPR	1.098	0.955	0.730	0.959	0.658	0.938
BPPR	1.098	0.890	0.730	0.859	0.658	0.831
PWY	1.313	0.910	0.783	0.850	0.715	0.882

TABLE 3
Empirical coverage probabilities and empirical lengths of the two types of confidence bands with nominal level 0.90

		T(1)		T(3)		T(8)	
	θ	**Length**	**CP**	**Length**	**CP**	**Length**	**CP**
S-method	0.25	2.674	0.978	1.176	0.953	0.967	0.931
	0.50	1.263	0.976	0.918	0.949	0.842	0.941
	0.85	5.574	0.992	1.598	0.965	1.175	0.958
Direct method	0.25	2.294	0.851	1.110	0.914	0.914	0.898
	0.50	1.057	0.908	0.848	0.909	0.793	0.904
	0.85	6.537	0.765	1.599	0.848	1.169	0.868

tization, which has the form $(x^*\hat{\beta}_{\rm ls} \pm (1 + s_{\rm ls} n^{-1/2}) z_\alpha)$, where $\hat{\beta}_{\rm ls}$ and $s_{\rm ls}$ are the least squares estimators of the coefficients β and standard error σ, and x^* is the predictor. Our simulation shows that for sample size 50 the direct method tends to underestimate the coverage levels. However, when sample size increases to 100, this phenomenon disappeared. Table 4 shows that the studentized prediction confidence intervals are generally too long with respect to the nominal level 0.9. In contrast, the direct prediction intervals give 0.89 empirical level. In practice, it is recommended that $(x^*\hat{\beta}(\theta_1 - \delta_n), x^*\hat{\beta}(\theta_2 + \delta_n))$ be used as the $(1 - 2\alpha)100\%$ confidence bands for the θth conditional quantile if the sample size is small. Typically, θ_1 and θ_2 are set symmetrically such that $1 - \theta_2 = \theta_1 = \alpha$ and the constant δ_n can be set to $z_\alpha \sqrt{\theta(1 - \theta)}\,/n$.

5. Summary. We have shown that the direct method is a robust and computationally efficient approach to constructing confidence intervals and confidence bands for quantiles and regression quantiles. In particular, predictions for future values are of practical interest. It should be pointed out that the i.i.d. assumption on the error terms of model (1) can be weakened to independence or even to stationary dependence. These cases suggest further research on extending the direct method to heteroscedastic linear models and ARMA models. Also, because regression quantiles are equivariant under

TABLE 4
Empirical coverage probabilities and empirical lengths of the two types of prediction intervals with nominal level 0.90

Predictors	**−1.333**	**−1.000**	**−0.667**	**0.000**	**0.667**	**1.000**	**1.333**
LS CP	0.923	0.923	0.923	0.923	0.923	0.923	0.923
LS length	5.528	5.513	5.503	5.495	5.503	5.513	5.527
Direct CP	0.884	0.890	0.894	0.896	0.893	0.889	0.883
Direct length	4.778	4.773	4.767	4.756	4.745	4.739	4.733

monotone transformations, the direct method may be an effective way to construct prediction intervals for future response variables in some nonlinear models which can be linearly approximated by monotone transformations.

APPENDIX

LEMMA A1. *Let* $W_n(\mathbf{t}, \theta)$ *be defined in Theorem* 2.1. *Assume conditions* (F1) *and* (X1). *Then, with probability* 1, $\|W_n(\hat{\beta}(\theta) - \beta(\theta), \theta\| = O(n^{-3/4})$ *uniformly for* $\theta \in (0, 1)$.

PROOF. The lemma follows easily from Lemmas A.2 and A.1 of Ruppert and Carroll (1980) [actually the gradient condition in Koenker and Bassett (1978)] and then condition (X1). □

LEMMA A2. *Let* $\Psi_\theta(x)$ *be defined in Theorem* 2.1. *Assume conditions* (F1), (X1) *and* (X2). *Let* $M_n = c_0 n^{-1/2}(\log n)^{1/2}$. *For fixed* $\theta \in (0, 1)$ *and* $n \to \infty$, *we have*

$$\sup_{\|\mathbf{t}\| \le M_n} \|W_n(\mathbf{t}, \theta) - W_n(0, \theta) + f(F^{-1}(\theta))Q\mathbf{t}\| = O_p\big(n^{-3/4}(\log n)^{3/4}\big). \tag{A.1}$$

PROOF. Let $H_n(\theta, \mathbf{t}) = W_n(\mathbf{t}, \theta) - W_n(0, \theta) - E[W_n(\mathbf{t}, \theta) - W_n(0, \theta)]$. Partition space $D = \{\mathbf{t}: \|\mathbf{t}\| \le ([1/\delta_n] + 1)\delta_n M_n\}$ as the union of a class $\mathscr{E}$ of closed cubes E_i, which have vertices on the set $(k_{i1}\delta_n M_n, \dots, k_{ip}\delta_n M_n)$, where $k_{ij} = \{0, \pm 1, \dots, \pm ([1/\delta_n] + 1)\}$ and $\delta_n = c_0 n^{-1/4}(\log n)^{1/2}$. Obviously, the number of cubes in $\mathscr{E}$ is no bigger than $(2[1/\delta_n] + 3)^p$. Using the monotonicity of Ψ_θ wrt $x_i'\mathbf{t}$, we have

$$\sup_{\|\mathbf{t}\| \le M_n} \|H_n(\mathbf{t}, \theta)\| \le P_n(\mathbf{t}, \theta) + Q_n(\mathbf{t}, \theta), \tag{A.2}$$

where

$$P_n(\mathbf{t}, \theta) = \max_{E_i \in \mathscr{E}} \{\|H_n(\mathbf{t}_i, \theta)\|, \mathbf{t}_i \text{ is the lowest vertex of } E_i\}, \tag{A.3}$$

and, with the notation $r_{ij} = u_j - F^{-1}(\theta) - x_j'\mathbf{t}_i$,

$$Q_n(\mathbf{t}, \theta) = \max_{E_i \in \mathscr{E}} \Big\{ \sup_{\mathbf{t} \in E_i} \|H_n(\mathbf{t}, \theta) - H_n(\mathbf{t}_i, \theta)\|, \ \mathbf{t}_i \text{ is the lowest vertex of } E_i \Big\} \tag{A.4}$$

$$\le \max_{E_i \in \mathscr{E}} \Big\{ \frac{1}{n} \sum_{j=1}^{n} x_j \big(\Psi_\theta(r_{ij} + \delta_n M_n S_j) - \Psi_\theta(r_{ij} - \delta_n M_n S_j)\big) \tag{A.5}$$

$$+ \frac{1}{n} E\Big[\sum_{j=1}^{n} x_j \big(\Psi_\theta(r_{ij} - \delta_n M_n S_j) - \Psi_\theta(r_{ij} + \delta_n M_n S_j)\big)\Big]\Big\}, \tag{A.6}$$

where $S_j = \sum_{i=1}^{p} |x_{ij}|$. Note that $P_n(\mathbf{t}, \theta) = (P_{n1}(\mathbf{t}, \theta), \dots, P_{np}(\mathbf{t}, \theta))$ is a vector. Set $\gamma_n = c_1 n^{-3/4} (\log n)^{3/4}$ for some constant c_1. We have

$$P\big[|P_{nj}(\mathbf{t}, \theta)| \geq \gamma_n\big] \leq \sum_{E_i \in \mathscr{E}} P\big[|H_{nj}(\mathbf{t}, \theta)| \geq \gamma_n\big]. \tag{A.7}$$

By condition (X1), there exists a Λ s.t. $\max_{i,j} |x_{ij}| \leq \Lambda n^{1/4}$. Set

$$Z_i = x_i\big(\Psi_\theta\big(u_i - x_i \mathbf{t} - F^{-1}(\theta)\big) - \Psi_\theta\big(u_i - F^{-1}(\theta)\big)\big), \tag{A.8}$$

then $|nH_n(\mathbf{t}, \theta)|$ is distributed as $|\sum_{i=1}^n Z_i - \sum_{i=1}^n EZ_i|$, where the Z_i's are independent Bernoulli random variables with mean $EZ_i = F(x_i'\mathbf{t} + F^{-1}(\theta)) - \theta$. By condition (F1) and Young's version of Taylor's series theorem we can write the covariance matrix of $\sum_{i=1}^n Z_i$ as

$$X'Xf\big(F^{-1}(\theta)\big)x_i'\mathbf{t} + O\big(X'X(x_i'\mathbf{t})^2\big) \leq a_1 \sum_{i=1}^n x_i x_i' x_i' \mathbf{t} + O\left(\sum_{i=1}^n x_i x_i' (x_i'\mathbf{t})^2\right),$$

where a_1 is a positive constant larger than $f(F^{-1}(\theta))$. Here, for two matrices A and B, $A \leq B$ means $B - A$ is nonnegative definite. Hence, by conditions (X1) and (X2),

$$\operatorname{Var}\left(\sum_{i=1}^n Z_{ij}\right) \leq pa_1 \sum_{i=1}^n \|x_i\|^3 \|\mathbf{t}\| + O\left(\sum_{i=1}^n \|x_i\|^3 \|\mathbf{t}\|^2\right) = O\big(n^{1/2} (\log n)^{1/2}\big).$$

Therefore, by the Bernstein inequality [Serfling (1980), page 95], we have

$$\begin{aligned} &P\big[n|H_{nj}(\mathbf{t}, \theta)| \geq n\gamma_n\big] \\ &\quad \leq 2\exp\left(\frac{n^2 \gamma_n^2}{2\sum_{i=1}^n \operatorname{Var}(Y_{ij}) + 2(\Lambda n^{1/4}) n\gamma_n/3}\right) \\ &\quad \leq c_2 \exp(-c_3 \log n)) \leq c_2 n^{-c_3}, \end{aligned} \tag{A.9}$$

where c_2 and c_3 are some constants. We can set c_0, c_1 and a_1 s.t. $c_3 > p/4$. Then for $n \to \infty$,

$$P\big[\|P_n(\mathbf{t}, \theta)\| \geq \gamma_n\big] \leq pc_2 \left(2\left[\frac{1}{\delta_n}\right] + 3\right)^p n^{-c_3} \to 0; \tag{A.10}$$

that is,

$$P_n(\mathbf{t}, \theta) = O_p\left(\left(\frac{\log n}{n}\right)^{3/4}\right). \tag{A.11}$$

Similarly, setting $\eta_n = c_4 n^{-3/4}$ for some constant c_4, we have

$$P\big[\|Q_n(\mathbf{t}, \theta)\| \geq \eta_n\big] \to 0. \tag{A.12}$$

Hence,

$$Q_n(\mathbf{t}, \theta) = O_p(n^{-3/4}). \tag{A.13}$$

Combining $P_n(\mathbf{t}, \theta)$ and $Q_n(\mathbf{t}, \theta)$, we have

$$\sup_{\|\mathbf{t}\| \le M_n} \|W_n(\mathbf{t}, \theta) - W_n(0, \theta) - E[W_n(\mathbf{t}, \theta) - W_n(0, \theta)]\| = O\left(\left(\frac{\log n}{n}\right)^{3/4}\right). \tag{A.14}$$

Also, by conditions (X1)–(X3) and Taylor's series theorem, we have

$$\begin{aligned} E(W_n(\mathbf{t}) - W_n(0)) &= \frac{1}{n}\sum_{i=1}^{n} x_i\big(F\big(x_i'\mathbf{t} + F^{-1}(\theta)\big) - \theta\big) \\ &= \frac{1}{n}\sum x_i x_i' \mathbf{t} f\big(F^{-1}(\theta)\big) + O\left(\frac{1}{n}\sum_{i=1}^{n} x_i (x'\mathbf{t})^2\right) \\ &= Q\mathbf{t} f\big(F^{-1}(\theta)\big) + O_p\big(n^{-3/4}(\log n)^{3/4}\big). \end{aligned}$$

Thus, the assertion of Lemma A2 follows. □

Acknowledgments. The authors would like to thank Professor Xuming He and Professor Roger W. Koenker for their helpful comments and suggestions. We are also indebted to the referees for numerous helpful remarks.

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BASSETT, G. W. and KOENKER, R. W. (1982). An empirical quantile function for linear models with i.i.d. errors. *J. Amer. Statist. Assoc.* **77** 407–415.

BASSETT, G. W. and KOENKER, R. W. (1986). Strong consistency of regression quantiles and related empirical processes. *Econometric Theory* **2** 191–201.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–433.

BOFINGER, E. (1975). Estimation of a density function using order statistics. *Austral. J. Statist.* **17** 1–7.

CSÖRGŐ, M. and RÉVÉSZ, P. (1984). Two approaches to constructing simultaneous confidence bounds for quantiles. *Probab. Math. Statist.* **4** 221–236.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

FALK, M. and KAUFMANN, E. (1991). Coverage probabilities of bootstrap confidence intervals for quantiles. *Ann. Statist.* **19** 485–495.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*. Wiley, New York.

GUTENBRUNNER, C. and JUREČKOVÁ, J. (1992). Regression rank-scores and regression quantiles. *Ann. Statist.* **20** 305–330.

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

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

HALL, P. and SHEATHER, S. J. (1988). On the distribution of a studentized quantile. *J. Roy. Statist. Soc. Ser. B* **50** 381–391.

JUREČKOVÁ, J. (1977). Asymptotic relations of *M*-estimates and *R*-estimates in linear regression model. *Ann. Statist.* **5** 464–472.

KOENKER, R. W. (1994). Confidence intervals for regression quantiles. In *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics* (P. Mandl and M. Huskova, eds.). 349–359. Physica, Heidelberg.

KOENKER, R. W. and BASSETT, G. W. (1978). Regression quantiles. *Econometrica* **46** 33–50.
KOENKER, R. W. and D'OREY, V. (1987). Computing regression quantiles. *J. Roy. Statist. Soc. Ser. C* **36** 383–393.
KOENKER, R. W. and PORTNOY, S. (1987). *L*-estimation linear model. *J. Amer. Statist. Assoc.* **82** 851–857.
PARZEN, M. I., WEI, L. J. and YING, Z. (1992). A resampling method based on pivotal estimating functions. Unpublished manuscript.
PORTNOY, S. (1988). Asymptotic behavior of regression quantiles in nonstationary, dependent cases. *J. Multivariate Anal.* **38** 100–113.
PORTNOY, S. (1991). Asymptotic number of regression quantile breakpoints. *SIAM J. Sci. Statist. Comput.* **12** 867–883.
PORTNOY, S. and KOENKER, R. W. (1989). Adaptive *L*-estimation for linear models. *Ann. Statist.* **17** 362–381.
RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.
SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
ZHOU, K. Q. (1995). Quantile regression and survival analysis. Ph.D. dissertation, Univ. Illinois, Urbana–Champaign.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS, URBANA–CHAMPAIGN
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820