Quantile Curve Estimation and Visualization for Nonstationary Time Series

Dana DRAGHICESCU, Serge GUILLAS, and Wei Biao WU

There is an increasing interest in studying time-varying quantiles, particularly for environmental processes. For instance, high pollution levels may cause severe respiratory problems, and large precipitation amounts can damage the environment, and have negative impacts on the society. In this article we address the problem of quantile curve estimation for a wide class of nonstationary and/or non-Gaussian processes. We discuss several nonparametric quantile curve estimates, give asymptotic results, and propose a data-driven procedure for the selection of smoothing parameters. This methodology provides a statistically reliable and computationally efficient graphical tool that can be used for the exploration and visualization of the behavior of time-varying quantiles for nonstationary time series. A Monte Carlo simulation study and two applications to ozone time series illustrate our method.

R codes with the algorithm for selection of smoothing parameters (described in Section 3) are available in the online supplements.

Key Words: Nonstationary time series; Ozone; Quantile estimation; Quantile visualization; Smoothing.

1. INTRODUCTION

Much research has been done in the past decades to extend quantile estimation methods for independent and identically distributed (iid) data to more general processes. Quantiles can be estimated by inverting estimates of the distribution function (Sheather and Marron 1990; Huang and Brill 1999), or directly via quantile regression (Koenker and Bassett 1978). For both approaches parametric and nonparametric methods are available. Inversion methods allow inference on the whole evolution of the process, the drawback being the computational time. Quantile estimates are useful when the problem of interest is to make inference only on a specific quantile.

Dana Draghicescu is Assistant Professor, Department of Mathematics and Statistics, Hunter College of the City University of New York, New York, NY 10065 (E-mail: *dana.draghicescu@hunter.cuny.edu*). Serge Guillas is Lecturer, Department of Statistical Science, University College London, London WC1E 6BT, U.K. (E-mail: *serge@stats.ucl.ac.uk*). Wei Biao Wu is Associate Professor, Department of Statistics, University of Chicago, Chicago, IL 60637 (E-mail: *wbwu@galton.uchicago.edu*).

For iid data parametric approaches were discussed by Koenker and Portnoy (1987) and the references therein. Nonparametric estimation of the quantile function was also considered under certain smoothness assumptions. For example, Koenker, Pin, and Portnoy (1994) and Cox (1983) used smoothing splines. Kernel smoothing was proposed, among others, by Bhattacharya and Gangopadhyay (1990); Chaudhuri (1991); and Yu and Jones (1998). Other approaches include rank nearest-neighbor methods (Yang 1981), penalized likelihood (Cole and Green 1992), and neural networks (White 1990). A comprehensive review of the applications of quantile regression and current research areas was given by Yu, Lu, and Stander (2003) and Koenker (2005).

Recent studies considered the dependent case as well. Quantile processes were analyzed by Mehra et al. (1992); Fotopoulos, Ahn, and Cho (1994); and Csörgő and Yu (1996). Estimation of quantiles under strong mixing conditions was dealt with by Degenhardt et al. (1996); Olsson and Rootzen (1996); Yu and Jones (1998); Mukherjee (1999); Cai (2002); and Abberger and Heiler (2002). Wu (2005b) gave asymptotic representations of sample quantiles for stationary processes.

For a random variable X with distribution function F, the αth quantile is defined as $q_\alpha(X) = \inf\{x : F(x) \geq \alpha\}$, $0 < \alpha < 1$. Given a sequence of random variables $X_1, \ldots, X_k$, the αth sample quantile is defined as $\hat{q}_\alpha(X_1, \ldots, X_k) = \inf\{x : F_k(x) \geq \alpha\}$, where $F_k(x) = k^{-1} \sum_{i=1}^{k} \mathbf{1}_{X_i \leq x}$ is the empirical distribution function. When the observations are temporally inhomogeneous, the marginal probability distribution (and thus the quantiles) of the X_i's may vary with time as well. In this situation $F(x; t) = P[X(t) \leq x]$, assumed to be a continuous function in the second argument, can be estimated more reliably by using a weighted average of the above indicators. The dependence on time is implicitly carried in the index i of each realization X_i, which is viewed as $X_i = X(t_i)$, with $t_i = i/k$, $i = 1, \ldots, k$, being rescaled time points. We can then estimate F by $\hat{F}(x; t) = \sum_{i=1}^{k} w_i \mathbf{1}_{X_i \leq x}$, $\sum_{i=1}^{k} w_i = 1$, the weights $w_i = w(t_i; t)$ accounting for the temporal inhomogeneity. One possibility for assigning these weights w_i is via kernel smoothing, a technique that was introduced by Rosenblatt (1956) for estimating density functions of iid data and widely applied since, more recently for dependent data as well. A kernel estimator can be viewed as the convolution of a smooth, known function (the kernel) with a rough empirical estimator, to produce a smooth estimator. Well-known kernel estimators are the *Nadaraya–Watson* (Nadaraya 1964; Watson 1964), *Gasser–Müller* (Gasser and Müller 1979), and *Priestley–Chao* (Priestley and Chao 1972) kernel estimators. An example of a kernel estimator for $F(x; t)$ is given by (4.2) in Section 4. For details on kernel estimation we refer to Wand and Jones (1995).

One way to generate dependent, nonstationary processes is via a time-varying transformation $G(t; Z_t)$ of a stationary process Z_t. By allowing the unknown transformation G to vary with time, the probability distribution function of the resulting process may also change, and therefore the process need not be stationary. The nonstationary process $G(t; Z_t)$ is then subordinated by a stationary process. Ghosh, Beran, and Innes (1997) studied asymptotic properties of a nonparametric conditional quantile estimator (obtained by inverting a kernel estimator of the probability distribution function) in this setting, where the underlying process Z_t was assumed to be Gaussian and having long memory. A similar

estimator was analyzed by Draghicescu and Ghosh (2003) for the case when the underlying Gaussian process has short memory (under the general assumption that the correlations are summable). A data-driven procedure for optimal bandwidth selection for these kernel quantile estimators was proposed by Ghosh and Draghicescu (2002).

In this article we propose quantile estimation methods for a wide class of stochastic processes, that allow for nonstationarity and/or non-Gaussianity. Section 2 introduces the theoretical model. We consider time-varying transformations of stationary processes (without assuming Gaussianity) and obtain asymptotic properties of a moving window quantile curve estimator. An automatic, data-driven procedure for the selection of smoothing parameters is introduced in Section 3. Sections 4 and 5 present a simulation study, and applications to ground-level and stratospheric ozone data, respectively. A brief discussion is given in Section 6.

2. QUANTILE CURVES OF NONSTATIONARY TIME SERIES

Quantile estimation for time series data involves the analysis of order statistics of dependent random variables. It is unfortunately quite challenging to obtain asymptotic properties of the order statistics if dependence is present. The assumption of independent order statistics is widely adopted in the literature; see, for example, Shorack and Wellner (1986). Recently Wu (2005b) established Bahadur representations for order statistics of stationary processes including linear processes and some widely used nonlinear processes.

Beside the dependence, the issue of nonstationarity further complicates the related study. To the best of our knowledge, there is virtually no prior work in the literature concerning asymptotic behavior of order statistics of nonstationary processes under general dependence structures. The goal of this section is to establish an asymptotic theory for quantile curve estimates of nonstationary time series. It is certainly necessary to impose structural assumptions on the underlying processes. We introduce a model for nonstationary processes which we call *nearly stationary processes*, based on which we shall present an asymptotic theory.

Definition 1 (Nearly stationary processes)**:**   Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be iid random variables, $Z_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots)$, and G be a measurable function such that

$$X_i = X_{i,n} = G(i/n; Z_i), \qquad 1 \le i \le n, \tag{2.1}$$

are proper random variables. Let $F(x; t) = P[G(t; Z_i) \le x], x \in \mathbb{R}, 0 \le t \le 1$. We say that the process $(X_i)_{i=1}^n$ is *nearly stationary* if there exists a constant $L < \infty$ such that for all $0 \le t, t' \le 1$,

$$\sup_{x\in\mathbb{R}} |F(x; t) - F(x; t')| \le L|t - t'|. \tag{2.2}$$

If the process $(X_i)_{i=1}^n$ is stationary, then $F(x; t) = P(X_i \le x) = F(x)$. Since the latter does not depend on t, one can let $L = 0$ in (2.2). Intuitively, the uniform Lipschitz continuity condition (2.2) implies that the distributions of X_i and X_j are close if $|i/n - j/n|$ is small, hence suggesting near stationarity.

The class of nearly stationary processes is quite general. Wiener (1958) claimed that, for every stationary ergodic process $(X_i)_{i \in \mathbb{Z}}$, there exist iid standard uniform$(0, 1)$ random variables ε_k, $k \in \mathbb{Z}$, and a measurable function G such that the distributional equality

$$(X_i)_{i \in \mathbb{Z}} =_{\mathcal{D}} (G(Z_i))_{i \in \mathbb{Z}}, \quad \text{where } Z_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots) \tag{2.3}$$

holds (see also Kallianpur 1981; Priestley 1988; Tong 1990; Borkar 1993). In this sense, by letting Z_i be an *infinite-dimensional* shift (one-sided) process instead of a one-dimensional process, one can allow for a more general class of processes. To account for nonstationarity with slowly changing data-generating mechanisms, it is natural to let $G(Z_i)$ in (2.3) depend on the time index i in the manner of (2.1), with $F(\cdot, \cdot)$ satisfying (2.2).

An important example of nearly stationary processes is the mean nonstationary model $X_i = m(i/n) + e_i$, where $(e_i)_{i \in \mathbb{Z}}$ are iid with distribution function $F_e(x) = P(e_i \leq x)$ and bounded density function $f_e = F_e'$. Then $F(x; t) = F_e(x - m(t))$. If $m(\cdot)$ is Lipschitz continuous, then (2.2) holds, and the process $(X_i)_{i=1}^n$ is nearly stationary. Fan and Yao (2003, p. 226) argued that modeling the mean trend as $m(i/n)$ is a simple way to capture the feature that the trend is much more slowly varying than the noise. Hall and Hart (1990) and Johnstone and Silverman (1997) considered estimating mean trends of mean nonstationary models with dependent errors. Another example is given by nonlinear heteroscedastic processes. In Section 2.2 we show that such processes are nearly stationary.

The concept of near stationarity is closely related to locally stationary processes; see Priestley (1965); Dahlhaus (1997); Adak (1998); Ombao et al. (2002); Giurcanu and Spokoiny (2002). However, most work on locally stationary processes concerns covariance based inference (such as the estimation of time-varying spectrum), whereas near stationarity imposes conditions on the distribution function F instead of the second moments [relation (2.2)], and could allow for infinite variances. Indeed, our primary goal is to estimate quantiles, and we do not need the finite second-moment assumption.

As pointed out by Yu, Lu, and Stander (2003) and Koenker (2005), quantile regression provides more detailed distributional information about the underlying statistical model than the mean regression fits. In our setting, by estimating the quantile curves

$$q_\alpha(t) = \inf\{x : F(x; t) \geq \alpha\}, \qquad 0 \leq t \leq 1, \tag{2.4}$$

at different values of $\alpha \in (0, 1)$, we can obtain detailed distributional information such as changes in the variability of the process. Note that $q_\alpha(t)$ is the αth quantile of $G(t; Z_i)$, $0 \leq t \leq 1$. Condition (2.2) in the definition of nearly stationary processes suggests that $q_\alpha(t)$ is a continuous function of t. Hence, for a fixed $t_0 \in (0, 1)$, we can use observations X_i with i close to $\lfloor nt_0 \rfloor$ to estimate $q_\alpha(t_0)$ via smoothing. Specifically, let b_n be a bandwidth sequence satisfying

$$b_n \to 0 \qquad \text{and} \qquad nb_n \to \infty. \tag{2.5}$$

Let $n_1 = \lfloor n(t_0 - b_n) \rfloor$, $n_2 = \lfloor n(t_0 + b_n) \rfloor$, and $m = n_2 - n_1 + 1$. Based on X_i, $n_1 \leq i \leq n_2$, we can estimate $q_\alpha(t_0)$ by the sample quantile

$$\hat{q}_\alpha(t_0) = \hat{q}_\alpha(X_{n_1}, \ldots, X_{n_2}) = \inf\left\{ x : \frac{1}{m} \sum_{i=n_1}^{n_2} \mathbf{1}_{X_i \leq x} \geq \alpha \right\}. \tag{2.6}$$

Theorem 1 provides a bound for the error $\hat{q}_\alpha(t_0) - q_\alpha(t_0)$, $0 < t_0 < 1$. For a random variable ξ define the $\mathcal{L}^2$ norm $\|\xi\| = [E(|\xi|^2)]^{1/2}$. Recall $Z_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots)$. Define the projection operator $\mathcal{P}_k$, $k \in \mathbb{Z}$, by

$$\mathcal{P}_k \cdot = E(\cdot | Z_k) - E(\cdot | Z_{k-1}). \tag{2.7}$$

Theorem 1. *Let* $(X_i)_{i=1}^n$ *be a nearly stationary process. Assume that*

$$\sum_{j=1}^\infty \omega_j < \infty, \quad \text{where } \omega_j = \sup_{0 \le t \le 1} \sup_{x \in \mathbb{R}} \|\mathcal{P}_0 \mathbf{1}_{G(t;Z_j) \le x}\|. \tag{2.8}$$

Further assume that $f(x; t) = \partial F(x; t)/\partial x$, *the density of* $G(t; Z_j)$, $0 \le t \le 1$, *is a continuous function in* (x, t). *Let* $t_0 \in (0, 1)$ *and assume*

$$f(q_\alpha(t_0); t_0) > 0. \tag{2.9}$$

Then under (2.5),

$$\hat{q}_\alpha(t_0) - q_\alpha(t_0) = O_\mathbb{P}[b_n + (nb_n)^{-1/2}]. \tag{2.10}$$

Before we prove the theorem, we shall comment on conditions (2.8) and (2.9). By the continuity of $f(x; t)$, condition (2.9) implies that, for sufficiently small $\delta > 0$, $\min_{|t-t_0| \le \delta} f(q_\alpha(t_0); t) > 0$. Hence, among $X_{n_1}, \ldots, X_{n_2}$, there will be considerably many data points that are contained in a neighborhood of $q_\alpha(t_0)$, thus ensuring the reliability of the estimate $\hat{q}_\alpha(t_0)$ in (2.6). Condition (2.8) is basically a short-range dependence assumption. As argued by Wu (2005a), $\sup_{x \in \mathbb{R}} \|\mathcal{P}_0 \mathbf{1}_{G(t;Z_j) \le x}\|$ measures the contribution of ε_0 in predicting $G(t; Z_j)$. Therefore (2.8) implies that the cumulative contribution of ε_0 in predicting future values is finite, thus suggesting short-range dependence. Because ω_j is directly related to the data-generating mechanism of the underlying process, (2.8) is easy to deal with. Sections 2.1 and 2.2 give applications of this result to nonstationary linear processes and nonlinear heteroscedastic processes, respectively.

Proof of Theorem 1: Let $\theta_n = b_n + (nb_n)^{-1/2}$, $m = n_2 - n_1 + 1$, and define

$$S_n(x) = \frac{1}{m} \sum_{i=n_1}^{n_2} \mathbf{1}_{X_i \le x} \quad \text{and} \quad s_n(x) = \frac{1}{m} \sum_{i=n_1}^{n_2} F(x; i/n).$$

Note that $\mathbf{1}_{X_i \le x} - F(x; i/n) = \sum_{k=0}^\infty \mathcal{P}_{i-k} \mathbf{1}_{X_i \le x}$ and, for fixed k, $\mathcal{P}_{i-k} \mathbf{1}_{X_i \le x}$, $i = n_1, \ldots, n_2$, are martingale differences. By the triangle inequality and the orthogonality of martingale differences,

$$m \|S_n(x) - s_n(x)\| = \left\| \sum_{i=n_1}^{n_2} \sum_{k=0}^\infty \mathcal{P}_{i-k} \mathbf{1}_{X_i \le x} \right\|$$

$$\le \sum_{k=0}^\infty \left\| \sum_{i=n_1}^{n_2} \mathcal{P}_{i-k} \mathbf{1}_{X_i \le x} \right\|$$

$$= \sum_{k=0}^\infty \left[\sum_{i=n_1}^{n_2} \|\mathcal{P}_{i-k} \mathbf{1}_{X_i \le x}\|^2 \right]^{1/2} \le \sum_{k=0}^\infty \sqrt{m} \omega_j.$$

Thus $S_n(x) - s_n(x) = O_P(m^{-1/2}) = O_P((nb_n)^{-1/2})$. To show (2.10), let τ_n be a positive sequence such that $\tau_n \to \infty$ and $\theta_n \tau_n \to 0$. Let $c_* = f(q_\alpha(t_0); t_0) > 0$. By the continuity of $f(x; t)$, we can choose $\delta > 0$ such that $\min_{|t-t_0| \le \delta} f(q_\alpha(t_0); t) > c_*/2$. By (2.2),

$$s_n(q_\alpha(t_0) + \theta_n \tau_n) - F(q_\alpha(t_0); t_0)$$

$$\ge F(q_\alpha(t_0) + \theta_n \tau_n; t_0) - F(q_\alpha(t_0); t_0)$$

$$- \frac{1}{m} \sum_{i=n_1}^{n_2} \left| F(q_\alpha(t_0) + \theta_n \tau_n; i/n) - F(q_\alpha(t_0) + \theta_n \tau_n; t_0) \right|$$

$$\ge (c_*/2)\theta_n \tau_n - L b_n.$$

Because $S_n(q_\alpha(t_0) + \theta_n \tau_n) - s_n(q_\alpha(t_0) + \theta_n \tau_n) = O_P(m^{-1/2})$ and $F(q_\alpha(t_0); t_0) = \alpha$, we have with probability approaching 1 that

$$S_n(q_\alpha(t_0) + \theta_n \tau_n) - \alpha = s_n(q_\alpha(t_0) + \theta_n \tau_n) + O_P(m^{-1/2})$$

$$\ge (c_*/2)\theta_n \tau_n - L b_n + O_P(m^{-1/2}) > (c_*/3)\theta_n \tau_n.$$

Note that $S_n(\cdot)$ is nondecreasing, $P[\hat{q}_\alpha(t_0) \le q_\alpha(t_0) + \theta_n \tau_n] \to 1$. Similarly, $P[\hat{q}_\alpha(t_0) \ge q_\alpha(t_0) - \theta_n \tau_n] \to 1$. Thus $\hat{q}_\alpha(t_0) - q_\alpha(t_0) = O_P(\theta_n)$ because the sequence $\tau_n \to \infty$ is arbitrarily chosen. $\quad \square$

2.1 NEARLY STATIONARY LINEAR PROCESSES

Let $a_j(t)$, $j \ge 0$, be real functions on $[0, 1]$ and let ε_k, $k \in \mathbb{Z}$, be iid random variables with distribution and density functions F_ε and f_ε, respectively. Consider the nonstationary process

$$X_i = X_{i,n} = \sum_{j=0}^{\infty} a_j(i/n)\varepsilon_{i-j}, \qquad i = 1, \dots, n. \tag{2.11}$$

In the context of spectral analysis of locally stationary processes, such models were also considered in Dahlhaus (1997). To apply Theorem 1 to estimate quantile curves, we need to impose some regularity conditions on $a_j(\cdot)$ and F_ε.

Assume that $\sup_{x \in \mathbb{R}} |x f_\varepsilon(x)| < \infty$, $E(|\varepsilon_j|^p) < \infty$, $1 \le p \le 2$, and

$$\sum_{j=0}^{\infty} \sup_{t \in [0,1]} |a_j(t)|^{p/2} < \infty. \tag{2.12}$$

Further assume that $a_j(\cdot)$ are Lipschitz continuous, namely, there exists $l_j < \infty$ such that $|a_j(t) - a_j(t')| \le l_j |t - t'|$ holds for all $t, t' \in [0, 1]$, $\min_{t \in [0,1]} |a_0(t)| > 0$, and

$$\sum_{j=0}^{\infty} l_j < \infty. \tag{2.13}$$

To show that $(X_i)_{i=1}^n$ is nearly stationary, let $Y_i(t) = \sum_{j=1}^{\infty} a_j(t)\varepsilon_{i-j}$ and $G(t; Z_i) = \sum_{j=0}^{\infty} a_j(t)\varepsilon_{i-j} = a_0(t)\varepsilon_i + Y_i(t)$. Because $a_0(\cdot)$ is a continuous function and $\min_{t \in [0,1]} |a_0(t)| > 0$, we have either $\min_{t \in [0,1]} a_0(t) > 0$ or $\max_{t \in [0,1]} a_0(t) < 0$. Without loss of generality we assume the former. Let $c_0 = \sup_{x \in \mathbb{R}} (|x| + 1) f_\varepsilon(x) < \infty$

and $\rho(t, t') = a_0(t)/a_0(t')$. Elementary calculations show that $|F_\varepsilon(x) - F_\varepsilon(x\rho(t, t'))| \leq c_0|1 - \rho(t, t')|/\min(1, \rho(t, t'))$. Hence

$$\begin{aligned}
|F(x; t) - F(x; t')| &= \left|E[a_0(t)\varepsilon_i + Y_i(t) \leq x|Z_{i-1}] - E[a_0(t')\varepsilon_i + Y_i(t') \leq x|Z_{i-1}]\right| \\
&\leq E\left|F_\varepsilon\big((x - Y_i(t))/a_0(t)\big) - F_\varepsilon\big((x - Y_i(t'))/a_0(t')\big)\right| \\
&\leq E\left|F_\varepsilon\big((x - Y_i(t))/a_0(t)\big) - F_\varepsilon\big((x - Y_i(t))/a_0(t')\big)\right| \\
&\quad + E\left|F_\varepsilon\big((x - Y_i(t))/a_0(t')\big) - F_\varepsilon\big((x - Y_i(t'))/a_0(t')\big)\right| \\
&\leq c_0|1 - \rho(t, t')|/\min(1, \rho(t, t')) + c_0 E|Y_i(t) - Y_i(t')|/|a_0(t')|.
\end{aligned}$$

Because $a_j(\cdot)$ are Lipschitz continuous and $E|\varepsilon_k| < \infty$, the above inequality and assumption (2.13) yield condition (2.22), and thus the process X_i given by (2.11) is nearly stationary.

To check (2.8), let $\varepsilon_0', \varepsilon_k, k \in \mathbb{Z}$, be iid and let $Z_i^* = (\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \ldots)$ be a coupled process of Z_i with ε_0 replaced by ε_0'. Let $Y_i^*(t) = Y_i(t) - a_i(t)\varepsilon_0 + a_i(t)\varepsilon_0'$. For $i \geq 1$, $E[\mathbf{1}_{G(t;Z_i)\leq x}|Z_{i-1}] = F_\varepsilon((x - Y_i(t))/a_0(t))$ and, by independence,

$$\begin{aligned}
E\left[F_\varepsilon\big((x - Y_i(t))/a_0(t)\big)|Z_{-1}\right] &= E\left[F_\varepsilon\big((x - Y_i^*(t))/a_0(t)\big)|Z_{-1}\right] \\
&= E\left[F_\varepsilon\big((x - Y_i^*(t))/a_0(t)\big)|Z_0\right].
\end{aligned}$$

Because $\min(1, |t|) \leq |t|^{p/2}$, $1 \leq p \leq 2$, and $E(|\varepsilon_j|^p) < \infty$, (2.8) follows from (2.12) in view of

$$\begin{aligned}
\left\|\mathcal{P}_0 \mathbf{1}_{G(t;Z_i)\leq x}\right\| &= \left\|\mathcal{P}_0 F_\varepsilon\big((x - Y_i(t))/a_0(t)\big)\right\| \\
&= \left\|E\left[F_\varepsilon\big((x - Y_i(t))/a_0(t)\big) - F_\varepsilon\big((x - Y_i^*(t))/a_0(t)\big)|Z_0\right]\right\| \\
&\leq c_0 \left\|\min\left[2, |Y_i(t) - Y_i^*(t)|/|a_0(t)|\right]\right\| \\
&\leq 2c_0 \left\|\left[|Y_i(t) - Y_i^*(t)|/|a_0(t)|\right]^{p/2}\right\| = O\left[|a_i(t)|^{p/2}\right]
\end{aligned}$$

via Jensen's inequality. We emphasize that our conditions allow processes with infinite variances. For example, if ε_i are iid standard symmetric-α-stable with characteristic function $E(\exp(\sqrt{-1}u\varepsilon_i)) = \exp(-|u|^\alpha)$ with $1 < \alpha < 2$, then $E(\varepsilon_i^2) = \infty$ and $E(|\varepsilon_i|^p) < \infty$ for $0 < p < \alpha$ (see Samorodnitsky and Taqqu 1994).

2.2 NONLINEAR HETEROSCEDASTIC PROCESSES

Let $\mu(t)$ and $\sigma(t)$, $t \in [0, 1]$ be Lipschitz continuous functions and let ε_i, $i \in \mathbb{Z}$, be iid random variables with $E(|\varepsilon_i|^p) < \infty$ for some $p \geq 1$. Denote by F_ε and f_ε the distribution and density functions of ε_i, respectively. Recall $Z_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots)$. Consider the heteroscedastic model

$$X_i = \mu(i/n) + \sigma(i/n)e_i, \tag{2.14}$$

where $(e_i)_{i\in\mathbb{Z}}$ is a stationary process recursively defined by

$$e_i = m(e_{i-1}) + \varepsilon_i. \tag{2.15}$$

Here m is a Lipschitz continuous function satisfying $l_0 = \sup_{a\neq b} |m(a) - m(b)|/|a - b| < 1$. Then (e_i) has a unique stationary distribution and, by iterating (2.15), e_i can

be expressed as $e_i = g(Z_i)$ for some measurable function g. Recall Section 2.1 for Z_i^*. By theorem 2 in Wu and Shao (2004), $\|g(Z_i)\|_p < \infty$ and $\|g(Z_i) - g(Z_i^*)\|_p \leq l_0^i C$ for some $C > 0$. Let $G(t; Z_i) = \mu(t) + \sigma(t)g(Z_i)$. Then $E[\mathbf{1}_{G(t;Z_i)\leq x}|Z_{i-1}] = F_\varepsilon((x - \mu(t))/\sigma(t) - m(e_{i-1}))$ and, as in Section 2.1,

$$
\begin{aligned}
&\big\| \mathcal{P}_0 \mathbf{1}_{G(t;Z_i)\leq x} \big\| \\
&= \big\| \mathcal{P}_0 F_\varepsilon\big((x - \mu(t))/\sigma(t) - m(e_{i-1})\big) \big\| \\
&= \big\| E\big[F_\varepsilon\big((x - \mu(t))/\sigma(t) - m(e_{i-1})\big) - F_\varepsilon\big((x - \mu(t))/\sigma(t) - m(e_{i-1}^*)\big)|Z_0\big] \big\| \\
&\leq c_0 \big\| \min[2, m(e_{i-1}) - m(e_{i-1}^*)] \big\| \\
&\leq 2c_0 \|m(e_{i-1}) - m(e_{i-1}^*)\|_p^{p/2} = O\big(l_0^{ip/2}\big).
\end{aligned}
$$

Thus (2.8) is verified. To show (2.2), assume further that $\sup_{x\in\mathbb{R}} |x f_\varepsilon(x)| < \infty$ and $\min_{0\leq t\leq 1} \sigma(t) > 0$. Let F_e and f_e be the distribution and density functions of $e_i = g(Z_i)$, respectively. Because $E(|e_i|) < \infty$ and $E(|\varepsilon_i|) < \infty$, we have $E(|m(e_i)|) < \infty$ and

$$
\begin{aligned}
|x f_e(x)| &= |x| E f_\varepsilon(x - m(e_i)) \\
&\leq E\big\{ |x - m(e_i)| f_\varepsilon(x - m(e_i)) + |m(e_i)| f_\varepsilon(x - m(e_i)) \big\} < \infty.
\end{aligned}
$$

Note that $F(x; t) = F_e((x - \mu(t))/\sigma(t))$. By the Lipschitz continuity of $\mu(\cdot)$ and $\sigma(\cdot)$, elementary calculations show that (2.2) holds, and hence the process $(X_i)_{i=1}^n$ is also nearly stationary.

2.3   IMPLEMENTATION ISSUES

As the proof of Theorem 1 indicates, the term b_n in the bound of $\hat{q}_\alpha(t_0) - q_\alpha(t_0)$ in (2.10) is due to the error of the approximation of nearly stationarity by stationarity, whereas the other term $(nb_n)^{-1/2}$ can be loosely interpreted as the variability of sample quantiles to their theoretical counterparts. Relation (2.10) suggests the optimal bandwidth of the form

$$
b_n^{opt} = cn^{-1/3}, \tag{2.16}
$$

where c is a positive constant which may depend on the process $(X_i)_{i=1}^n$. Section 3 proposes a data-driven scheme for the choice of b_n.

Note that, as t changes from 0 to 1, $\hat{q}_\alpha(t, b_n) := \hat{q}_\alpha(X_{\lfloor n(t-b_n)\rfloor}, \ldots, X_{\lfloor n(t+b_n)\rfloor})$ may not be a continuous function of t, and thus an extra smoothing step is required. By using the Nadaraya–Watson method, we propose the smoothed estimator

$$
\tilde{q}_\alpha(t) = \frac{\sum_{i=1}^n K((t - i/n)/h_n)\hat{q}_\alpha(i/n, b_n)}{\sum_{i=1}^n K((t - i/n)/h_n)}, \tag{2.17}
$$

where $h_n \to 0$ is another sequence of bandwidths, and the kernel K is a nonnegative probability density function. We also assume that $nh_n \to \infty$ to ensure that the estimated curves are sufficiently smooth.

It is well known that the kernel estimator of a Lipschitz continuous function has the mean squared error (MSE) of order $O(n^{-2/3})$, and the optimal bandwidth is proportional to $n^{-1/3}$ (see, e.g., Gasser and Müller 1979). As the Monte Carlo simulation study in Section 4 shows, the smoothing step (2.17) does not significantly contribute to the MSE;

however, it can be used as a visual tool, as it provides nicer pictures. As another remark, the kernel method could also deal with potential boundary problems. A common approach is to use modified boundary kernels with asymmetric support. For examples of such kernels and a detailed discussion we refer to Müller (1991).

3. SELECTION OF SMOOTHING PARAMETERS

The choice of smoothing parameters (bandwidths) plays a crucial role in nonparametric inference. Commonly used criteria for bandwidth selection are based on cross-validation or minimization of mean squared errors. For time-dependent data, the leave-one-out principle used in cross-validation could alter the dependence structure, thus yielding unstable estimates. The other criterion relies on the analytical expression of the mean squared error of the nonparametric estimator, that is minimized with respect to the smoothing parameter under certain regularity conditions. For quantile estimates, the mean squared error can be expressed analytically in terms of the probability distribution function F. Because F is unknown, this criterion would involve an estimator of F, together with its bias and variance. Our purpose is to avoid the estimation of the whole probability distribution function, as it adds error and inflates computational time considerably (see next section for an illustration).

In this section we propose an automatic, data-driven scheme for choosing smoothing parameters in nonparametric quantile curve estimation. In practice, we cut up the observed time series into blocks, assuming that the time series in each block are approximately stationary, and carry out the two-step nonparametric procedure described in the previous section. Selection of the number of blocks is equivalent to selection of the bandwidth b_n in Theorem 1. Let $X_1, \ldots, X_n$ be the observed time series, assumed to be nearly stationary. Our algorithm involves splitting the series sequentially into (approximate) halves. The idea is to select the biggest block size that will enable us to estimate accurately enough the quantile evolution. The procedure is carried out in three steps:

1. Divide the series into blocks of approximate length 2^i, $i = p_0, \ldots, m - 1$, and compute the sample quantile in each of these blocks ($m \geq 1$ is such that $2^{m-1} < n \leq 2^m$). The choice of the initial p_0 depends on the frequency with which the data are sampled. A rule of thumb would be the minimum size for which the process seems to behave with some stationarity, and such that it is not too small to be able to compute quantiles. Moreover, p_0 may depend on the level α of the quantile to be estimated. Indeed, for α close to 0 or 1, it is intuitively clear that the starting minimal size should be larger (e.g., 2^{p_0} at least 16 for $0.1 \leq \alpha \leq 0.9$, and at least 32 for $\alpha > 0.9$ or $\alpha < 0.1$).

2. Compute the mean squared errors of the resulting estimates. These MSE_i's are computed as the sum of the squared differences between the sample quantiles and the sample quantiles in the smallest blocks. The intuition is that the smallest blocks, under the right assumptions, should give the most precise results. However, the variability is large with quantiles computed with small blocks, making it necessary to introduce the penalty in the next step.

3. Choose the block size that minimizes the criterion $(1 + \lambda 2^i)\mathrm{MSE}_i$ for the largest number of smoothing parameters λ, where the sequence of λ's is of length 10 and ranging from 0.01 to 0.1. The penalization for the number of parameters is aimed to avoid oversmoothing. In this way, a small reduction of the mean squared error with two times more parameters (i.e., with a block size divided by 2) does not imply a choice of a smaller block size. This penalization requires only the choice of a quantile level, and it is fast on both simulations and real data, as shown in the sequel.

A step further would be to extend the piecewise constant quantile estimator based on the above scheme to a piecewise linear estimator. The heuristic argument is that we use a first-order approximation of the true quantile curve locally. Even though this can be done easily in practice by locally using quantile regression (see next section), the theoretical derivations become cumbersome.

Another way to obtain quantile curve estimates is to use a rolling window approach, namely to consider at each time point the sample quantile over a moving window centered at the respective time point. Intuitively, the moving window approach would yield better estimates, because the nonstationarity could arise at irregular time points. However, the computing time, although still reasonable, is inflated. The problem then is how to choose the optimal window width. Theoretically this can be done by minimizing the mean squared error of the resulting quantile estimate, which is expressed analytically in terms of the probability distribution function. As mentioned previously, because we do not want to involve estimation of the whole probability distribution function, we need to get such a bandwidth automatically (directly from the data). The above scheme can be used in this setting as well. Thus the optimal bandwidth for the moving window approach can be computed by dividing the length of the time series by the optimal number of blocks. As explained in Section 2, we further smooth this piecewise constant quantile curve by using the kernel method; see (2.17).

To summarize, the automatic procedure for bandwidth selection described in this section can be used to produce piecewise constant (PC), piecewise linear (PL), and moving window (MW) quantile curve estimates, which can be further smoothed via the kernel method. The PC estimators are generated by taking the sample quantiles (constant) in each block. For the PL estimators, the linear quantiles are taken in each block instead of the sample quantiles. Linear quantiles are constructed via quantile regression (Koenker and Bassett 1978), a method that has been implemented in R (package *quantreg*, *http://www.econ.uiuc.edu/roger/research/rq/rq.html*). In contrast to these two approaches, where the blocks (windows) are fixed, the estimates given by (2.6) are obtained by taking the sample quantiles in moving windows, and therefore referred to as MW estimates. By using the kernel method [relation (2.17)], we get the smoothed versions of these estimates, PCS, PLS, and MWS, respectively.

4. MONTE CARLO SIMULATIONS

The following simulation study illustrates the behavior of the proposed quantile curve estimates for small and moderate sample sizes. We compare various quantile estimation
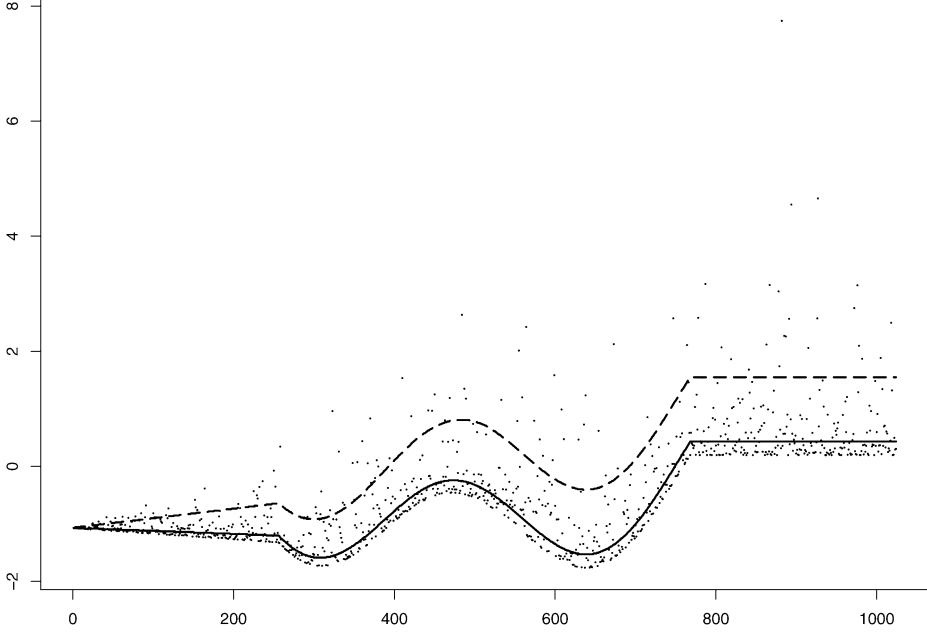
Figure 1. One realization, true median (solid line) and true 0.9 quantile (dashed line), for the simulated process (4.1), $n = 1024$.

methods, and show that the smoothing step improves the mean squared error. We simulated nonstationary non-Gaussian time series

$$X_t = \mu(t/n) + \sigma(t/n) \cdot (Z_t^2 - 1), \tag{4.1}$$

where Z_t is standard normal. We used $\mu(t/n) = \cos(7t/n) + \sin(17t/n)$ for $1/4 < t/n < 3/4$, and its linear extension by continuity on $[0, 1]$, $\mu(0) = -1.07$, $\mu(1) = 0.69$, and $\sigma(t/n) = \min(t/n, 0.5)$; see Figure 1. Process (4.1) satisfies conditions in Section 2.2.

Table 1 displays the integrated mean squared errors $\text{IMSE}(\tilde{q}_\alpha) = \int_0^1 \text{MSE}(\tilde{q}_\alpha(t)) \, dt$ for two choices of the quantile level $\alpha = 0.5$, 0.9, and for the six direct quantile estimators described at the end of Section 3. We generated 500 simulated processes (4.1) in each case.

Before discussing the performance of these estimators, we comment on the advantages of the direct approach versus a plug-in method [involving inverting an estimator of $F(x; t)$]. As mentioned in the Introduction, for processes that change smoothly with time, $F(x; t)$ can be estimated by using different weights for the indicator variables $\mathbf{1}_{X_i \leq x}$. Consider, for example, the Nadaraya–Watson estimate

$$\hat{F}(x; t) = \frac{\sum_{i=1}^n K((t - i/n)/d_n)\mathbf{1}_{X_i \leq x}}{\sum_{i=1}^n K((t - i/n)/d_n)}, \tag{4.2}$$

where K is a kernel, and d_n is a sequence of bandwidths such that $d_n \to 0$ and $nd_n \to 0$ as $n \to \infty$. Then construct the plug-in quantile estimator

$$\check{q}_\alpha(t) = \inf\{x : \hat{F}(x; t) \geq \alpha\}. \tag{4.3}$$

Table 1. Integrated mean squared errors (multiplied by 10^3 and rounded) of 0.5 and 0.9 quantile estimates for simulated nearly stationary time series (4.1): piecewise constant (PC), piecewise constant, smoothed (PCS), piecewise linear (PL), piecewise linear, smoothed (PLS), moving window (MW), moving window, smoothed (MWS). The moving window is applied for two kernels: rectangular (MWr, MWSr), and Gaussian (MWg, MWSg).

	$\alpha = 0.5$				$\alpha = 0.9$			
n	128	256	512	1024	128	256	512	1024
PC	163	153	148	146	335	269	248	229
PCS	137	132	128	127	282	221	203	187
PL	140	146	163	170	282	221	203	187
PLS	131	140	159	167	230	187	170	169
MWr	79	77	75	73	200	167	137	129
MWSr	81	77	75	73	185	161	136	129
MWg	84	75	75	73	210	155	137	127
MWSg	87	76	75	73	194	149	135	127

A major drawback of this approach is the fact that taking the above infimum adds numerical errors and increases computational time. We applied this method to the same 500 simulated processes for sample size $n = 128$. We used the truncated Gaussian kernel for K, and selected d_n by using the algorithm in Ghosh and Draghicescu (2002). The integrated mean squared errors of $\breve{q}_\alpha$ (multiplied by 10^3 and rounded, to be compared with the values in Table 1) were 206 for the median estimator and 470 for the 0.9 quantile estimator, respectively. The infimum was taken over 200 values of x, and the computational time for each simulation was about one minute (as opposed to the direct estimates, for which all the 500 runs took less than one minute). It is thus clear that the direct approach for estimating time-varying quantiles is not only significantly faster, but more accurate as well.

With regard to the behavior of the proposed direct quantile estimates, Table 1 shows that the moving window approach does much better than the two other techniques (piecewise constant and piecewise linear). Moreover, the piecewise constant estimate does better than the piecewise linear one for relatively large sample sizes and the median, because the piecewise linear estimate introduces the unnecessary complexity of linearly varying quantiles when the sample quantiles are nearly constant. Note that for high quantiles and the same sample sizes, the piecewise linear estimator does better in this example (and possibly in many real examples), as these quantiles tend to vary more than the median. However, the piecewise constant and piecewise linear estimates evenly divide the sample, whereas the moving window has the ability to adjust to complex nonstationarities (such as seasonality, or heteroscedasticity with an irregular frequency). Furthermore, it can be seen that smoothing improves the mean squared error in nearly all cases, with better results for the 0.9 quantiles. Indeed, there is more variation in the higher quantiles and therefore it is more difficult to find good estimates with small sample sizes. For the 0.5 and 0.9 quantile estimates based on the smoothed moving window with a rectangular kernel, when the sample sizes increase from 128 to 1024, the error goes from 81 to 73 and from 185 to 129, respectively. The other methods exhibit the same behavior. This shows that a larger

sample improves the estimation of higher quantiles much more than for the median. The best estimates are the ones produced by the moving window approach.

5. APPLICATIONS

There is a growing demand for new flexible and informative statistical tools (such as summaries and graphs) for the exploration of large datasets with complex structures. The methodology proposed in this article has a wide area of applicability in many fields, such as environmental science, atmospheric sciences, ecology, and epidemiology. For instance, in the study of climatic data, as pointed out by Vinnikov and Robock (2002), an important but difficult problem is to determine whether the climate is getting more or less variable. Vinnikov and Robock (2002) considered moments in residuals to test changes in variability. Here, the quantile curve estimation procedure previously described provides a very natural way to assess the variability in time series, by presenting detailed distributional information. Our method can be used as a graphical and visualization tool in time series analysis, having the advantages of being flexible, fast, accurate, and informative. Quantile estimates can also be further used in modeling spatial-temporal processes.

The next applications to ground-level and stratospheric ozone illustrate the potential of our approach. The smooth physical evolution of these processes makes nearly stationarity a natural working assumption. It is known that high ozone values have negative effects, and therefore modeling higher quantiles is very important. Moreover, our method can be used to describe ozone dispersion, by modeling measures of variability such as the interquartile range (IQR). The IQR is defined as the difference between the 0.75 quantile and the 0.25 quantile, and the IQR curve is defined as $IQR(t) = q_{0.75}(t) - q_{0.25}(t)$. In all examples below we use the moving window, smoothed approach, with a Gaussian kernel.

5.1 GROUND-LEVEL OZONE

In this application we consider the time series of daily maximum eight-hour averages of ground-level ozone mixing ratios at five monitoring sites in metropolitan Chicago (source: Illinois EPA). We used records of hourly averages of ozone in parts per billion (ppb). For each day we computed the 16 possible eight-hour averages (starting at 0:00, 1:00, ..., 15:00 for eight consecutive hours), and retained the maximum of these eight-hour averages. This measure is part of the National Air Quality Standard (NAAQS) for ground-level ozone issued by the United States Environmental Protection Agency (EPA). Typical questions of interest related to air pollution data concern understanding the behavior of high values, and also the space–time variability of the underlying process. Our method provides an informative exploratory tool, that may lead to building relevant statistical models, to capture the observed space–time dependencies. As an illustration, Figure 2 displays the 0.9 quantile curve estimates, and the IQR curve estimates for the aforementioned time series. We show the moving window, smoothed curve estimates for the period June 1–September 6, 1998. Even though Chicago had met the EPA ozone standards during this period, it is interesting to note the differences in these curves.
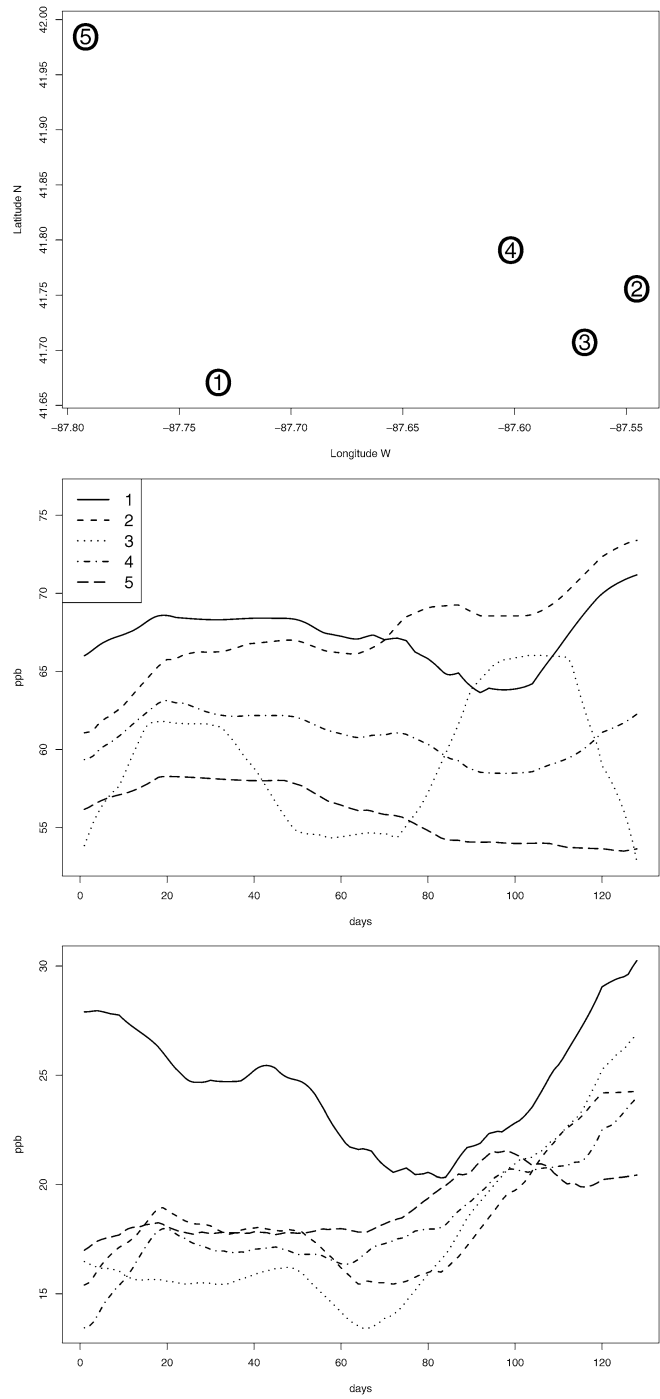
Figure 2.   Locations of five air pollution monitors in metropolitan Chicago (top); moving window, smoothed 0.9 quantile curve estimators for daily maximum eight-hour averages of ground-level ozone at these sites (middle); and moving window, smoothed estimators of interquartile range curves (bottom), period June 1–September 6, 1998.

For example, the monitoring sites 2 and 3 on the shore of Lake Michigan are 5.74 km apart, but display different patterns in the behavior of higher ozone (as measured by the 0.9 quantile), with a maximum difference of about 20 ppb, whereas the sites 4 and 5 are 26.67 km apart and show a similar pattern, with a shift of about 5 ppb in the 0.9 quantile curves. With regard to the ozone variability (measured by the IQR curves), whereas all sites show increased values in the second half of the period, site 1 in South East Chicago displays a completely different pattern in the first half of the period, with more than 10 ppb increase compared to the other monitors. These plots may indicate temporal nonstationarity and spatial-temporal interactions.

5.2 STRATOSPHERIC OZONE

An important environmental concern of the past three decades is stratospheric ozone depletion. The Total Ozone Mapping Spectrometer (TOMS) Nimbus 7 satellite (*http://toms.gsfc.nasa.gov/eptoms/ep.html*) provides daily total column ozone measurements on a 180 by 288 latitude by longitude grid. Figure 3 shows the observations overpassing a 1° by 1.25° latitude by longitude cell centered at 45.5°N, 0.625°W, covering the period July 17, 1990–May 5, 1993. This location was chosen for illustration purposes because midlatitude variations are the most relevant to study ozone recovery (Guillas et al. 2004). The estimated 0.1, 0.5, and 0.9 quantile curves are also displayed. This time period is of great interest because of the Mount Pinatubo eruption in June 1991 (Robock 2002).
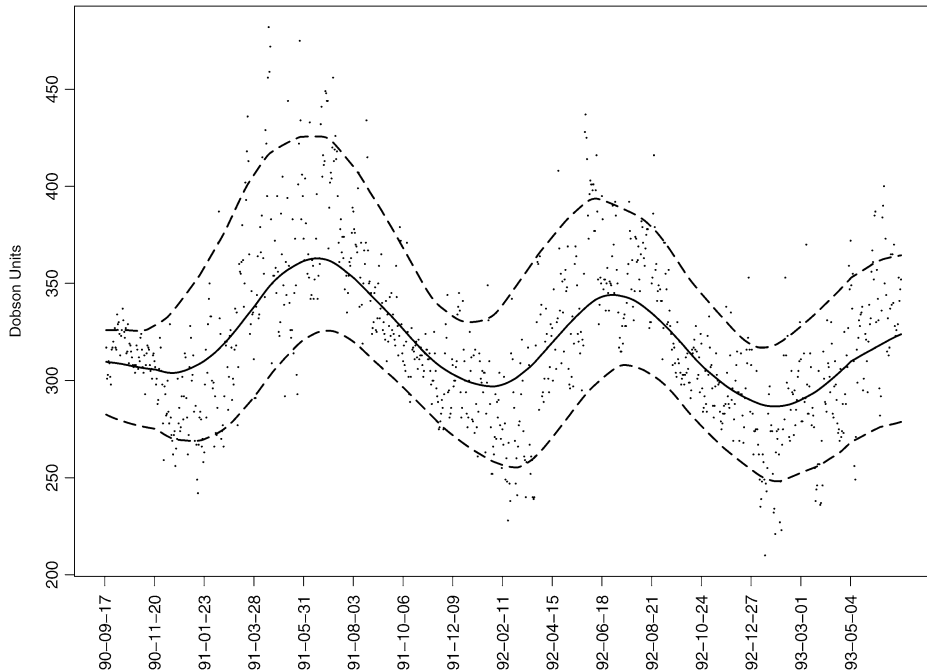

Figure 3.   Daily measurements of total column ozone at 45.5°N latitude, 0.625°W longitude. Total Ozone Mapping Spectrometer, Nimbus 7 Satellite, July 17, 1990–May 5, 1993: moving window, smoothed median (solid line), 0.1 and 0.9 quantile estimates (dashed lines).
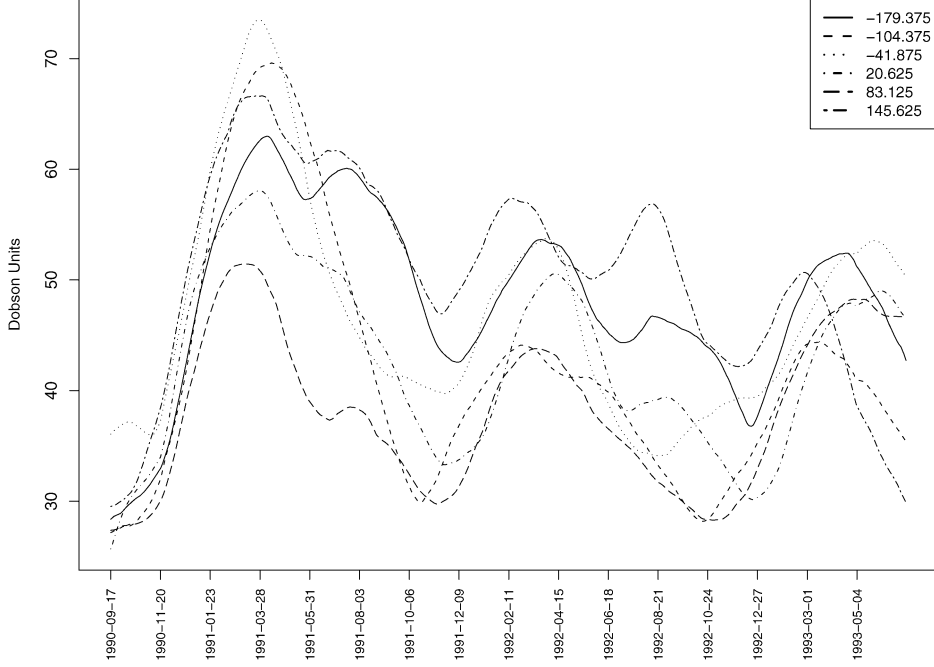
Figure 4.   Moving window, smoothed IQR curve estimates for latitude 45.5°N, longitudes ranging from −179.375° to 145.625°.

It appears that the annual variability is very well captured. Figure 4 displays the moving window, smoothed IQR curve estimates for six longitudes ranging from −179.375° to 145.625°, at latitude 45.5°N. Before the Mount Pinatubo eruption, starting in June 1991, the variability measured through the IQR at different longitudes has a large spread, with peak values in the 50–75 Dobson Units (DU) interval. After the eruption, the spread of the IQRs for all of these longitudes decreases and the IQRs are generally lower. This illustrates the overall dampening effect of Mount Pinatubo's eruption across the globe.

Instead of studying a particular location at a specific latitude, geophysicists focus on a zonal mean, which is usually an average over a latitude band. We can take advantage of all the data in a specific latitude band to get insight on the uncertainty in our quantile curve estimates. Thus, we compute 288 quantile curve estimates and rank them according to a functional criterion, the so-called functional depth introduced in Fraiman and Muniz (2001). Consider the univariate depth $D(x) = 1 - |1/2 - F(x)|$. If F is the cumulative distribution function of a real random variable and x is a data point, the depth measures the nearness to the median *med*, for which we have $D(med) = 1$. Consider functional data $Y(t), t \in [0, 1]$ and a sample $Y_1(t), \ldots, Y_p(t)$ having the same distribution as $Y(t)$. For each sample point x, let $F_{p,t}(x) = p^{-1} \sum_{j=1}^{p} \mathbf{1}_{Y_j(t) \leq x}$ and $D_{p,t}(x) = 1 - |1/2 - F_{p,t}(x)|$. As in Fraiman and Muniz (2001), we classify our curves according to the integrated index $I(Y) = \int_0^1 D_{p,t}(Y(t)) \, dt$. The index I globally measures the closeness to the empirical functional median, for which I attains its maximum. We estimate the median curves $\tilde{q}_{0.5}(t)$ for these $p = 288$ time series. In Figure 5 we display the functional median of these esti-
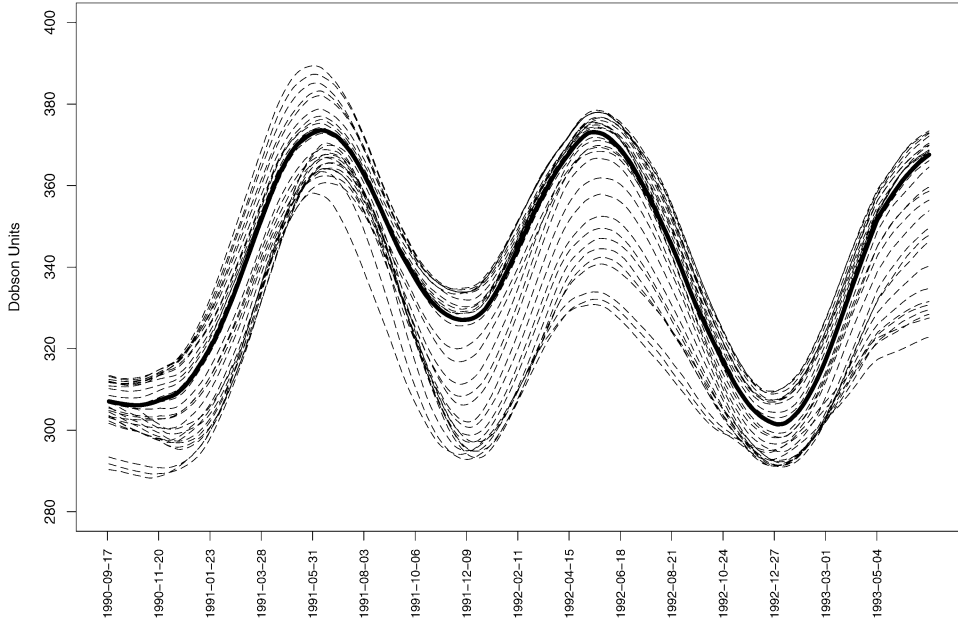
Figure 5. Spread of the median curve estimate for daily measurements of total column ozone at 45.5°N latitude and 288 locations ranging from 179.375°W to 179.375°E longitude, July 17, 1990–May 5, 1993. Solid line represents the functional median of the 288 curves. Dashed lines are the 10% least close to the median ordered curves according to the functional depth criterion.

mated median curves (bold line) together with the last 10% of the ordered median curves (dashed).

The functional median can be interpreted as the most representative pattern of the temporal changes in the daily total column ozone for this particular latitude band and time period. It can be also seen that for this particularly downward trend in total column ozone, at some longitudes total column ozone underwent a severe depletion after Mount Pinatubo's eruption, that is not shown in the most representative median curve. This means that ozone depletion occurred with drastically different impacts at different latitudes. Note that for changes associated with less localized events than volcanic eruption (such as halocarbons emissions), one may expect less variability across longitude. Our technique could be applied to detect other geographically dependent phenomena.

6. DISCUSSION

We introduced several quantile curve estimation methods for a very general class of stochastic processes. In particular we allow for nonstationarity, non-Gaussianity, and processes that may have infinite second moments. We assumed that the observed time series has a nearly stationary structure and developed a data-driven algorithm for the choice of the optimal block size. Our procedure proves to be fast in both simulations and applications. Another possibility for the choice of smoothing parameters would be to use resampling methods to estimate $\text{MSE}(\hat{q}_\alpha(t))$. For nearly stationary time series new meth-

ods need to be developed, for example in the spirit of Künsch (1989). Alternatively, one could focus on the choice of the optimal c in (2.16), for example by using the "second generation" methods discussed by Jones, Marron, and Sheather (1996), such as plug-in and smoothed bootstrap techniques. The performance of the smoothed estimators could be improved by considering time-varying bandwidths h_n or by using modified boundary kernels in (2.17). Future work will also include the extension of the concept of near stationarity to spatial and space–time processes. From the applied point of view, it would be interesting to examine the long-term trends of stratospheric ozone based on the direct median estimator.

ACKNOWLEDGMENTS

REFERENCES

Abberger, K., and Heiler, S. (2002), "Nonparametric Quantile Regression With Applications to Financial Time Series," in *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, Basel: Birkhäuser, pp. 149–160.

Adak, S. (1998), "Time-Dependent Spectral Analysis of Nonstationary Time Series," *Journal of the American Statistical Association*, 93, 1488–1501.

Bhattacharya, P., and Gangopadhyay, A. (1990), "Kernel and Nearest-Neighbor Estimation of Condition Quantiles," *Annals of Statistics*, 18, 1400–1412.

Borkar, V. S. (1993), "White-Noise Representations in Stochastic Realization Theory," *SIAM Journal on Control and Optimization*, 31, 1093–1102.

Cai, Z. (2002), "Regression Quantiles for Time Series," *Econometric Theory*, 18, 169–192.

Chaudhuri, P. (1991), "Nonparametric Estimates of Regression Quantiles and Their Bahadur Representations," *Annals of Statistics*, 19, 760–777.

Cole, T., and Green, P. (1992), "Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood," *Statistics in Medicine*, 11, 1305–1319.

Cox, D. (1983), "Asymptotic M-Type Smoothing Splines," *Annals of Statistics*, 11, 530–551.

Csörgö, M., and Yu, H. (1996), "Weak Approximations for Quantile Processes of Stationary Sequences," *The Canadian Journal of Statistics*, 24 (4), 403–430.

Dahlhaus, R. (1997), "Fitting Time Series Models to Nonstationary Processes," *Annals of Statistics*, 25 (1), 1–37.

Degenhardt, H. J. A., Puri, M. L., Sun, S., and van Zuijlen, M. C. A. (1996), "Characterization of Weak Convergence for Smoothed Empirical and Quantile Processes Under ϕ-Mixing," *Journal of Statistical Planning and Inference*, 53 (3), 285–295.

Draghicescu, D., and Ghosh, S. (2003), "Smooth Nonparametric Quantiles," in *Proceedings of MENP-2*, Bucharest: Geometry Balkan Press, pp. 45–52.

Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer.

Fotopoulos, S. B., Ahn, S. K., and Cho, S. (1994), "Strong Approximation of the Quantile Processes and Its Applications Under Strong Mixing Properties," *Journal of Multivariate Analysis*, 51, 17–45.

Fraiman, R., and Muniz, G. (2001), "Trimmed Means for Functional Data," *Test*, 10 (2), 419–440.

Gasser, T., and Müller, H. (1979), "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation. Lecture Notes in Mathematics*, Vol. 757, New York: Springer, pp. 23–68.

Ghosh, S., and Draghicescu, D. (2002), An Algorithm for Optimal Bandwidth Selection for Smooth Nonparametric Quantile Estimation," in *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, Basel: Birkhäuser, pp. 161–168.

Ghosh, S., Beran, J., and Innes, J. (1997), "Nonparametric Conditional Quantile Estimation in the Presence of Long Memory," *Student*, 2, 109–117.

Giurcanu, M., and Spokoiny, V. (2002), "Confidence Estimation of the Covariance Function of Stationary and Locally Stationary Processes," preprint. Available at *http://www.stat.ufl.edu/~giurcanu/document.pdf*.

Guillas, S., Stein, M. L., Wuebbles, D. J., and Xia, J. (2004), "Using Chemistry-Transport Modeling in Statistical Analysis of Stratospheric Ozone Trends From Observations," *Journal of Geophysical Research*, 109, D22303, doi:10.1029/2004JD005049.

Hall, P., and Hart, J. D. (1990), "Nonparametric Regression With Long-Range Dependence," *Stochastic Processes and Their Applications*, 36 (2), 339–351.

Huang, M. L., and Brill, P. (1999), "A Level Crossing Quantile Estimation Method," *Statistics & Probability Letters*, 45, 111–119.

Johnstone, I. M., and Silverman, B. W. (1997), "Wavelet Threshold Estimators for Data With Correlated Noise," *Journal of the Royal Statistical Society*, Ser. B, 59 (2), 319–351.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996), "A Brief Survey of Bandwidth Selection for Density Estimation," *Journal of the American Statistical Association*, 91, 401–407.

Kallianpur, G. (1981), "Some Ramifications of Wieners Ideas on Nonlinear Prediction," in *Norbert Wiener*, Vol. III, ed. P. Masini, Cambridge, MA: MIT Press, pp. 402–424.

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

Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and Portnoy, S. (1987), "*L*-Estimation for the Linear Model," *Journal of the American Statistical Association*, 82, 851–857.

Koenker, R., Pin, N., and Portnoy, S. (1994), "Quantile Smoothing Splines," *Biometrika*, 81, 673–680.

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

Mehra, K. L., Krishnaiah, Y. S. R., and Rao, M. S. (1992), "Weak Convergence of Conditional Empirical Processes and Conditional Quantile Estimators for Weakly Dependent Sequences," in *Information Theory, Statistical Decision Functions, Random Processes: Transactions of the 11th Prague Conference, Prague, 1990*, pp. 201–226.

Mukherjee, K. (1999), "Asymptotics of Quantiles and Rank Scores in Nonlinear Time Series," *Journal of Time Series Analysis*, 20, 173–192.

Müller, H. G. (1991), "Smooth Optimum Kernel Estimators Near Endpoints," *Biometrika*, 78, 521–530.

Nadaraya, E. (1964), "On Estimating Regression," *Theory of Probability and Its Applications*, 9, 141–142.

Olsson, J., and Rootzen, H. (1996), "Quantile Estimation From Repeated Measurements," *Journal of the American Statistical Association*, 91, 1560–1565.

Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002), "The SLEX Model of Non-Stationary Random Process," *Annals of the Institute of Statistical Mathematics*, 54, 171–200.

Priestley, M. B. (1965), "Evolutionary Spectra and Nonstationary Processes," *Journal of the Royal Statistical Society*, Ser. B, 27, 204–237.

———— (1988), *Nonlinear and Nonstationary Time Series Analysis*, London: Academic Press.

Priestley, M. B., and Chao, M. T. (1972), "Nonparametric Function Fitting," *Journal of the Royal Statistical Society*, Ser. B, 34, 385–392.

Robock, A. (2002), "Pinatubo Eruption: The Climatic Aftermath," *Science*, 295, 1242–1244.

Rosenblatt, M. (1956), "Remarks on Some Non-Parametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 642–669.

Samorodnitsky, G., and Taqqu, M. (1994), *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*, New York: Chapman & Hall.

Sheather, S. J., and Marron, J. S. (1990), "Kernel Quantile Estimators," *Journal of the American Statistical Association*, 85, 410–416.

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes With Applications to Statistics*, New York: Wiley.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press.

Vinnikov, K. Y., and Robock, A. (2002), "Trends in Moments of Climatic Indices," *Geophysical Research Letters*, 29 (2), doi:10.1029/2001GL014025.

Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman & Hall.

Watson, G. (1964), "Smooth Regression Analysis," *Sankhyā*, Ser. A, 26, 359–372.

White, H. (1990), "Nonparametric Estimation of Conditional Quantiles Using Neural Networks," in *Proceedings of Computing Science and Statistics, 8th Annual Symposium of the Interface*, 22, New York: Springer, pp. 190–199.

Wiener, N. (1958), *Nonlinear Problems in Random Theory*, Cambridge, MA: MIT Press.

Wu, W. B. (2005a), "Nonlinear System Theory: Another Look at Dependence," *Proceedings of the National Academy of Science, USA*, 102, 14150–14154.

——— (2005b), "On the Bahadur Representation of Sample Quantiles for Dependent Sequences," *Annals of Statistics*, 33, 1934–1963.

Wu, W. B., and Shao, X. (2004), "Limit Theorems for Iterated Random Functions," *Journal of Applied Probability*, 41, 425–436.

Yang, S. (1981), "Linear Functions of Concomitants of Order Statistics With Applications to Nonparametric Estimation of Regression Functions," *Journal of the American Statistical Association*, 76, 658–662.

Yu, K., and Jones, M. C. (1998), "Local Linear Quantile Regression," *Journal of the American Statistical Association*, 93, 228–237.

Yu, K., Lu, Z., and Stander, J. (2003), "Quantile Regression: Applications and Current Research Areas," *The Statistician*, 52 (3), 331–350.