The Annals of Statistics
2018, Vol. 46, No. 3, 1138–1166
https://doi.org/10.1214/17-AOS1580
© Institute of Mathematical Statistics, 2018

A SMOOTH BLOCK BOOTSTRAP FOR QUANTILE REGRESSION WITH TIME SERIES

BY KARL B. GREGORY[1], SOUMENDRA N. LAHIRI[2] AND DANIEL J. NORDMAN[3]

University of South Carolina, North Carolina State University and Iowa State University

Quantile regression allows for broad (conditional) characterizations of a response distribution beyond conditional means and is of increasing interest in economic and financial applications. Because quantile regression estimators have complex limiting distributions, several bootstrap methods for the independent data setting have been proposed, many of which involve smoothing steps to improve bootstrap approximations. Currently, no similar advances in smoothed bootstraps exist for quantile regression with dependent data. To this end, we establish a smooth tapered block bootstrap procedure for approximating the distribution of quantile regression estimators for time series. This bootstrap involves two rounds of smoothing in resampling: individual observations are resampled via kernel smoothing techniques and resampled data blocks are smoothed by tapering. The smooth bootstrap results in performance improvements over previous unsmoothed versions of the block bootstrap as well as normal approximations based on Powell's kernel variance estimator, which are common in application. Our theoretical results correct errors in proofs for earlier and simpler versions of the (unsmoothed) moving blocks bootstrap for quantile regression and broaden the validity of block bootstraps for this problem under weak conditions. We illustrate the smooth bootstrap through numerical studies and examples.

1. Introduction. For a stationary time sequence $\{(Y_t, X_t) : Y_t \in \mathbb{R}, X_t \in \mathbb{R}^d\}_{t\in\mathbb{Z}}$, we consider the linear quantile regression model

$$Y_t = X_t'\beta_0(\theta) + \varepsilon_t(\theta, X_t), \tag{1.1}$$

where $\varepsilon_t \equiv \varepsilon_t(\theta, X_t)$ is an error term having θth quantile equal to zero conditional on X_t [for some $\theta \in (0, 1)$]. Thus $X_t'\beta_0(\theta)$ is the θth conditional quantile of Y_t

Received June 2016; revised February 2017.

[1]Supported in part by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany.

[2]Supported in part by NSF Grants DMS-1007703 and DMS-1310068.

[3]Supported in part by NSF Grant DMS-1406747.

MSC2010 subject classifications. Primary 62G09; secondary 62G20, 62J05, 62M10.

Key words and phrases. Kernel smoothing, jackknife after bootstrap, moving blocks, tapering, value at risk.

given X_t. For a set of observations $(Y_1, X_1), \dots, (Y_n, X_n)$, we consider the problem of estimating the sampling distribution of

$$\sqrt{n}(\hat{\beta}_n - \beta_0),$$

under time dependence, where $\hat{\beta}_n$ is the quantile regression (QR) estimator of $\beta_0 \equiv \beta_0(\theta)$ defined as

$$\hat{\beta}_n \equiv \arg\min_{\beta \in \mathbb{R}^d} S_n(\beta), \qquad S_n(\beta) \equiv n^{-1} \sum_{t=1}^{n} (Y_t - X_t'\beta) \operatorname{sign}_\theta(Y_t - X_t'\beta), \tag{1.2}$$

for $\theta \in (0, 1)$ and

$$\operatorname{sign}_\theta(x) = \theta I(x > 0) - (1 - \theta) I(x \le 0), \qquad x \in \mathbb{R}.$$

A challenge is that, under conditions of independence or weak dependence [8, 37], the distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is known to converge to a normal limit with an asymptotic covariance matrix depending intricately on a smooth population quantity, namely the (conditional) density of the error terms in (1.1). This asymptotic behavior resembles that of sample quantiles for independent data $Z_1, \dots, Z_n$ having distribution function F with density f, where the limiting variance of the θth sample quantile is $\theta(1 - \theta)/f^2(F^{-1}(\theta))$ (cf. [31]). The appearance of the error density in the asymptotic variance complicates normal approximations to the distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$. Particularly for independent data, this has occasioned ongoing research since the birth of QR ([21]) about covariance estimation as well as resampling methods for the QR estimator $\hat{\beta}_n$.

For covariance estimation, Koenker [19] proposed an estimator of the reciprocal of the error density based on inverting a rank test from Gutenbrunner et al. [12]. Powell [30] suggested plugging in a kernel estimate of the error density, which has become a standard approach. Hasan and Koenker [14] and He [15] proposed other plug-in methods. These estimators were originally developed under the assumption of independent observations, though Kato [18] established the consistency of Powell's kernel estimator for weakly dependent observations. For approximating the sampling distribution of QR estimators with independent data, several bootstrap procedures have also been proposed. Parzen, Wei and Ying [28] suggested using an asymptotically pivotal derivative of the QR criterion function to generate resamples of $\hat{\beta}_n - \beta_0$. Hahn [13] showed the validity of a residual bootstrap for the QR estimator by treating this as a type of nonsmooth M-estimator considered by Arcones and Giné [1]. Feng, He and Hu [7] developed a wild bootstrap to account for heteroscedasticity in QR. As the limit law in quantile estimation involves a smooth error density, bootstrap versions involving various types of smoothing have also been considered for independent data. For L_1 regression and sample quantiles, De Angelis, Hall and Young [3, 4] showed advantages to bootstraps based on kernel-smoothing. Buchinsky [2] applied "m-out-of-n" bootstrapping for QR, using small m for seeming computational ease, but smoothing of the resamples

was not considered. Horowitz [16] proposed bootstrapping a smoothed objective function in QR.

All of the above resampling works in QR are intended for *independent* data, and the time series scenario has received comparatively little consideration, particularly with regard to bootstrap smoothing. For weakly dependent data, Fitzenberger [8] studied the moving block bootstrap (MBB) of Künsch [22] and Liu and Singh [26] for estimating the distribution of the QR estimator. Simulations there indicate that the MBB does not improve greatly upon normal approximations, which is an issue related to the appearance of smooth densities in the limit of $\sqrt{n}(\hat{\beta}_n - \beta_0)$. As a remedy, we investigate a smooth extended tapered block bootstrap (SETBB) proposed by Gregory, Lahiri and Nordman [10]. For QR in particular, the SETBB approach offers improvements by incorporating *two* rounds of smoothing in resampling not present in the MBB: namely, smoothing of both data blocks and individual observations. That is, this bootstrap involves block tapering as a trademark of state-of-the-art block bootstraps for capturing time dependence, as originally proposed in the tapered block bootstrap of Paparoditis and Politis [27] and expanded by Shao [32] to the extended tapered block bootstrap (ETBB). The SETBB also includes additional kernel smoothing in resampling, analogously to techniques known to enhance the bootstrap for independent data [3, 4]. In this case, smoothing is performed on each observation to recreate a resampling distribution with smoothness properties that better mimic those of the population marginal distribution.

To help frame findings here, we note that Gregory, Lahiri and Nordman [10] originally considered the SETBB for time series estimators defined by differentiable functionals that exclude QR. Arguably, QR is a more meaningful context for the smooth bootstrap, where smoothing can provide greater methodological benefit due to the nature of QR (i.e., the distribution of the QR estimator depending on unknown smooth quantities). However, the QR problem requires a different bootstrap investigation. In establishing the SETBB for QR, technical complexities arise due to differing layers of resampling (both block- and kernel-based) as well as nonsmoothness issues with the QR criterion. To provide some contrast, Fitzenberger [8] considered a simpler block bootstrap (i.e., MBB) for QR, but the proof given is considerably incorrect. The argument in [8] fails to properly recreate objective functions in the bootstrap world and then misses critical steps in verifying bootstrap approximations. [For example, that proof (p. 274) defines an important bootstrap expected value in terms of unconditional expectation "E" instead of bootstrap expectation "E_*." This confuses (conditional) bootstrap probability with (unconditional) data-generating probability, and the resulting argument misses rounds of bootstrap approximation important for the QR problem. The motivation was probably that expectations under "E" can be appropriately smooth with QR but ordinary bootstrap expectations "E_*" are generally not, which is where analysis of bootstrap approximations is needed; we describe these issues in Section 3.3.] As resampling layers are even more intricate for the SETBB, developments here are

not eased by past bootstrap work; the proof in [8] is in error and the theory of [10] does not apply.

Our findings are broadly formulated to include both smooth and unsmoothed versions of the block bootstrap for QR with time series, which validates the SETBB and modified versions of the MBB and ETBB as special cases. Additionally, the SETBB work of Gregory, Lahiri and Nordman [10] depended on Gaussian kernels for smoothing and particular bandwith/block restrictions. The current results allow for both general kernels and bandwidth choices, covering a wider range of bootstrap implementations.

Section 2 describes the SETBB method and its formulation for QR. Section 3 provides the main distributional results as well as regularity conditions, and Section 4 discusses practical issues in implementation and computation (e.g., selection of tuning parameters). Numerical studies in Section 5 demonstrate performance improvements offered by the SETBB, and Section 6 illustrates the method for QR with two data examples, one involving value at risk estimation. Section 7 offers concluding remarks. Proofs of main results appear in Section 8 as well as in the Supplementary Material [11].

2. Smooth extended tapered block bootstrap method.

2.1. *Bootstrap empirical distributions*. Note that the objective function (1.2), which the QR estimator $\hat{\beta}_n$ minimizes, may be reexpressed as

$$S_n(\beta) \equiv \int \operatorname{sign}_\theta(y - \beta' x)(y - \beta' x)\, dF_n(y, x) \tag{2.1}$$

using the empirical distribution of the data

$$F_n = \frac{1}{n} \sum_{t=1}^{n} \delta_{(Y_t, X_t)}.$$

Rather than recreating bootstrap replicates of the original time series as in the MBB, the SETBB aims to reproduce bootstrap renditions F_n^* of the empirical distribution F_n which can be subsequently applied to formulate bootstrap versions of the QR estimator (cf. Section 2.2). For this, the SETBB combines smoothing of data blocks and individual observations as follows.

For some choice of block size ℓ, set $b = \lfloor n/\ell \rfloor$ and let $I_1^*, \dots, I_b^*$ denote a uniform random sample from the integers $\{1, \dots, n - \ell + 1\}$. The variables $\{I_j^*\}_{j=1}^b$ denote the starting points of blocks of length ℓ, for example, $(Y_{I_j^*}, X_{I_j^*}), \dots, (Y_{I_j^*+\ell-1}, X_{I_j^*+\ell-1})$, from the original series $(Y_1, X_1), \dots, (Y_n, X_n)$. Following [27] and [32] for block tapering, define weights

$$w_\ell(t) := w\left(\frac{t - 0.5}{\ell}\right), \qquad t \in \mathbb{R}, \ell = 1, 2, \dots, \tag{2.2}$$

based on a taper function $w : \mathbb{R} \to [0, 1]$, where $w(x)$ is symmetric about $x = 1/2$, positive in a neighborhood of $x = 1/2$, nondecreasing for $x \in [0, 1/2]$, and such that $w(x) = 0$ for $x \notin [0, 1]$. Then the empirical weight π_t^* assigned to time point t by the SETBB empirical distribution function F_n^* is equal to

$$\pi_t^* = \frac{1}{b\|w_\ell\|_1} \sum_{j=1}^{b} \sum_{k=1}^{\ell} w_\ell(k) \mathbb{I}\big(I_j^* = t - k + 1\big),$$

where $\|w_\ell\|_1 = \sum_{k=1}^{\ell} |w_\ell(k)|$ and $\mathbb{I}(\cdot)$ is the indicator function. Hence, π_t^* represents an average weighting over random blocks $j = 1, \dots, b$, where the jth weight contribution from $\{w_\ell(k)/\|w_\ell\|_1\}_{k=1}^{\ell}$ depends on the position of time point t among the indices $(I_j^*, \dots, I_j^* + \ell - 1)$ of the jth random block of length ℓ. Independently from the block selection, the data points $(Y_1, X_1), \dots, (Y_n, X_n) \in \mathbb{R}^{d+1}$ are also randomly perturbed by small amounts to induce additional smoothing. In particular, let $H \equiv H_n$ denote a $(d+1) \times (d+1)$ bandwidth matrix and let $\{Z_t^*\}_{t=1}^n$ represent i.i.d. random vectors in $\mathbb{R}^{d+1}$ drawn from some common distribution G_Z with mean zero to define perturbed data as $(Y_t^*, X_t^{*\prime})' = (Y_t, X_t')' + HZ_t^*$, $t = 1, \dots, n$. If G_Z had a kernel density g and $H > 0$, then each resample (Y_t^*, X_t^*) would be equivalent to a random selection from a multivariate kernel $g_{t,n}(y, x) = [\det(H)]^{-1} g[H^{-1}(y - Y_t, x' - X_t')']$, $y \in \mathbb{R}$, $x \in \mathbb{R}^d$. Finally, the SETBB version of the empirical distribution is defined as

$$F_n^* = \sum_{t=1}^{n} \pi_t^* \delta_{(Y_t^*, X_t^*)}, \tag{2.3}$$

associating point masses $\{\pi_t^*\}_{t=1}^n$ to smoothed variables $\{(Y_t^*, X_t^*)\}_{t=1}^n$.

The role of the block tapering is analogous to tapering lag windows in spectral density estimation. In particular, the MBB uses untapered data blocks, corresponding to a uniform weight function $w(t) = \mathbb{I}(t \in [0, 1])$. However, block bootstraps using smooth tapers $w(\cdot)$ in (2.2), which are zero $w(t) = 0$ at $t = 0$ and 1 and thereby give less weights to time points near the ends of a data block, are known to possess smaller-order bias properties in variance estimation (cf. [22, 27]); technically, this involves a taper with a self-convolution $w * w(t) = \int_{-1}^{1} w(x + |t|) w(x)\, dx$ being twice differentiable at $t = 0$. The data perturbation step added to the SETBB is intended to impart further smoothness features to the resampling distribution in order to mimic similar features at the population distribution level. Hence, viewed in our framework, the original MBB approach of Fitzenberger [8] for QR would correspond to two choices in the SETBB scheme: $H_n = 0$ and $w(t) = \mathbb{I}(t \in [0, 1])$, involving no kernel smoothing for individual observations and no taper smoothing for data blocks. A smooth block taper w combined with no observational smoothing ($H_n = 0$) would reproduce the extended tapered block bootstrap (ETBB) approach of Shao [32], which has not been considered for QR. Hence, the SETBB method here is general in its formulation

of smoothing at observation/block levels and contains the MBB/ETBB as special cases.

2.2. *Bootstrap for time series quantile regression.* The QR estimator $\hat{\beta}_n$ again minimizes a criterion $S_n(\beta) \equiv \int (y - \beta' x)\mathrm{sign}_\theta(y - \beta' x)\, dF_n(y, x)$ from (1.2) or (2.1) based on the empirical distribution of the original data $F_n = n^{-1}\sum_{t=1}^n \delta_{(Y_t, X_t)}$. The true parameter β_0 under the model (1.1) (i.e., defining some θth conditional quantile) minimizes a population counterpart $\mathrm{E}S_n(\beta) = \int (y - \beta' x)\mathrm{sign}_\theta(y - \beta' x)\, dF(y, x)$ based on the true marginal distribution F of $\{(Y_t, X_t)\}_{t=1}^n$. To estimate the sampling distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ for purposes of testing or confidence region calibration, we require SETBB analogs $\hat{\beta}_n^*$ and $\tilde{\beta}_n$ of $\hat{\beta}_n$ and β_0, respectively.

We define the bootstrap version $\hat{\beta}_n^*$ of $\hat{\beta}_n$ as

$$\hat{\beta}_n^* \equiv \arg\min_{\beta \in \mathbb{R}^d} S_n^*(\beta),$$

where the bootstrap criterion function

$$\begin{aligned} S_n^*(\beta) &\equiv \sum_{t=1}^n \pi_t^* (Y_t^* - X_t^{*\prime}\beta)\mathrm{sign}_\theta(Y_t^* - X_t^{*\prime}\beta) \\ &= \int (y - \beta' x)\mathrm{sign}_\theta(y - \beta' x)\, dF_n^*(y, x) \end{aligned}$$

mimics $S_n(\beta)$ using the SETBB distribution $F_n^* = \sum_{t=1}^n \pi_t^* \delta_{(Y_t^*, X_t^*)}$ in place of the original empirical distribution F_n. The bootstrap analog $\tilde{\beta}_n$ of β_0 is defined as $\tilde{\beta}_n \equiv \arg\min_{\beta \in \mathbb{R}^d} \mathrm{E}_* S_n^*(\beta)$, based on the bootstrap expected value

$$\mathrm{E}_* S_n^*(\beta) = \frac{1}{n - \ell + 1} \sum_{j=1}^{n-\ell+1} \sum_{k=1}^{\ell} \frac{\omega_\ell(k)}{\|\omega_\ell\|_1} h_{j+k-1,n}(\beta),$$

as the counterpart of $\mathrm{E}S_n(\beta)$ defining $\beta_0 \equiv \arg\min_{\beta \in \mathbb{R}^d} \mathrm{E}S_n(\beta)$, where above

$$h_{t,n}(\beta) = \int (Y_t - X_t'\beta + J_\beta H z)\mathrm{sign}_\theta(Y_t - X_t'\beta + J_\beta H z)\, dG_Z(z)$$

for the smoothing bandwidth matrix $H \equiv H_n$ and $J_\beta = (1, -\beta')$. The bootstrap expectation E_* and the consequent form of $\mathrm{E}_* S_n^*(\beta)$ are defined by the distribution of random block positions and (independently) the random sample $Z_1^*, \dots, Z_n^* \in \mathbb{R}^{d+1}$ of perturbations from G_Z. When blocks are untapered and observations are unsmoothed so that $H = 0$, it holds that $h_{t,n}(\beta) = (Y_t - X_t'\beta)\mathrm{sign}_\theta(Y_t - X_t'\beta)$, and the bootstrap expectation $\mathrm{E}_* S_n^*(\beta)$ will closely resemble the original sample criterion $S_n(\beta)$ from (1.2).

Finally, the SETBB version of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is given by

$$m_\ell^{1/2}\sqrt{n}(\hat{\beta}_n^* - \tilde{\beta}_n),$$

where $m_\ell = \|w_\ell\|_1^2/[\ell\|w_\ell\|_2^2]$ is a scalar adjustment for block tapering, using $\|w_\ell\|_1 = \sum_{k=1}^{\ell} |w_\ell(k)|$ and $\|w_\ell\|_2^2 = \sum_{k=1}^{\ell} |w_\ell(k)|^2$. This factor also appears in the TBB/ETBB methods of [27] and [32], and is not an order adjustment as $m_\ell \to [\int_0^1 w(x)\,dx]^2/\int_0^1 w^2(x)\,dx \in (0,\infty)$ holds when $\ell \to \infty$ as $n \to \infty$.

3. Main results.

3.1. *Assumptions and conditions.* We next frame some general conditions for establishing the validity of the SETBB method for QR. Under the model (1.1), weak dependence of the regressor/error process $\{(X_t, \varepsilon_t)\}$ is prescribed in terms of strong mixing coefficients defined as $\alpha(k) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty\}$, where P denotes process probability and $\mathcal{F}_{-\infty}^0, \mathcal{F}_k^\infty$ represent the σ-algebras generated by $\{(X_t, \varepsilon_t) : t \le 0\}$ and $\{(X_t, \varepsilon_t) : t \ge k\}$, respectively; see [5]. Let $\|\cdot\|$ denote the Euclidean norm applied to either vectors or matrices. We use the following conditions:

(C.1) $\{(X_t, \varepsilon_t)\}_{t\in\mathbb{Z}}$ is a $\mathbb{R}^d \times \mathbb{R}$-valued stationary process where $Y_t = X_t'\beta + \varepsilon_t$ and $\mathrm{E}X_1 \mathrm{sign}_\theta(\varepsilon_1) = 0$.

(C.2) The distribution of ε_t given X_t has a density $f(\varepsilon) \equiv f(\varepsilon|X_t)$ (with respect to the Lebesgue measure) such that:

- (i) $f(\cdot|X_t)$ is continuous at 0 (a.s. P).
- (ii) For some $a_0 > 0$ and some nonnegative X_t-measurable function C_{X_t}, it holds that

$$\sup_{|a|\le a_0} f(\varepsilon + a|X_t) \le C_{X_t}\big(1 + |\varepsilon|\big) \qquad \text{for each } \varepsilon \in \mathbb{R}$$

  (a.s. P) such that $\mathrm{E}[C_{X_t} + \|X_t\|^{4+\kappa_1} C_{X_t}] < \infty$ for some $\kappa_1 > 0$.
- (iii) $\inf_{|\varepsilon|\le a_0} f(\varepsilon|X_t) > a_1 > 0$ (a.s. P) for some $a_0, a_1 > 0$.

(C.3) For some $\kappa_2 > 0$, $\mathrm{E}\|X_t\|^{3+\kappa_2} < \infty$ and $\sum_{k=1}^\infty [\alpha(k)]^{\tilde\delta/(2+\tilde\delta)} < \infty$ hold for $\tilde\delta = \min\{\kappa_1, \kappa_2\}$ for κ_1 in (C.2).

(C.4) $\Sigma \equiv \sum_{k=-\infty}^\infty \mathrm{Cov}(X_0 \mathrm{sign}_\theta(\varepsilon_0), X_t \mathrm{sign}_\theta(\varepsilon_t))$ and $L \equiv \mathrm{E}X_t X_t' f(0|X_t)$ are positive definite.

(C.5) Vectors $\{(Y_t, X_t')' \in \mathbb{R}^{d+1}\}_{t=1}^n$ are in general position (a.s. P) for each n, where we say generic vectors $\{a_t \in \mathbb{R}^{d+1}\}_{t=1}^n$ are in general position if no hyperplane in $\mathbb{R}^{d+1}$ contains $d+2$ points among $a_1, \dots, a_n$.

(C.6) In the resampling, $Z_1^*, \dots, Z_n^*$ are i.i.d. random vectors in $\mathbb{R}^{d+1}$ with common distribution G_Z and, for $Z \sim G_Z$, $\mathrm{E}Z = 0$ and $\mathrm{E}\|Z\|^{4+\tilde\delta} < \infty$ hold for $\tilde\delta$ in (C.4). Also, for each n, resampling points $\{(Y_t^*, X_t^{*\prime})' \in \mathbb{R}^{d+1}\}_{t=1}^n$ are in general position with probability 1 (a.s. P).

(C.7) The resampling block ℓ satisfies $\ell^{-1} + \ell/n \to 0$ as $n \to \infty$ and the norm $h_n \equiv \|H_n\|$ of bandwidth matrix in smoothing satisfies $h_n \to 0$.

The conditions are similar to, but generally weaker than, those of Fitzenberger [8]. In particular, the moment/mixing assumptions are milder and Condition (C.2) also allows a broader variety of error distributions because the conditional densities f are not assumed to be bounded, supported on a compact interval, or to satisfy Lipschitz conditions. Under Condition (C.2), the density $f(\cdot|X_t)$ near the origin may be potentially unbounded in X_t, noting that L in (C.4) involves $f(0|X_t)$ and influences the distribution of the QR estimator (cf. Theorem 1). So, for example, a simple conditional distribution $\varepsilon_t|X_t \sim N(0, \sigma^2 X_t^2)$ with normally distributed errors having a standard deviation proportional to the regressor X_t is allowable, but is not admitted by assumptions in [8]. Unlike [8], we do not assume a compact parameter space for QR inference. Additionally, the block length condition in condition (C.7) is minimal (e.g., in contrast, $\ell^2/n = o(1)$ is required in [8]). Likewise, in the resample kernel smoothing, the bandwidth matrix is only required to converge to zero as $n \to \infty$. The choice $h_n = 0$ is allowed, which means that our assumptions admit the unsmoothed ETBB or MBB estimator (cf. Section 2.1). We make no strict assumption about the distribution of the resample perturbations in (C.6), other than that these have mean zero and a sufficient number of finite moments; this aspect also provides an extension over earlier SETBB results in [10], in which normal perturbations were assumed/required. We also assume that bootstrap samples $\{(Y_t^*, X_t^{*\prime})'\}_{t=1}^n \subset \mathbb{R}^{d+1}$ are in general position if the original data $\{(Y_t, X_t')'\}_{t=1}^n \subset \mathbb{R}^{d+1}$ are [e.g., a continuous G_Z in (C.6) is sufficient for this]. At the expense of additional complexity, the assumption regarding the stationarity of $\{(X_t, \varepsilon_t)\}_{t\in\mathbb{Z}}$ could be weakened to ones of first- and/or second-order stationarity of certain process quantities [e.g., $\{X_t X_t'\}$, $\{X_t \operatorname{sign}_\theta(\varepsilon_t)\}$, $\{\varepsilon_t \operatorname{sign}_\theta(\varepsilon_t)\}$].

3.2. *Bootstrap distributional results*. Under mild assumptions, Theorem 1 establishes asymptotic normality of the QR estimator $\hat{\beta}_n \equiv \arg\min_{\beta\in\mathbb{R}^d} S_n(\beta)$, where $S_n(\beta)$ denotes the QR objective function from (1.2). A consistency statement is also included for completeness, which technically follows under weaker conditions [e.g., (C.1), (C.2)(iii) and (C.3) with $\kappa_2 = \tilde{\delta} = 2$].

THEOREM 1. *Assume Conditions* (C.1)–(C.5). *Then, as* $n \to \infty$:

(i) $\hat{\beta}_n \xrightarrow{p} \beta_0$, *where* β_0 *is the unique minimizer of* $\mathrm{E}S_n(\beta)$, $\beta \in \mathbb{R}^d$;
(ii) *for* L, Σ *defined in* (C.4),

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N\big(0, L^{-1}\Sigma L^{-1}\big).$$

Hence, as mentioned in the Introduction (Section 1), a complication in approximating the distribution of the QR estimator $\hat{\beta}_n$ directly is that its limiting normal distribution involves an unknown, smooth population density. In particular, from Theorem 1(ii), note that the term L depends on the conditional density $f(0)$ of an

error variable evaluated at zero; see also Condition (C.2). This aspect again motivates the SETBB method for approximating the sampling distribution of the QR estimator $\hat{\beta}_n$.

The next result shows that the SETBB version $m_\ell^{1/2}\sqrt{n}(\hat{\beta}_n^* - \tilde{\beta}_n)$, involving a centering analog $\tilde{\beta}_n$ of the true parameter β_0, successfully estimates the QR distribution $\sqrt{n}(\hat{\beta}_n - \beta_0)$.

THEOREM 2. *Assume Conditions* (C.1)–(C.7). *Then, as* $n \to \infty$:

(i) *for the SETBB version* $m_\ell^{1/2}\sqrt{n}(\hat{\beta}_n^* - \tilde{\beta}_n)$ *of* $\sqrt{n}(\hat{\beta}_n - \beta_0)$,

$$\sup_{x\in\mathbb{R}^d} \left| P_*\big(m_\ell^{1/2}\sqrt{n}(\hat{\beta}_n^* - \tilde{\beta}_n) \le x\big) - P\big(\sqrt{n}(\hat{\beta}_n - \beta_0) \le x\big)\right| \xrightarrow{p} 0,$$

where P_* *denotes bootstrap probability*;

(ii) *for the SETBB version* $\tilde{\beta}_n \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}_* S_n^*(\beta)$ *of* $\beta_0 \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}S_n(\beta)$,

$$\tilde{\beta}_n = \beta_0 + O_p\big(n^{-1/2} + h_n\big),$$

where $h_n \equiv \|H_n\|$ *is the norm of the kernel bandwidth matrix*;

(iii) *for* $\beta_n \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}[\mathrm{E}_* S_n^*(\beta)]$,

$$\sqrt{n}(\tilde{\beta}_n - \beta_n) \xrightarrow{d} N\big(0, L^{-1}\Sigma L^{-1}\big).$$

From Theorem 2(i), the general SETBB method is consistent for the distribution of the QR estimator under minimal assumptions on the resampling block length ℓ, smoothing bandwidth parameter H and the block taper function w from (2.2). Again, the choice $H = 0$ is allowed in combination with either a smooth block taper w or untapered blocks $w(t) = \mathbb{I}(t \in [0, 1])$, so that Theorem 2 also includes the ETBB/MBB methods.

In Theorem 2(i), note that the centering $\tilde{\beta}_n \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}_* S_n^*(\beta)$ for the SETBB rendition $\hat{\beta}_n^* \equiv \arg\min_{\beta\in\mathbb{R}^d} S_n^*(\beta)$ of the QR estimator is not the original QR estimator $\hat{\beta}_n \equiv \arg\min_{\beta\in\mathbb{R}^d} S_n(\beta)$. Recall $\tilde{\beta}_n$ imitates the true parameter $\beta_0 \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}S_n(\beta)$ in the bootstrap world where, under smooth resampling, $\mathrm{E}_* S_n^*(\beta)$ aims to mimic smooth features of $\mathrm{E}S_n(\beta)$ that the empirical criterion $S_n(\beta)$ generally does not [due to additional kernel-smoothing with bandwidth H_n in $\mathrm{E}_* S_n^*(\beta)$]. Consequently, the minimizers $\tilde{\beta}_n$ and $\hat{\beta}_n$ of $\mathrm{E}_* S_n^*(\beta)$ and $S_n(\beta)$, respectively, have distinct behaviors as well as different focal points of convergence given by $\beta_n \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}[\mathrm{E}_* S_n^*(\beta)]$ and $\beta_0 \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}S_n(\beta)$. Theorem 2(ii)–(iii) outlines some relevant properties of the SETBB centering $\tilde{\beta}_n$ for reference. Like the QR estimator $\hat{\beta}_n$, the SETBB centering $\tilde{\beta}_n$ is consistent for β_0 but not at a guaranteed $\sqrt{n}$-rate; instead, the norm $\|H_n\|$ of the kernel smoothing bandwidth [which is general under Condition (C.7)] additionally impacts $\tilde{\beta}_n$.

This illustrates that, particularly under smoothing, bootstrap centering can differ dramatically from the original QR estimator $\hat{\beta}_n$. With unsmoothed resampling $H_n = 0$, our proofs show that $\sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) \xrightarrow{p} 0$ holds so that, in this case, the original QR estimator $\hat{\beta}_n$ could be validly used to define bootstrap centering, as done by [8] for the MBB. However, under kernel smoothing steps ($H_n \neq 0$) for the SETBB, the original QR estimator $\hat{\beta}_n$ is generally inappropriate for bootstrap centering and, if so used, general consistency of the SETBB method does not follow (e.g., consider, under Theorems 1–2, that $\hat{\beta}_n$ and $\tilde{\beta}_n$ will vary from the true parameter β_0 by different orders if $\sqrt{n}\|H_n\| \to 0$ fails to hold).

3.3. *On proving the main results.* In determining the limit distribution of $\hat{\beta}_n$, the near derivative $D_n(\beta) = n^{-1}\sum_{t=1}^{n} X_t \operatorname{sign}_\theta(Y_t - X_t'\beta)$ of the QR objective function $S_n(\beta)$ is nonsmooth in β. [Technically, $S_n(\beta)$ from (1.2) is also nonsmooth but has left/right partial derivatives which asymptotically match $D_n(\beta)$.] However, the expectation function $\lambda_n(\beta) \equiv \mathrm{E}D_n(\beta) = \mathrm{E}X_1 \operatorname{sign}_\theta(Y_1 - X_1'\beta)$ is differentiable, at least at the true parameter β_0. Hence, an argument for establishing the limit distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ involves an indirect expansion of $D_n(\hat{\beta}_n)$ combined with a distributional limit for $\sqrt{n}D_n(\beta_0)$, rather than direct expansion of $D_n(\hat{\beta}_n)$ around β_0 as commonly encountered with smooth-parameter estimating functions. Namely, under the mixing/moment conditions, the consistency of the QR estimator $\hat{\beta}_n$ for β_0 may be shown along with a central limit theorem for the sum $D_n(\beta_0)$, that is, $\sqrt{n}D_n(\beta_0) \xrightarrow{d} N(0, \Sigma)$ as $n \to \infty$ where $\mathrm{E}D_n(\beta_0) = \lambda_n(\beta_0) = 0$ holds. One can additionally establish a strong approximation result,

$$\sqrt{n}\big[\{D_n(\hat{\beta}_n) - \lambda_n(\hat{\beta}_n)\} - \{D_n(\beta_0) - \lambda_n(\beta_0)\}\big] \xrightarrow{p} 0,$$

and that the criterion $S_n(\beta)$ at its minimizer $\hat{\beta}_n$ has a near derivative $D_n(\hat{\beta}_n)$ converging to zero, that is, $\sqrt{n}D_n(\hat{\beta}_n) \xrightarrow{p} 0$. These aspects combine to yield

$$\sqrt{n}\big[\lambda_n(\hat{\beta}_n) - \lambda_n(\beta_0)\big] = -\sqrt{n}D_n(\beta_0) + o_p(1) \xrightarrow{d} N(0, \Sigma).$$

Consequently, because $\lambda_n(\cdot)$ is differentiable at β_0, with $\partial\lambda_n(\beta_0)/\partial\beta = -L$ for all n using the positive definite matrix L from Condition (C.4), the asymptotic normality of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ in Theorem 1 can be established by Taylor expansion and using $\hat{\beta}_n \xrightarrow{p} \beta_0$. This general strategy is well known [8]. For example, [17] established asymptotic normality of maximum likelihood estimators based on i.i.d. data, without assuming differentiability of log-likelihood functions. For independent data, [29] established asymptotic normality of M-estimators based on nonsmooth criterion functions having differentiable expectations; this includes the QR estimator. For weakly dependent data, [37] considered median regression estimators, following arguments of Huber to establish asymptotic normality.

However, to establish the validity of the bootstrap, the counterpart arguments for the QR estimator are more difficult to recreate in the bootstrap realm. Similarly to the above, the proof of Theorem 2 is based on the SETBB version of $D_n(\beta)$ given by

$$D_n^*(\beta) = \sum_{t=1}^{n} \pi_t^* X_t^* \operatorname{sign}_\theta(Y_t^* - X_t^{*\prime}\beta). \tag{3.1}$$

Compared to its counterpart $\lambda_n(\beta) = \mathrm{E}D_n(\beta)$, the analog expectation in the bootstrap world, say $\tilde{D}_n(\beta) \equiv \mathrm{E}_* D_n^*(\beta)$ [i.e., the bootstrap rendition of $\lambda_n(\beta)$], may not be sufficiently smooth at the bootstrap version $\tilde{\beta}_n \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}_* S_n^*(\beta)$ of the true parameter $\beta_0 \equiv \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}S_n(\beta)$. However, the expectation $\tilde{\lambda}_n(\beta) \equiv \mathrm{E}\tilde{D}_n(\beta)$ of this bootstrap expected value is differentiable at $\tilde{\beta}_n$. Consequently, one may establish two rounds of approximation for the SETBB estimator $\hat{\beta}_n^*$ and its centering $\tilde{\beta}_n$ as

$$\sqrt{n}[\{D_n^*(\hat{\beta}_n^*) - \tilde{D}_n(\hat{\beta}_n^*)\} - \{D_n^*(\tilde{\beta}_n) - \tilde{D}_n(\tilde{\beta}_n)\}] \xrightarrow{p*} 0,$$
$$\sqrt{n}[\{\tilde{D}_n(\hat{\beta}_n^*) - \tilde{\lambda}_n(\hat{\beta}_n^*)\} - \{\tilde{D}_n(\tilde{\beta}_n) - \tilde{\lambda}_n(\tilde{\beta}_n)\}] \xrightarrow{p*} 0,$$

in order to pair a smooth quantity $\sqrt{n}\{\tilde{\lambda}_n(\hat{\beta}_n^*) - \tilde{\lambda}_n(\tilde{\beta}_n)\}$ to the difference $\sqrt{n}\{D_n^*(\hat{\beta}_n^*) - D_n^*(\tilde{\beta}_n)\}$; above $\xrightarrow{p*}$ loosely denotes convergence in bootstrap probability conditional on the data. One may also show that $m_\ell^{1/2}\sqrt{n}D_n^*(\tilde{\beta}_n)$ at $\tilde{\beta}_n$ mimics the normal limit of $\sqrt{n}D_n(\beta_0)$ at β_0 and that, analogously to $\sqrt{n}D_n(\hat{\beta}_n)$, the bootstrap counterpart $\sqrt{n}D_n^*(\hat{\beta}_n^*)$ converges to zero. By a Taylor expansion of $m_\ell^{1/2}\sqrt{n}[\tilde{\lambda}_n(\hat{\beta}_n^*) - \tilde{\lambda}_n(\tilde{\beta}_n)]$ around the bootstrap centering $\tilde{\beta}_n$, one can then establish that the bootstrap limit of $m_\ell^{1/2}\sqrt{n}(\hat{\beta}_n^* - \tilde{\beta}_n)$ matches the normal limit of $\sqrt{n}(\hat{\beta}_n - \beta_0)$. Further details appear in Section 8. In contrast, the argument for the (unsmooth) MBB in the seminal QR paper of [8] [p. 274 (IV)] incorrectly defines the bootstrap expected version of $\tilde{D}_n^*(\beta)$ as $\mathrm{E}D_n^*(\beta)$ (an unconditional expectation), and not as $\tilde{D}_n(\beta) \equiv \mathrm{E}_* D_n^*(\beta)$ (a bootstrap expectation). As a consequence, that argument misses the two rounds of bootstrap approximation explained above for rigorously justifying the MBB, and mistakenly confounds bootstrap probability with that of the original data-generating process.

4. Practical implementation.

4.1. *Bandwidth and block size selection.* As a tuning parameter, the SETBB method requires a bandwidth matrix H_n associated with a distribution G_Z for defining resampling perturbations (Section 2.1). We prescribe a data-driven approach for selecting H_n, and in the numerical studies to follow we choose the $(d+1)$-dimensional standard Gaussian distribution for the kernel G_Z. As resampling here is connected to a kernel density estimate, a data-based choice of the

bandwidth may be suggested via the following heuristic. Note the (conditional) density appearing in the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ owes to the error terms in the model (1.1). Therefore, we choose a bandwidth appropriate for kernel estimation of this density based on a kernel function g. Here, g denotes a selected probability density (e.g., standard normal) for defining the perturbation distribution G_Z in resampling. Namely, we compute and use the fitted residuals $\hat{e}_t = Y_t - X_t'\hat{\beta}_n$ from the model and, with the kernel g, apply a standard data-driven choice for bandwidth—namely the [33] approach, which selects a bandwidth $\hat{h}_n \asymp n^{1/5}$. We then set $H_n = \hat{h}_n I_{d+1}$, where I_{d+1} is the $(d+1) \times (d+1)$ identity matrix, and this remains fixed during all bootstrap resampling. This approach performed well in our simulation studies (Section 5).

In the SETBB method, we empirically choose a block length ℓ using the nonparametric plug-in (NPPI) approach developed in [25], which aims to minimize the MSE of MBB or tapered block bootstrap (TBB) estimators. The block size, once chosen, remains fixed for all bootstrap resamples. In particular, for estimating the variance φ_n of a real-valued smooth functional of a sample mean, the MSE of the TBB estimator $\hat{\varphi}_n(\ell)$, with block length ℓ, admits an expansion

$$\mathrm{MSE}\{\hat{\varphi}_n(\ell)\} = v\ell n^{-1} + B^2\ell^{-4} + o\big(\ell^{-4} + \ell n^{-1}\big)$$

in terms of bias $\mathrm{Bias}\{\hat{\varphi}_n(\ell)\} = B\ell^{-2} + o(\ell^{-2})$ and variance $\mathrm{Var}\{\hat{\varphi}_n(\ell)\} = v\ell n^{-1} + o(\ell n^{-1})$ depending on unknown process quantities B, v [27]. Minimizing this MSE provides the form of an optimal block size

$$\ell_n^{\mathrm{opt}} = \big(4B^2/v\big)^{1/5} n^{1/5}.$$

The NPPI then produces an estimate $\hat{\ell}_n^{\mathrm{opt}}$ of ℓ_n^{opt} by plugging estimates $\hat{B}$ and $\hat{v}$ of B and v into the above expression.

For a pilot block size ℓ_1, an estimate of B can be constructed as

$$\hat{B} = \ell_1^2(4/3)\big\{\hat{\varphi}_n(\ell_1) - \hat{\varphi}_n(2\ell_1)\big\},$$

for which $\mathrm{E}\hat{B} = B + o(1)$ holds. An estimate of v can be obtained using the jackknife-after-bootstrap (JAB) described in [23]. This is a block jackknife procedure which sequentially removes m consecutive blocks of length ℓ_1 and computes the bootstrap estimator on the remaining data. When computing $\hat{\varphi}_n(\ell_1)$ in $\hat{B}$ above from Monte Carlo approximation (i.e., resampling $\lfloor n/\ell_1 \rfloor$ blocks of length ℓ_1 to build block bootstrap versions of a sample-mean-based statistic), the removal of m consecutive blocks of length ℓ_1 can be performed from the initial block resamples by simply considering subsets of the $\lfloor n/\ell_1 \rfloor$ bootstrap block selections for which none of the first m, second m, and so on, blocks were chosen. The NPPI requires choices of ℓ_1 and m, for which we use $\ell_1 = \mathrm{nint}(n^{1/5})$, where $\mathrm{nint}(x)$ is the nearest integer to x, and $m = \lfloor n^{1/3}\ell^{2/3} \rfloor$, following [24]; see [25] for details. Finally, because the QR estimator $\hat{\beta}_n$ is not a smooth function of a sample mean itself, we carry out our block selection by considering

instead a vector $\sqrt{n}D_n(\beta_0)$ of sample means having the asymptotic covariance matrix Σ from (C.4). We then calculate the NPPI block size for the estimator $\hat{\varphi}_n(\ell) = \mathrm{tr}\{\mathrm{Cov}_*(\sqrt{n}\tilde{D}_n^*(\tilde{\beta}_n))\}$ of $\mathrm{tr}(\mathrm{Cov}(\sqrt{n}D_n(\beta_0))$, based on $\sqrt{n}\tilde{D}_n^*(\tilde{\beta}_n)$ in (3.1). The above block length selections are intended for the SETBB with a smooth block taper (2.2); if blocks are untapered (e.g., MBB), then we instead estimate the block length $\hat{\ell}_n^{\mathrm{opt}} = [2\hat{B}^2/\hat{v}]^{1/3}n^{1/3}$ with $\hat{v}$ as above and a modification $\hat{B} = 2\ell_1\{\hat{\varphi}_n(\ell_1) - \hat{\varphi}_n(2\ell_1)\}$; see also [24].

Lastly, for the SETBB, we base our block tapering on the trapezoidal function $w(t) = (t/c)\mathbb{I}(t \in [0, c]) + \mathbb{I}(t \in [c, 1-c]) + \{(1-t)/c\}\mathbb{I}(t \in [1-c, 1])$ with $c = 0.43$ as in [27] and [32].

4.2. *Computation*. The SETBB method for QR may be implemented readily with standard software packages. For example, after generating resampling weights $\{\pi_t^*\}_{t=1}^n$ and perturbed data $\{(Y_t^*, X_t^*)\}_{t=1}^n$, the corresponding SETBB version $\hat{\beta}_n^*$ of the QR estimator is easily obtained from the R package `quantreg` [20] by

```
rq(perturbed.y ~ perturbed.x, tau = theta,
                weights = pi.star)
```

for a quantile θ of interest. The SETBB also requires one determination of the bootstrap centering $\tilde{\beta}_n$ of $\hat{\beta}_n^*$, found by minimizing $\mathrm{E}_* S_n^*(\beta)$. The latter function often admits a smooth closed form. For instance, if $H_n = h_n I_{d+1}$ with $h_n > 0$ with G_Z as the Normal$(0, I_{d+1})$ distribution, then $\mathrm{E}_* S_n^*(\beta)$ is expressed in terms of the standard normal distribution Φ and density ϕ as

$$\sum_{t=1}^{n} \frac{\tilde{w}_t}{n-\ell+1}\left[(Y_t - X_t'\beta)\left\{\theta - \Phi\left(\frac{-Y_t + X_t'\beta}{\sqrt{\delta(\beta)}}\right)\right\} + \delta(\beta)\phi\left(\frac{Y_t - X_t'\beta}{\sqrt{\delta(\beta)}}\right)\right],$$

for $\delta(\beta) = h_n^2(1 + \|\beta\|^2)$ and defining $\tilde{w}_t$ as $\sum_{k=1}^{t}\omega_\ell(k)/\|\omega_\ell\|_1$, 1 or $\sum_{k=1}^{n-t+1}\omega_\ell(k)/\|\omega_\ell\|_1$ in the cases $t < \ell$, $\ell \le t \le n-\ell+1$, or $t > n-\ell+1$, respectively. The R function `nlm`, for example, can then be used to find $\tilde{\beta}_n$ using the original QR estimator $\hat{\beta}_n$ as a starting value. We use the above implementation in our simulation studies and data examples in the following sections.

5. Simulation studies. We here examine the performance of the SETBB approach for (i) estimating the covariance matrix of $\sqrt{n}\hat{\beta}_n$, (ii) estimating the cumulative distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$, and (iii) constructing confidence intervals for individual quantile regression coefficients.

We simulated 500 data sets of size $n = 100$ according to $Y_t = X_t'\beta + \varepsilon_t$, with $X_t = (1, \tilde{X}_t')'$, $\tilde{X}_t \in \mathbb{R}^4$ such that

$$\tilde{X}_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \Psi U_t, \qquad U_t \sim \mathrm{Normal}(0, I_4),$$
$$\varepsilon_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + V_t, \qquad V_t \sim F_V,$$

$t = 1, \dots, 100$, using independent sets $\{V_t\}$ and $\{U_t\}$ of i.i.d. random variables. For innovations V_t of the error series, we chose F_V to be (i) a Normal$(0, 1)$ distribution, (ii) a centered/scaled Chi-square distribution with unit variance or (iii) a t_3 distribution scaled to have unit variance. Note that the error distributions satisfy (C.2) and the autogressive model has an exponential mixing rate. Under these choices, $X_t'\beta$ is the conditional median of Y_t given X_t. We set $\beta = (0, 1, -1, 1, -2)'$ and considered values (ϕ_1, ϕ_2) equal to $(0.7, 0.1)$ or $(0.8, 0.1)$. The matrices (Φ_1, Φ_2, Ψ) were either $(0.7I_4, 0.1I_4, I_4)$ or $(0.8I_4, 0.1I_4, I_4)$, under which $\{\tilde{X}_t\}_{t=1}^n$ are four independent time series, or the matrices (Φ_1, Φ_2, Ψ) were chosen to respectively have (l', l) entry equal to

$$0.4^{1+|l'-l|}, \qquad 0.1^{1+|l'-l|}(-1)^{|l'-l|}, \quad \text{and} \quad 0.5^{|l'-l|}, \qquad 1 \le l', l \le 4, \tag{5.1}$$

under which $\{\tilde{X}_t\}_{t=1}^n$ are realizations of a vector autoregressive time series.

To produce the SETBB distribution estimator, 2500 bootstrap resamples were used for each data set, with choices $\ell_1 = 3$ and $m = 10$ in the NPPI block size selection from Section 4.1. The smoothing perturbations for the SETBB were generated from a Normal$(0, I_{d+1})$ distribution, and the bandwidth matrix H was set to $H = \hat{h} I_{d+1}$ based on the Sheather–Jones selection $\hat{h}$ for kernel density estimation from the fitted residuals $\{\hat{\varepsilon}_t = Y_t - X_t'\hat{\beta}_n\}_{t=1}^n$.

For comparison, we also included block bootstraps with various combinations of smoothing in resampling based on the use of a smooth block taper $w(t)$ (or not) and/or the use of a data smoothing perturbation step (or not). That is, in addition to the SETBB (i.e., smoothed data blocks and data perturbation), we considered the moving block bootstrap (MBB) [i.e., unsmoothed data blocks $w(t) = I_{t\in[0,1]}$ and no data smoothing perturbation], a smooth moving block bootstrap (SMBB) [i.e., unsmoothed data blocks $w(t) = I_{t\in[0,1]}$ but with data perturbation as in the SETBB] and the extended tapered block bootstrap (ETBB) [i.e., smoothed data blocks $w(t)$ and no data perturbation]. For comparison to the bootstrap methods, Powell's kernel estimator $\hat{\Omega}_n^{(PK)}$ of $L^{-1}\Sigma L^{-1} = \lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\hat{\beta}_n)$ was also computed and used for a normal approximation to the distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$. The estimator $\hat{\Omega}_n^{(PK)}$ was obtained using the function `summary.rq()` from the `R` package `quantreg` on the fitted model with the option `se = "ker"` and otherwise default settings [20].

5.1. *SETBB estimation of the covariance matrix of* $\sqrt{n}\hat{\beta}_n$. To compare performance in covariance matrix (5×5) estimation, we evaluated the average estimation errors $\|\mathrm{Cov}_*(\sqrt{n}\hat{\beta}_n^*) - \mathrm{Cov}(\sqrt{n}\hat{\beta}_n)\|_F^2/25$ for the four bootstrap methods over all simulated datasets, where $\|\cdot\|_F$ is the Frobenius norm, as well as the average error $\|\hat{\Omega}_n^{(PK)} - \mathrm{Cov}(\sqrt{n}\hat{\beta}_n)\|_F^2/25$ from Powell's estimator. We refer to these as mean squared errors (MSEs), where the true covariance matrix $\mathrm{Cov}(\sqrt{n}\hat{\beta}_n)$ of the QR estimator was approximated from 10,000 simulations. Table 1 shows the ratios of bootstrap MSEs to the MSE of Powell's kernel estimator

TABLE 1
Ratios of bootstrap MSEs at data-adaptive bandwidth and block size choices to MSE of Powell's kernel estimator of Cov($\sqrt{n}\hat{\beta}_n$) over 500 simulated data sets

	$(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ $(\phi_1, \phi_2) = (0.7, 0.1)$			$(\Phi_1, \Phi_2, \Psi) = (0.8I_4, 0.1I_4, I_4)$ $(\phi_1, \phi_2) = (0.8, 0.1)$		
	Normal	**Cent/scaled χ_1^2**	**Scaled t_3**	**Normal**	**Cent/scaled χ_1^2**	**Scaled t_3**
MBB	1.12	1.21	1.10	0.82	0.82	0.85
ETBB	0.95	0.93	0.97	0.83	0.82	0.88
SMBB	0.66	0.70	0.77	0.73	0.75	0.74
SETBB	0.64	0.71	0.74	0.76	0.77	0.76
P's Ker	1.00	1.00	1.00	1.00	1.00	1.00

under the models with $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$, $(\phi_1, \phi_2) = (0.7, 0.1)$ and $(\Phi_1, \Phi_2, \Psi) = (0.8I_4, 0.1I_4, I_4)$, $(\phi_1, \phi_2) = (0.8, 0.1)$. Under both settings, block bootstraps with a smoothing pertubation (SETBB, SMBB) perform much better than Powell's estimator and their unsmoothed bootstrap counterparts (ETBB, MBB). For $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ and $(\phi_1, \phi_2) = (0.7, 0.1)$, the MBB does not perform as well as Powell's kernel, and the ETBB performs only slightly better, but all bootstrap methods perform much better than Powell's kernel when the dependence is strengthened to $(\Phi_1, \Phi_2, \Psi) = (0.8I_4, 0.1I_4, I_4)$ and $(\phi_1, \phi_2) = (0.8, 0.1)$. Hence, in QR, block bootstraps without data smoothing may not easily perform better than direct kernel estimation techniques due to the smooth population quantity (conditional error density) in the limiting variance of the QR estimator; this supports findings in [8]. However, data smoothing steps introduced in the SETBB method do consistently improve performance in variance estimation. We remark that this performance is achieved under data-driven block size and bandwidth selections.

Figure 1 depicts the behavior of the four bootstrap estimators of $\mathrm{Cov}(\sqrt{n}\hat{\beta}_n)$ across the block sizes $\ell = 1, \dots, 30$ under the model with Chi-square innovations and with $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ and $(\phi_1, \phi_2) = (0.7, 0.1)$. The vertical bars in each panel depict the relative frequency of the NPPI block size selections. The horizontal dotted line in each panel is positioned at the MSE over the 500 simulated data sets under the NPPI-selected block sizes, and the horizontal dashed line gives the MSE of Powell's kernel estimator over the 500 simulated data sets. We see that the data smoothing perturbation reduces the MSE of block bootstraps across a wide range of block sizes, with additional improvements created by smooth block tapering. The SETBB method exhibited the best results, and the associated NPPI method also made good empirical block size selections.

Figure 2 depicts MSE results for estimating $\mathrm{Cov}(\sqrt{n}\hat{\beta}_n)$ in the same format as in Figure 1 under the model in which the covariate realizations $\{\tilde{X}_t\}$ come from the vector autoregressive time series in (5.1) with Normal(0, 1) innovations for generating the error series. We see that the SETBB under data-based block selection

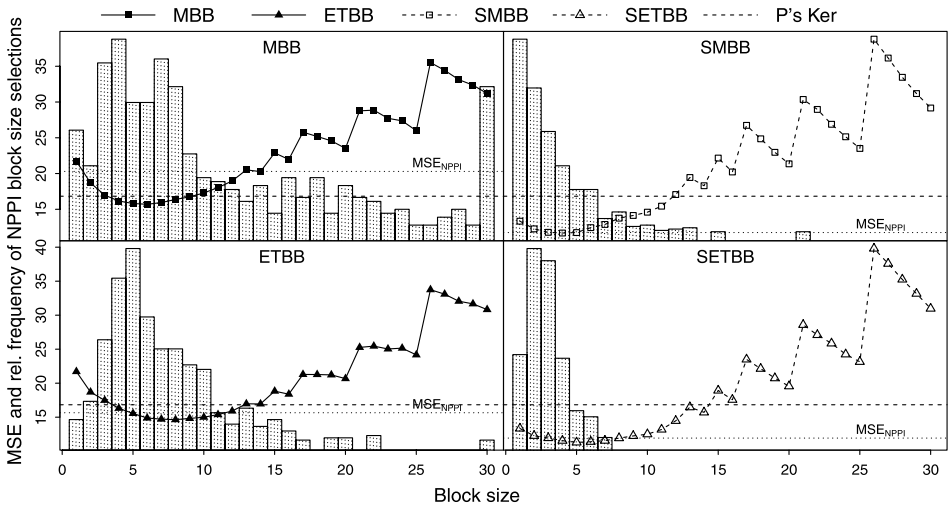

FIG. 1. *MSE across block sizes and relative frequency of NPPI block size selections for the MBB, SMBB, ETBB and SETBB under the centered and scaled Chi-square errors in the* $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ *and* $(\phi_1, \phi_2) = (0.7, 0.1)$ *model over* 500 *simulated data sets. Dashed line shows MSE of Powell's kernel and dotted line shows MSE of bootstrap method under NPPI block size selections.*

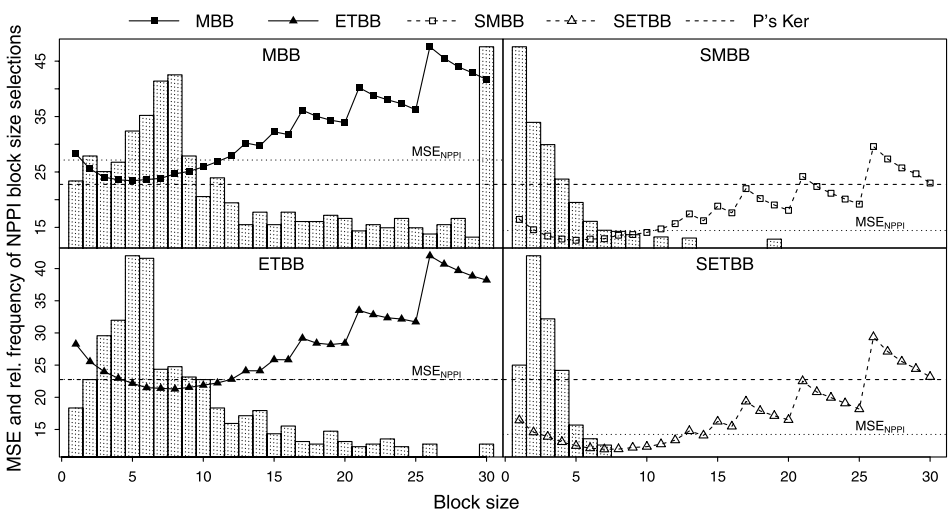

FIG. 2. *MSE across block sizes and relative frequency of NPPI block size selections for the MBB, SMBB, ETBB and SETBB under Normal*$(0, 1)$ *errors in the vector autoregressive covariates model in* (5.1) *over* 500 *simulated data sets. Dashed line shows MSE of Powell's kernel and dotted line shows MSE of bootstrap method under NPPI block size selections.*

estimates $\mathrm{Cov}(\sqrt{n}\hat{\beta}_n)$ with a much lower MSE than its unsmooth counterpart and Powell's kernel estimator. Remarkably, the MBB with no smoothing/tapering performs more poorly than Powell's kernel over all block sizes. These results demonstrate that the SETBB can accommodate a degree of multicollinearity among the covariates.

5.2. *SETBB estimation of the distribution function of* $\sqrt{n}(\hat{\beta}_n - \beta_0)$. To assess the performance of the SETBB for distribution estimation, we considered the QR estimator $\hat{\beta}_{n1}$ of the first slope parameter β_{01}, and for several $x \in \mathbb{R}$ values we computed the MSE of the bootstrap estimator $P_*(m_\ell^{1/2}\sqrt{n}(\hat{\beta}_{n1}^* - \tilde{\beta}_{n1}) \leq x)$ of $P(\sqrt{n}(\hat{\beta}_{n1} - \beta_{01}) \leq x)$ from the simulation runs; the true probabilities $P(\sqrt{n}(\hat{\beta}_{n1} - \beta_{01}) \leq x)$ were approximated from 10,000 simulations. The results under the scaled t_3-distributed errors are displayed in Figure 3, in which the upper three panels show the average cumulative distribution function estimators over 500 simulation runs across a range of x values in the left tail, middle region and right tail of the distribution. The lower panels show the MSE. For comparison, Powell's kernel variance estimator was used for a normal approximation to the distribution of $\sqrt{n}(\hat{\beta}_{n1} - \beta_{01})$. The filled and unfilled squares and triangles are arbitrarily positioned and serve only to label the curves.

We see from the lower panels that the MBB and ETBB are not competitive with the normal approximation based on Powell's estimator in the middle of the distribution; in the tails, however, the normal approximation based on Powell's estimator

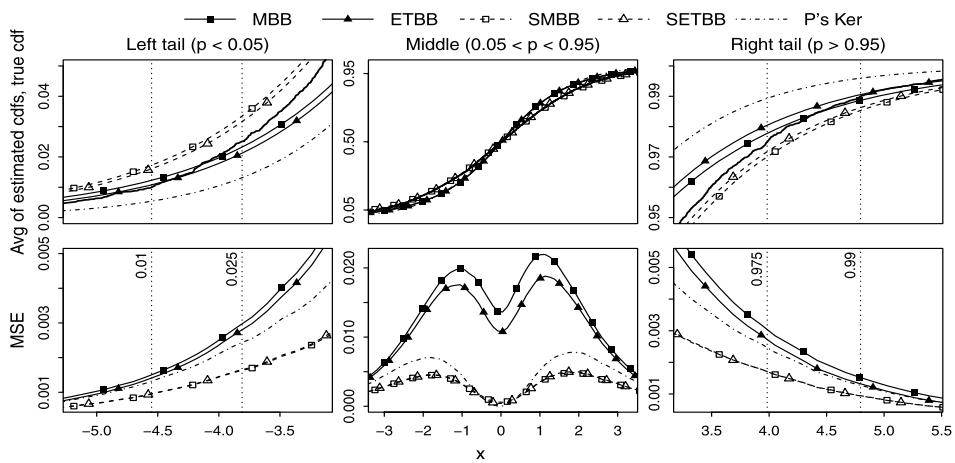

FIG. 3. *Upper panels*: *Average of bootstrap estimators* $P_*(m_\ell^{1/2}\sqrt{n}(\hat{\beta}_{n1}^* - \tilde{\beta}_{n1}) \leq x)$ *of* $P(\sqrt{n}(\hat{\beta}_{n1} - \beta_{01}) \leq x)$ *and of a normal approximation based on Powell's kernel estimator across a range of* x *values under the scaled* t_3 *errors in the* $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$, $(\phi_1, \phi_2) = (0.7, 0.1)$ *model*; *solid curve denotes true simulated distribution*. *Lower panels*: *MSE achieved by bootstrap estimators and the normal approximation from Powell's kernel*.

appears to be too lightly-tailed, whereas the MBB and ETBB distributions appear to be less biased in the tails and have a similar MSE. Both the SMBB and SETBB achieve lower MSEs in estimating the sampling distribution across nearly the entire range of x, importantly achieving much lower MSEs in the tails of the distribution. Additionally, when considering normal approximations to $P(\sqrt{n}(\hat{\beta}_{n1} - \beta_{01}) \le x)$ based on bootstrap estimators of $\mathrm{Cov}(\sqrt{n}\hat{\beta}_n)$, all four bootstrap methods also turn out to perform better than Powell's kernel, with the SETBB performing better by far; see the Supplementary Material [11] for further summaries.

5.3. *Estimation performance under heteroscedastic error variances.* We also consider the performance of the SETBB for quantile regression when the error variance is heteroscedastic. We thus define an error term

$$\varepsilon_t^{\mathrm{het}} = \varepsilon_t \sqrt{0.5 + 0.5\tilde{X}_{t1}^2}, \tag{5.2}$$

such that the variance of the error depends on the value of one of the covariates, and we set $Y_t = X_t'\beta + \varepsilon_t^{\mathrm{het}}$, where $\{X_t\}_{t=1}^n$ and $\{\varepsilon_t\}_{t=1}^n$ are generated with $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ and $(\phi_1, \phi_2) = (0.7, 0.1)$ and Normal(0, 1) innovations. We see from Figure 4, which shows MSE results for estimating $\mathrm{Cov}(\sqrt{n}\hat{\beta}_n)$ in the same format as in Figures 1 and 2, that the SETBB still performs well under this heteroscedastic model. Additional simulation results under this type of heteroscedasticity are provided in the Supplementary Material [11].

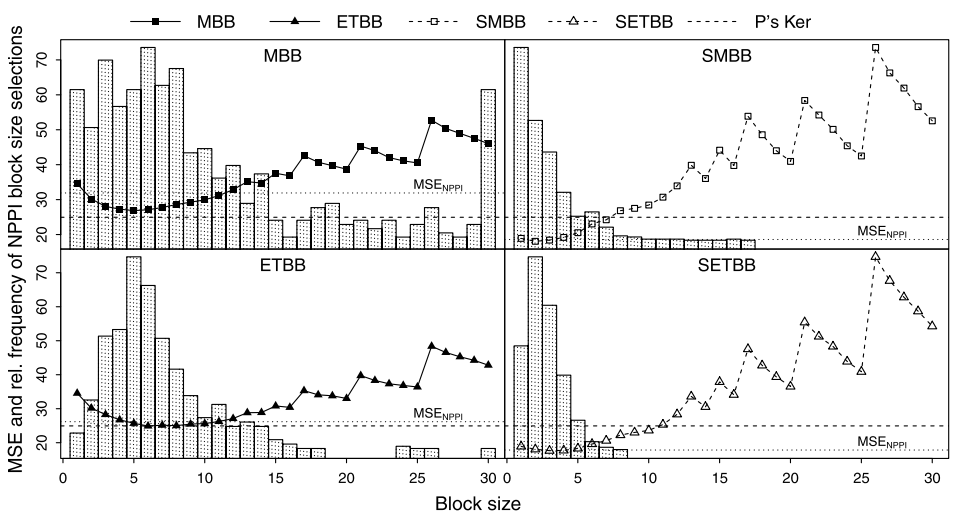

FIG. 4. *MSE across block sizes and relative frequency of NPPI block size selections for the MBB, SMBB, ETBB and SETBB under the heteroscedastic errors from* (5.2) *constructed with Normal*(0, 1) *innovations in the* $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ *and* $(\phi_1, \phi_2) = (0.7, 0.1)$ *model over* 500 *simulated data sets. Dashed line shows MSE of Powell's kernel and dotted line shows MSE of bootstrap method under NPPI block size selections.*

5.4. *Coverage of confidence intervals based on the SETBB.* As another assessment of distribution estimation performce, we also study the coverage and average width of confidence intervals based on quantiles of the MBB, ETBB, SMBB and SETBB estimators of the distribution of $\sqrt{n}(\hat{\beta}_{nj} - \beta_{0j})$ for each of the quantile regression coefficients $\beta_{01} = 1$, $\beta_{02} = -1$, $\beta_{03} = 1$, and $\beta_{04} = 2$ under each of the model settings. The block size and bandwidth are chosen according to the data-based procedures discussed in Section 4.1. For comparison, we include confidence intervals from normal-approximations to the distribution of $\sqrt{n}(\hat{\beta}_{nj} - \beta_{0j})$ using Powell's kernel estimator of the variance. We also consider confidence intervals constructed via the self-normalization procedure introduced by Zhou and Shao (ZS) in [38] for making inference on parameters in a linear model with dependent errors. For the latter procedure, a tuning parameter r must be selected by the user; we choose $r = 0.1, 0.2$ in accordance with the discussion in Section 3.1 of [38].

Table 2 displays the coverage results in the setting with Chi-square innovations and with $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ and $(\phi_1, \phi_2) = (0.7, 0.1)$. The coverage of the unsmooth bootstrap methods is far below the nominal coverage of 95%, but both the SMBB and SETBB achieve the nominal coverage for each of the regression coefficients. We emphasize that this coverage is obtained under data-based bandwidth and block-size choices. The coverage of the confidence intervals based on the normal approximation with Powell's kernel estimator of the variance plugged in lies rather below the nominal level. The coverage of the ZS confidence intervals is quite close to nominal, although tending, perhaps, toward over-coverage.

We point out that the average widths of the ZS confidence intervals are much wider than those of the SMBB and SETBB confidence intervals, which suggests

TABLE 2

Coverage and (average width) of bootstrap-quantile 95% confidence intervals under data-based choices of block and bandwidth as well as normal-approximation confidence intervals with Powell's kernel estimator and Zhou and Shao [38] confidence intervals with $r = 0.1, 0.2$. Computed over 500 simulated data sets from the model with Chi-square innovations and $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ and $(\phi_1, \phi_2) = (0.7, 0.1)$

β_{0j}	MBB	ETBB	SMBB	SETBB	P's Ker	ZS ($r = 0.1$)	ZS ($r = 0.2$)
1	0.76	0.78	0.95	0.94	0.87	0.97	0.96
	(0.67)	*(0.67)*	*(0.82)*	*(0.81)*	*(0.64)*	*(1.30)*	*(1.29)*
−1	0.77	0.79	0.95	0.93	0.84	0.97	0.97
	(0.67)	*(0.66)*	*(0.81)*	*(0.80)*	*(0.64)*	*(1.31)*	*(1.33)*
1	0.75	0.77	0.93	0.92	0.86	0.98	0.97
	(0.66)	*(0.66)*	*(0.81)*	*(0.80)*	*(0.64)*	*(1.35)*	*(1.34)*
−2	0.80	0.84	0.95	0.94	0.89	0.99	0.97
	(0.67)	*(0.67)*	*(0.82)*	*(0.81)*	*(0.65)*	*(1.36)*	*(1.34)*

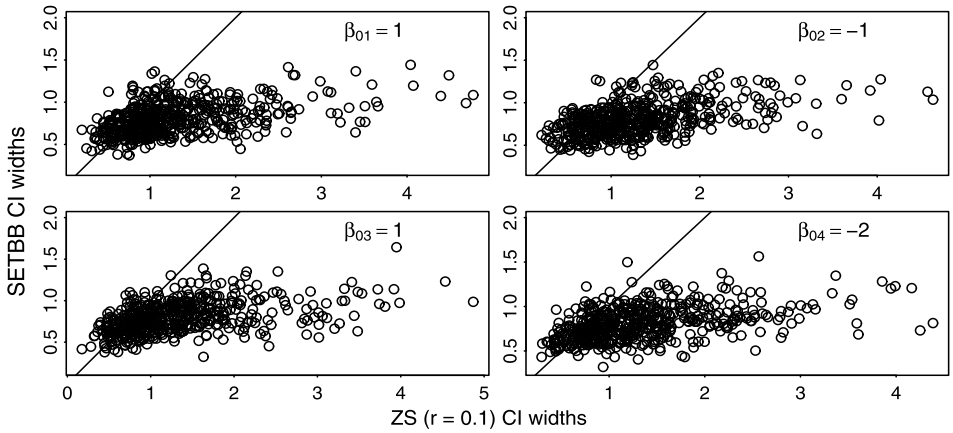

FIG. 5. *Widths of confidence intervals from the method of* [38] *with* $r = 0.1$ *against those of the SETBB under data-based block size and bandwidth choices for each quantile regression coefficient over* 500 *simulated data sets from the model with Chi-square errors and with* $(\Phi_1, \Phi_2, \Psi) = (0.7I_4, 0.1I_4, I_4)$ *and* $(\phi_1, \phi_2) = (0.7, 0.1)$.

that inference based on the ZS intervals would have low-power under these settings. Figure 5 plots the widths of the ZS confidence intervals against those of the SETBB confidence intervals over the 500 simulated datasets. The plot reveals a tendency of the ZS procedure to produce very wide confidence intervals. In this simulation, the ZS confidence intervals for $(\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}) = (1, -1, 1, -2)$ included zero (committed Type II errors) for 19%, 19%, 23% and 2.0% of the simulated data sets while those of the SETBB included zero for only 1%, 1%, 1% and 0% percent of the data sets—while maintaining nominal coverage.

In total, the results for covariance matrix estimation and distribution function estimation demonstrate clear advantages in combining block bootstrap methods with data smoothing in resampling, as the bootstrap procedures using the data perturbation performed much better than their unsmooth counterparts under all settings considered. See the Supplementary Material [11] for further simulation summaries.

6. Data examples.

6.1. *Weekly gas and oil prices* 2000–2005. Using weekly oil and gas price data from the R package `astsa` from [34] over 2002–2005 ($n = 260$), we model the 0.9 quantile of changes in the gas price conditional on the change in oil price. The left-hand panel of Figure 6 displays the weekly data, and the right-hand panel shows a scatterplot of the differenced gas price series against the differenced oil price series. The solid line is the QR fit for the 0.9 conditional quantile. According to the model, given a change in the oil prices, the change in the gas price in the

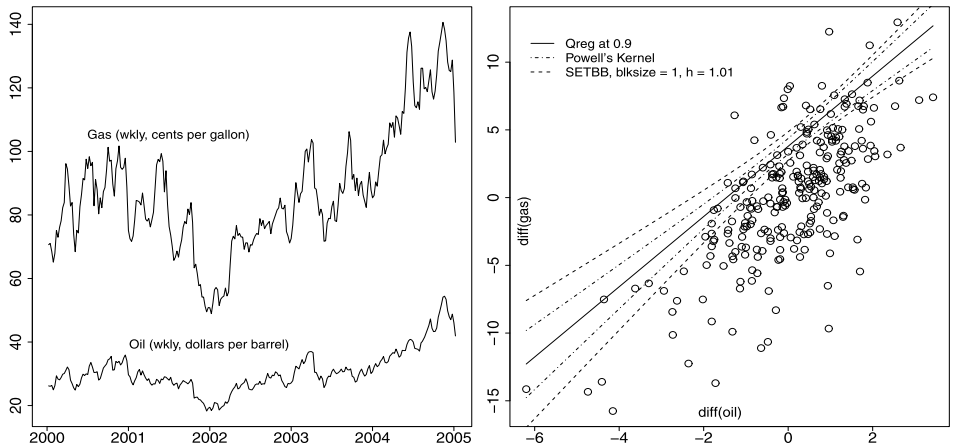

FIG. 6. *Weekly gas and oil prices from* 2000–2005. *Scatterplot of differenced series with* $\theta = 0.9$ *QR fit and confidence intervals from Powell's kernel and the SETBB overlaid.*

same week will exceed the height of the solid line with a probability of at most 0.1. Using the normal approximation based on Powell's kernel estimate for the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$, we construct 95% confidence intervals around the fitted line. We do the same using the estimated covariance matrix from the SETBB at the data-adaptive bandwidth and block size choices, getting wider confidence intervals for the position of the regression line.

6.2. *Value at risk for Dow Jones Industrial Average.* A common application of quantile regression in time series is the estimation of the value at risk (VaR) of a financial asset; see, for example, [6, 9] and [35]. The $100\theta\%$ VaR for a holding period k is the value which the incurred loss will exceed with a probability of at most $1 - \theta$ if the asset is held for a period of length k. For example, if the 90% one-day VaR of an asset is \$1000, then there is at most a 10% chance of losing \$1000 or more if the asset is held for one day. The $100\theta\%$ VaR is thus the θth quantile of the loss distribution.

Regulators are interested in the VaR of assets held by financial institutions, and traders often make decisions based on estimates of the future VaR based on present and past data. To this end, QR may be used to fit, for example, an autogressive model such as $Y_t = X_t'\beta_0 + \varepsilon_t(\theta, X_t)$, where Y_t is the loss in time period t, regressors $X_t = (1, Y_{t-1}, \dots, Y_{t-p})'$ are formulated in terms of past losses, and the error $\varepsilon_t(\theta, X_t)$ has θth quantile equal to zero. This model is a simple instance of the CAViaR model proposed in [6]. After obtaining $\hat{\beta}_n$ from QR based on n observations, the fitted value $X_{n+1}'\hat{\beta}_n$ is typically applied as an approximate lower $100(1-\theta)\%$ prediction bound for the future loss Y_{n+1} value (i.e., $X_{n+1}'\hat{\beta}_n$ is taken as the $100\theta\%$ VaR for Y_{n+1}), in light of the fact that $P(Y_{n+1} \geq X_{n+1}'\beta_0) = 1 - \theta$ under the model and $\hat{\beta}_n$ consistently estimates β_0.

Here, we consider the SETBB method for improving such one-step ahead VaR estimates based on QR, which also illustrates potential extensions of the SETBB approach (i.e., QR-based forecasting). In particular, for smaller samples, the bootstrap may be applied to enhance the estimate $X'_{n+1}\hat{\beta}_n$ as as an approximate lower $100(1-\theta)\%$ prediction limit for the future loss Y_{n+1}. Specifically, we use the bootstrap to estimate the value a_n for which

$$P\big(Y_{n+1} \geq X'_{n+1}\hat{\beta}_n + a_n\big) = 1 - \theta.$$

While $a_n \xrightarrow{p} 0$ holds as $n \to \infty$ by the consistency of the QR estimator, the term a_n may differ from zero in small samples due to errors associated with QR estimation in this problem (e.g., induced by using the same data for parameter estimation and one-step ahead prediction). We consequently approximate a_n from the SETBB estimator for the distribution of prediction errors $Y_{n+1} - X'_{n+1}\hat{\beta}_n = \varepsilon_{n+1} - X'_{n+1}(\hat{\beta}_n - \beta_0)$. A SETBB rendition of $\varepsilon_{n+1} - X'_{n+1}(\hat{\beta}_n - \beta_0)$ may be constructed as $\varepsilon^*_{n+1} - X^{*\prime}_{n+1}(\hat{\beta}^*_n - \tilde{\beta}_n)$, where, having defined perturbed observations $\{(Y^*_t, X^{*\prime}_t)'\}_{t=1}^n$ in the SETBB resampling steps for obtaining $\hat{\beta}^*_n = \arg\min_{\beta\in\mathbb{R}^d} S^*_n(\beta)$, the variables ε^*_{n+1} and X^*_{n+1} are drawn from resampling distributions for model errors and regressors, respectively, defined as

$$\varepsilon^*_{n+1} | \{(Y^*_t, X^*_t)\}_{t=1}^n \sim \textstyle\sum_{t=1}^n \tilde{w}_t \delta_{Y^*_t - X^{*\prime}_t \tilde{\beta}_n}$$
$$X^*_{n+1} | \{(Y^*_t, X^*_t)\}_{t=1}^n \sim \textstyle\sum_{t=1}^n \tilde{w}_t \delta_{(1, Y^*_t, \ldots, Y^*_{t-p+1})'}.$$

Above $\tilde{w}_t \equiv \mathrm{E}_*\pi^*_t = (n+\ell-1)^{-1}\|w_\ell\|_1^{-1}\sum_{j=1}^{n-\ell+1}\sum_{k=1}^{\ell} w_\ell(k) I(t = j+k-1)$ denotes a selection probability for observation $t = 1, \ldots, n$ under the SETBB resampling distribution and the SETBB version $\tilde{\beta}_n$ of β_0 is, as before, $\tilde{\beta}_n = \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}_* S^*_n(\beta)$. From this, bootstrap versions of prediction errors $\varepsilon^*_{n+1} - X^{*\prime}_{n+1}(\hat{\beta}^*_n - \tilde{\beta}_n)$ may be approximated via Monte Carlo simulation through the following algorithm:

Input: Data $\{(Y_t, X'_t)'\}_{t=1}^n$, number of bootstrap replicates B, block length ℓ, perturbation density G_Z with bandwidth H_n.

1: Compute SETBB version $\tilde{\beta}_n$ of β_0.
2: **For all** $b = 1, \ldots, B$**, repeat 3-7:**
3: Draw $Z_1^{*(b)}, \ldots, Z_n^{*(b)}$ independently from G_Z.
4: Set $(Y_t^{*(b)}, X_t^{*(b)\prime})' = (Y_t, X'_t)' + H_n Z_t^{*(b)}$ for $t = 1, \ldots, n$.
5: Draw bootstrap weights $\pi_1^{*(b)}, \ldots, \pi_n^{*(b)}$.
6: Compute $\hat{\beta}_n^{*(b)} = \arg\min_{\beta\in\mathbb{R}^d} \sum_{t=1}^n \pi_t^{*(b)}(Y_t^{*(b)} - X_t^{*(b)\prime}\beta)\mathrm{sign}_\theta(Y_t^{*(b)} - X_t^{*(b)\prime}\beta)$
7: Set $\varepsilon_t^{*(b)} = Y_t^{*(b)} - X_t^{*(b)\prime}\tilde{\beta}_n$ for $t = 1, \ldots, n$.

As an estimator of a_n, the θth quantile of the SETBB error distribution $Y^*_{n+1} - X^{*\prime}_{n+1}\hat{\beta}^*_n$ is then approximated as

$$\hat{a}_n = \inf_{a\in\mathbb{R}}\left\{\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{n}\sum_{s=1}^{n}\tilde{w}_t\tilde{w}_s\mathbb{I}\big(\varepsilon_t^{*(b)} - X_s^{*(b)\prime}(\hat{\beta}_n^{*(b)} - \tilde{\beta}_n) \le a\big) \ge \theta\right\},$$

and the SETBB-calibrated one-step ahead $100\theta\%$ VaR [i.e., a SETBB $(1-\theta)100\%$ lower prediction bound for future loss Y_{n+1}] is then given by $X'_{n+1}\hat{\beta}_n + \hat{a}_n$. We may also consider finding a conditional calibrator b_n of the VaR limit such that $P(Y_{n+1} \ge X'_{n+1}\hat{\beta}_n + b_n | X_{n+1}) = 1-\theta$ holds, given the immediate past $X_{n+1} = (1, Y_n, \dots, Y_{n-p})$ of loss observations. In which case, we set

$$\hat{b}_n = \inf_{b\in\mathbb{R}}\left\{\frac{1}{B}\sum_{b=1}^{B}\sum_{t=1}^{n}\tilde{w}_t\mathbb{I}\big(\varepsilon_t^{*(b)} - X'_{n+1}(\hat{\beta}_n^{*(b)} - \tilde{\beta}_n) \le b\big) \ge \theta\right\},$$

(i.e., set $X^*_{n+1} = X_{n+1}$ in creating bootstrap prediction errors) and define a conditionally calibrated VaR limit as $X'_{n+1}\hat{\beta}_n + \hat{b}_n$ using the SETBB.

For illustration, we consider predicting the 95% VaR for a one-day holding period of an asset with the same composition as the Dow Jones Industrial Average (DJIA) stock index, following [36]. We conduct a historical simulation in which we choose a day in the past and use the foregoing 50, 100 and 200 days to fit an AR(2) quantile regression model with which we predict the 0.95 quantile of the conditional loss distribution on our chosen day [we set $Y_t = -100^*(\log V_t - \log V_{t-1})$, where V_t is the value of the DJIA stock index at time t]. We repeat this for 1000 consecutive days beginning on April 11, 1996, and record whether or not the loss exceeded the 0.95 prediction limit as given by (i) $X'_{n+1}\hat{\beta}_n + \hat{b}_n$, (ii) $X'_{n+1}\hat{\beta}_n + \hat{a}_n$, and (iii) $X'_{n+1}\hat{\beta}_n$, considering SMBB and SETBB versions of $\hat{a}_n$ and $\hat{b}_n$. The NPPI method is used to select block size for each bootstrap and the bandwidth is chosen as described in Section 4.1.

Table 3 records the proportion of the 1000 historical days for which the loss exceeded the 95% VaR limit under conditional/unconditional bootstrap calibrations as well as with no bootstrap calibration. The standard prediction limit $X'_{n+1}\hat{\beta}_n$ for future loss Y_{n+1} was exceeded at the rates 0.09, 0.07 and 0.06 for the sample sizes $n = 50$, 100 and 200. This exceedance rate approaches the nominal rate of 0.05 as the sample size grows, suggesting that the model is not misspecified and that deviations from the 0.05 nominal rate are due to finite-sample effects. In comparison, the bootstrap-calibrated prediction limits performed better, with exceedance rates much closer to the nominal 0.05 level for all of the sample sizes. In particular, conditional prediction limits from the SETBB had excellent coverage rates: 0.06, 0.05 and 0.05 for the sizes $n = 50$, 100 and 200. This demonstrates the potential usefulness of the SETBB method in QR-based forecasting problems in addition to quantile parameter estimation.

TABLE 3
Proportion of historical days out of 1000 *where loss exceeded the* 95% *VaR limit* (*lower* 5% *prediction bound*) *under conditional and unconditional bootstrap calibrations under from SMBB/SETBB methods, as well as without bootstrap calibration* $X'_{n+1}\hat{\beta}_n$

		$X'_{n+1}\hat{\beta}_n + \hat{b}_n$	$X'_{n+1}\hat{\beta}_n + \hat{a}_n$	$X'_{n+1}\hat{\beta}_n$
$n = 50$	SMBB	0.07	0.07	0.09
	SETBB	0.06	0.06	0.09
$n = 100$	SMBB	0.06	0.06	0.07
	SETBB	0.05	0.06	0.07
$n = 200$	SMBB	0.06	0.05	0.06
	SETBB	0.05	0.05	0.06

7. Conclusions. Under mild conditions, we have established the consistency of the smooth extended tapered block bootstrap (SETBB) for the sampling distribution of the quantile regression (QR) estimator with time series. Our results also apply to unsmooth block bootstraps (i.e., the moving blocks bootstrap and extended tapered block bootstrap) and extend their validity to QR under weaker conditions than previously considered. Our simulation studies show that smoothing steps in the SETBB remarkably improve estimation over unsmoothed bootstrap counterparts. For the QR estimator, the SETBB also exhibited better performance in distribution and variance estimation than Powell's kernel method, and confidence intervals for quantiles from the SETBB were shown to achieve the nominal coverage while being narrower than intervals produced by the self-normalization method of [38]—all under data-based choices of the block size and the bandwidth. Our data examples also demonstrated the SETBB in practical applications.

8. Proofs of main results. We outline some details for proving the bootstrap results of Section 3; the Supplementary Material [11] offers complete proofs. Lemmas 1–3 provide supporting technical results, where we state only the most relevant parts of these lemmas for our purposes here; fuller versions appear in [11]. All limits are taken as $n \to \infty$. Lemma 1 first gives distributional and approximation results for the QR estimator $\hat{\beta}_n$.

LEMMA 1. *Assume Conditions* (C.1)–(C.5) *with* Σ, L *from* (C.4).

(iii) *For*

$$S_n(\beta) = n^{-1} \sum_{t=1}^{n} (Y_t - X'_t\beta)\operatorname{sign}_\theta(Y_t - X'_t\beta),$$

$\hat{\beta}_n \equiv \arg\min_{\beta \in \mathbb{R}^d} S_n(\beta)$ *exists for all* $n \geq 1$ (*a.s.* P) *and* $\hat{\beta}_n \xrightarrow{p} \beta_0$.

(iv) *Let* $D_n(\beta) = n^{-1}\sum_{t=1}^n X_t \operatorname{sign}_\theta(Y_t - X_t'\beta)$ *and* $\lambda_n(\beta) \equiv \mathrm{E}D_n(\beta) = \mathrm{E}X_1 \operatorname{sign}_\theta(Y_1 - X_1'\beta)$, $\beta \in \mathbb{R}^d$. *There exists some* $d_0 > 0$ *such that*

$$\sup_{\|\beta-\beta_0\|\le d_0} \frac{\|D_n(\beta) - D_n(\beta_0) - \lambda_n(\beta)\|}{n^{-1/2} + \|\lambda_n(\beta)\|} \xrightarrow{p} 0.$$

(v) $\sqrt{n} D_n(\hat\beta_n) = o_p(1)$.

(vi) $\sqrt{n} D_n(\beta_0) - \sqrt{n}\lambda(\hat\beta_n) \xrightarrow{p} 0$.

(vii) $\lambda_n(\hat\beta_n) = -L(\hat\beta_n - \beta_0) + o_p(\|\hat\beta_n - \beta_0\|)$.

(viii) $\sqrt{n}(\hat\beta_n - \beta_0) \xrightarrow{p} N(0, L^{-1}\Sigma L^{-1})$.

For $\beta \in \mathbb{R}^d$, recall the SETBB versions $S_n^*(\beta) = \sum_{t=1}^n \pi_t^*(Y_t^* - X^{*\prime}\beta) \times \operatorname{sign}_\theta(Y_t^* - X^{*\prime}\beta)$ and $\tilde S_n(\beta) \equiv \mathrm{E}_* S_n^*(\beta)$ of $S_n(\beta)$ and $\mathrm{E}S_n(\beta)$. Additionally, for the SETBB rendition $D_n^*(\beta) = \sum_{t=1}^n \pi_t^* X_t^* \operatorname{sign}_\theta(Y_t^* - X^{*\prime}\beta)$ from (3.1) of the derivative-like criterion $D_n(\beta) \equiv n^{-1}\sum_{t=1}^n X_t \operatorname{sign}_\theta(Y_t - X'\beta)$ (see Section 3.3), define $\tilde D_n(\beta) \equiv \mathrm{E}_* D_n^*(\beta)$. Lemma 2 next establishes distributional convergence for quantities related to the bootstrap expectation $\tilde D_n(\beta)$, having parallels to results for $D_n(\beta)$ in Lemma 1. Lemma 3 then provides distributional results, in the bootstrap world, for the resampling version $D_n^*(\beta)$ of $D_n(\beta)$.

LEMMA 2. *Assume Conditions* (C.1)–(C.7) *with* $\tilde\delta$ *from* (C.3):

(i) $\tilde\beta_n = \arg\min_{\beta\in\mathbb{R}^d} \tilde S_n(\beta)$ *in probability for large* n *and* $\tilde\beta_n \xrightarrow{p} \beta_0$.

(ii) $\beta_n = \arg\min_{\beta\in\mathbb{R}^d} \mathrm{E}\tilde S_n(\beta)$ *exists for all large* n, *satisfies* $\beta_n \longrightarrow \beta_0$, *and* $\tilde\lambda_n(\beta_n) = 0$ *holds for all large* n *where* $\tilde\lambda_n(\beta) \equiv \mathrm{E}\tilde D_n(\beta)$.

(vi) $\sqrt{n}\|\tilde D_n(\tilde\beta_n)\| = o_p(1)$, $\sqrt{n}\tilde D_n(\beta_n) \xrightarrow{d} N(0, \Sigma)$, $\sqrt{n}\tilde D_n(\beta_n) - \sqrt{n}[\tilde\lambda_n(\tilde\beta_n) - \tilde\lambda_n(\beta_n)] \xrightarrow{p} 0$ *and* $\sqrt{n}(\tilde\beta_n - \beta_n) \xrightarrow{d} N(0, L^{-1}\Sigma L^{-1})$.

(vii) *There exists some* $d_1 > 0$ *such that*

$$\sup_{\|\beta-\tilde\beta_n\|\le d_1} \frac{\|\tilde D_n(\beta) - \tilde D_n(\tilde\beta_n) - [\tilde\lambda_n(\beta) - \tilde\lambda_n(\tilde\beta_n)]\|}{n^{-1/2} + \|\tilde\lambda_n(\beta) - \tilde\lambda_n(\tilde\beta_n)\|} \xrightarrow{p} 0.$$

(x) $\tilde\beta_n - \beta_0 = O_p(n^{-1/2} + h_n)$ *for* $h_n = \|H_n\| \to 0$.

LEMMA 3. *Assume Conditions* (C.1)–(C.7). *Let* $\varrho_d[\cdot,\cdot]$ *denote a metric on distributions* (*the laws of random vectors*) *in* $\mathbb{R}^d$ *for some given* $d \ge 1$:

(i) $\varrho_d[\hat\beta_n^*, \beta_0] \xrightarrow{p} 0$ *and* $\varrho_d[\hat\beta_n^*, \tilde\beta_n] \xrightarrow{p} 0$.

(iii) $\varrho_d[m_\ell^{1/2} n^{1/2} D_n^*(\tilde\beta_n), W] \xrightarrow{p} 0$, *where* W *denotes a normal* $N(0, \Sigma)$ *random vector with* Σ *from* (C.3) *and* $m_\ell = \|w_\ell\|_1^2/[\ell\|w_\ell\|_2^2]$.

(v) $\varrho_1[\sqrt{n}\|D_n^*(\hat\beta_n^*)\|, 0] \xrightarrow{p} 0$.

(vi) *There exists some* $d_2 > 0$ *such that* $\varrho_1[\Delta_n^*, 0] \xrightarrow{p} 0$ *for*

$$\Delta_n^* \equiv \sup_{\|\tau-\tilde{\beta}_n\|\le d_2} \frac{\|D_n^*(\tau) - D_n^*(\tilde{\beta}_n) - [\tilde{D}_n(\tau) - \tilde{D}_n(\tilde{\beta}_n)]\|}{n^{-1/2} + \|\tilde{D}_n(\tau) - \tilde{D}_n(\tilde{\beta}_n)\|}.$$

PROOF OF THEOREMS 1–2. Theorem 1 follows from Lemma 1, while Theorem 2(ii)–(iii) follows from Lemma 2. To show Theorem 2(i), for any subsequence $\{n_j\} \subset \{n\}$, choose a further subsequence $\{k \equiv n_k\} \subset \{n_j\}$ where

$$\hat{\beta}_k^* - \tilde{\beta}_k \xrightarrow{p_*} 0, \tag{8.1}$$

$$\sqrt{k} D_k^*(\hat{\beta}_k^*) \xrightarrow{p_*} 0, \tag{8.2}$$

$$\sup_{\|\tau-\tilde{\beta}_k\|\le d_2} \frac{\|D_k^*(\tau) - D_k^*(\tilde{\beta}_k) - [\tilde{D}_k(\tau) - \tilde{D}_k(\tilde{\beta}_k)]\|}{k^{-1/2} + \|\tilde{D}_k(\tau) - \tilde{D}_k(\tilde{\beta}_k)\|} \xrightarrow{p_*} 0, \tag{8.3}$$

$$m_\ell^{1/2} k^{1/2} D_k^*(\tilde{\beta}_k) \xrightarrow{d_*} N(0, \Sigma), \tag{8.4}$$

$$\sup_{\|\beta-\tilde{\beta}_k\|\le d_1} \frac{\|\tilde{D}_k(\beta) - \tilde{D}_k(\tilde{\beta}_k) - [\tilde{\lambda}_k(\beta) - \tilde{\lambda}_k(\tilde{\beta}_k)]\|}{k^{-1/2} + \|\tilde{\lambda}_k(\tau) - \tilde{\lambda}_k(\tilde{\beta}_k)\|} \longrightarrow 0, \tag{8.5}$$

hold (a.s. P) by Lemma 3 and Lemma 2(vii). Above $\xrightarrow{p_*}$ and $\xrightarrow{d_*}$ denote convergence in bootstrap probability P_*; fixed $d_2, d_1 > 0$ are from Lemma 3(vi) and Lemma 2(vii); and $\tilde{\lambda}_n(\beta) \equiv \mathrm{E}\tilde{D}_n(\beta)$, $\beta \in \mathbb{R}^d$, is from Lemma 2. In the following, all convergence (a.s. P) is along the subsequence $k \equiv n_k \to \infty$.

From (8.1) and the proof of Lemma 2(v), we have that

$$\begin{aligned}\tilde{\lambda}_k(\hat{\beta}_k^*) - \tilde{\lambda}_k(\tilde{\beta}_k) + L(\hat{\beta}_k^* - \tilde{\beta}_k) &= -G_k(\hat{\beta}_k^* - \tilde{\beta}_k) + o_{p_*}(\|\hat{\beta}_k^* - \tilde{\beta}_k\|) \\ &= o_{p_*}(\|\hat{\beta}_k^* - \tilde{\beta}_k\|),\end{aligned}$$

where $L \equiv \mathrm{E}X_t X_t' f(0|X_t)$ and $G_n \equiv J_1 H_n \mathrm{E}ZZ'H_n' J_1' \mathrm{E}f(0|X_t) = o(1)$ [for $Z \sim G_Z$ from (C.6) and diagonal $(d+1) \times (d+1)$ matrix J_1 with entries $(0, 1_d')$] by $h_n = \|H_n\| \to 0$. By this, (8.1) and (8.5), we have

$$\sqrt{k}[\tilde{D}_k(\hat{\beta}_k^*) - \tilde{D}_k(\tilde{\beta}_k)] + L\sqrt{k}(\hat{\beta}_k^* - \tilde{\beta}_k) = o(1) + o_{p_*}(\sqrt{k}\|\hat{\beta}_k^* - \tilde{\beta}_k\|). \tag{8.6}$$

By (8.1)–(8.3), it holds that $\sqrt{k} D_k^*(\tilde{\beta}_k) - \sqrt{k}[\tilde{D}_k(\hat{\beta}_k^*) - \tilde{D}_k(\tilde{\beta}_k)] \xrightarrow{p_*} 0$ directly from the proof of [17], Theorem 3. By this and (8.6), we then have

$$\sqrt{k} D_k^*(\tilde{\beta}_k) + L\sqrt{k}(\hat{\beta}_k^* - \tilde{\beta}_k) = o(1) + o_{p^*}(\sqrt{k}\|\hat{\beta}_k^* - \tilde{\beta}_k\|).$$

As L is positive definite from (C.4) and $m_\ell = O(1)$, it now follows from (8.4) that $\sqrt{k}\|\hat{\beta}_k^* - \tilde{\beta}_k\| = O_{p_*}(1)$ and

$$m_\ell^{1/2}\sqrt{k}(\hat{\beta}_k^* - \tilde{\beta}_k) = -L^{-1} m_\ell^{1/2}\sqrt{k} D_k^*(\tilde{\beta}_k) + o_{p_*}(1) \xrightarrow{d} N(0, L^{-1}\Sigma L^{-1}).$$

As the subsequence $\{n_j\}$ was arbitrary, the proof is now complete by Pólya's theorem and Theorem 1(ii). $\square$

Acknowledgments. The authors are very grateful to the Editor Ed George, an Associate Editor and several referees for their time and thoughtful comments, which greatly helped to improve the manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "A smooth block bootstrap for quantile regression with time series" (DOI: 10.1214/17-AOS1580SUPP; .pdf). Details of proofs and additional simulation results.

REFERENCES

[1] ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990) (R. LePage and L. Billard, eds.) 13–47. Wiley, New York. MR1197777

[2] BUCHINSKY, M. (1994). Changes in the U.S. wage structure 1963–1987: Application of qunatile regression. *Econometrica* **62** 405–458.

[3] DE ANGELIS, D., HALL, P. and YOUNG, G. A. (1993). A note on coverage error of bootstrap confidence intervals for quantiles. *Math. Proc. Cambridge Philos. Soc.* **114** 517–531. MR1235999

[4] 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

[5] DOUKHAN, P. (1994). *Mixing*: *Properties and Examples*. *Lecture Notes in Statistics* **85**. Springer, New York. MR1312160

[6] ENGLE, R. F. and MANGANELLI, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* **22** 367–381. MR2091566

[7] FENG, X., HE, X. and HU, J. (2011). Wild bootstrap for quantile regression. *Biometrika* **98** 995–999. MR2860339

[8] FITZENBERGER, B. (1998). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *J. Econometrics* **82** 235–287. MR1613422

[9] GAGLIANONE, W. P., LIMA, L. R., LINTON, O. and SMITH, D. R. (2011). Evaluating value-at-risk models via quantile regression. *J. Bus. Econom. Statist.* **29** 150–160. MR2789438

[10] GREGORY, K. B., LAHIRI, S. N. and NORDMAN, D. J. (2015). A smooth block bootstrap for statistical functionals and time series. *J. Time Series Anal.* **36** 442–461. MR3343010

[11] GREGORY, K. B., LAHIRI, S. N. and NORDMAN, D. J. (2018). Supplement to "A smooth block bootstrap for quantile regression with time series." DOI:10.1214/17-AOS1580SUPP.

[12] 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

[13] HAHN, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory* **11** 105–121. MR1325103

[14] HASAN, M. N. and KOENKER, R. W. (1997). Robust rank tests of the unit root hypothesis. *Econometrica* **65** 133–161. MR1433687

[15] HE, X., ZHU, Z.-Y. and FUNG, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89** 579–590. MR1929164

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

[17] HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I*: *Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620

[18] KATO, K. (2012). Asymptotic normality of Powell's kernel estimator. *Ann. Inst. Statist. Math.* **64** 255–273. MR2878905

[19] KOENKER, R. (1994). Confidence intervals for regression quantiles. In *Asymptotic Statistics* (*Prague*, 1993) (P. Mandl and M. Hušková, eds.) 349–359. Physica, Heidelberg. MR1311953

[20] KOENKER, R. (2013). quantreg: Quantile regression. R package version 5.05.

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

[22] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147

[23] LAHIRI, S. N. (2002). On the jackknife-after-bootstrap method for dependent data and its consistency properties. *Econometric Theory* **18** 79–98. MR1885351

[24] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York. MR2001447

[25] LAHIRI, S. N., FURUKAWA, K. and LEE, Y.-D. (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Stat. Methodol.* **4** 292–321. MR2380557

[26] LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990) 225–248. Wiley, New York. MR1197787

[27] PAPARODITIS, E. and POLITIS, D. N. (2001). Tapered block bootstrap. *Biometrika* **88** 1105–1119. MR1872222

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

[29] POLLARD, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1** 295–313.

[30] POWELL, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (*Durham*, *NC*, 1988) 357–384. Cambridge Univ. Press, Cambridge. MR1174980

[31] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

[32] SHAO, X. (2010). Extended tapered block bootstrap. *Statist. Sinica* **20** 807–821. MR2682643

[33] SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **53** 683–690. MR1125725

[34] STOFFER, D. (2014). astsa: Applied statistical time series analysis. R package version 1.3.

[35] UMANTSEV, L. and CHERNOZHUKOV, V. (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empir. Econom.* **26** 271–292.

[36] VAN DEN GOORBERGH, R. W. J. and VLAAR, P. J. G. (1999). Value-at-risk analysis of stock returns historical simulation, variance techniques or tail index estimation? DNB Staff Reports (discontinued) No. 40, Netherlands Central Bank. Available at http://ideas.repec.org/p/dnb/staffs/40.html.

[37] WEISS, A. A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* **7** 46–68. MR1101211

[38] ZHOU, Z. and SHAO, X. (2013). Inference for linear models with dependent errors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 323–343. MR3021390

K. B. GREGORY
DEPARTMENT OF STATISTICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SOUTH CAROLINA 29201
USA
E-MAIL: gregorkb@stat.sc.edu

S. N. LAHIRI
DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695-8203
USA
E-MAIL: snlahiri@ncsu.edu

D. J. NORDMAN
DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011
USA
E-MAIL: dnordman@iastate.edu