The University of Chicago
Center for Integrating Statistical and Environmental Science
www.stat.uchicago.edu/~cises

Chicago, Illinois USA

TECHNICAL REPORT NO. 2

OPTIMAL BANDWIDTH SELECTION FOR KERNEL ESTIMATORS OF TIME-VARYING QUANTILES*

D. Draghicescu
January 2003

*Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through STAR Cooperative Agreement #R-82940201-0 to The University of Chicago, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

This research is funded by
U.S. EPA - Science To Achieve Results (STAR) Program
Grant # R-82940201-0

Optimal bandwidth selection for kernel estimators of time-varying quantiles

Dana Draghicescu
University of Chicago
Center for Integrating Statistical and Environmental Science
5734 South Ellis Avenue, 60637 Chicago, IL, USA

Keywords: time series, quantiles, kernel smoothing, bandwidth selection

Mathematical Subject Classification: 62M10, 62G20

Abstract: In this paper we address the problem of optimal bandwidth selection for kernel estimators of probability distribution functions leading to quantile estimators for transformations of Gaussian processes with short and long memory. Specifically, locally optimal bandwidths are obtained by minimizing the asymptotic mean squared error of the initial kernel estimator. Two data-driven procedures for optimal bandwidth selection are discussed. The first procedure is very general and utilizes a nonparametric approach, whereas the second is semiparametric, being based on the long-memory asymptotic assumption. These two methods are compared by means of a simulation study. Applications to real data are also included.

1 Introduction

The problem of quantile estimation for dependent data has received much attention in the past years due to its practical importance. In particular, extreme quantiles are of interest for many environmental and economic processes. The natural approach is to generalize existing quantile estimation methods for independent identically distributed data. One way of getting quantile estimates is based on estimation of the probability distribution function of the underlying process. In this paper we use this approach. An alternative is given by quantile regression, method introduced in Koenker and Bassett (1978). More recently quantile regression was extended for dependent data (see for example Abberger and Heiler (2002), Cai (2002) and the references therein). In these both approaches parametric or nonparametric models can be used. In many practical applications nonparametric models are preferred due to their flexibility and smoothing techniques are used for the estimation of the quantities of interest. For example, an overview of kernel smoothing for quantile estimation is given in Sheather and Marron (1990). Another problem is then how to characterize the dependence of the data. A widely used approach is to consider stationary Gaussian time series. While much is understood about the behavior of such time series, many data sets are neither stationary nor Gaussian, therefore there is a need for more general approaches. As pointed out in Beran (1994), two natural ways to generalize Gaussian time series are given by linear processes and transformation of Gaussian processes. In this paper we focus on the latter case. Thus in Section 2 we introduce a very general model and give relevant examples. In Section 3 kernel smoothing in the time domain is used to estimate time-varying probability distribution functions leading to plug-in quantile estimators. Asymptotic properties of these estimators are discussed for time-dependent transformations of Gaussian time series with short and long memory. Optimal selection of the smoothing parameters is discussed in detail in Section 4, where a data-driven procedure for consistent estimation of these optimal bandwidths is analyzed. An alternative method is also given in a particular case (linear transformations of long-memory Gaussian time series). These two methods are compared by means of the simulation study in Section 5. Section 6 shows applications of the proposed methodology to time series with short and long memory.

2 The model

Consider discrete-parameter stochastic processes of the form

$$Y_i := Y(x_i) = G(Z_i, x_i), \quad i = 1, 2, \dots \tag{1}$$

defined on rescaled time points $x_i = \frac{i}{n}$. We assume that $\{Z_i\}_{i\geq 1}$ is a standardized stationary Gaussian process with covariance function $\gamma_Z(l) := cov(Z_i, Z_{i+l})$, $i \geq 1$, $l = 0, \pm 1, \pm 2, \dots$. The unknown function $G : \mathbf{R} \times [0,1] \to \mathbf{R}$ in (1) is assumed Lebesgue measurable and such that

$$1_{\{G(Z_i,x_i)\leq y\}} - F_{x_i}(y) \;=\; \sum_{k=m}^{\infty} c_k(x_i, y) H_k(Z_i), \tag{2}$$

where $H_k(\cdot)$ is the Hermite polynomial of degree k and the coefficients $c_k(x, y)$

are assumed twice continuously differentiable with respect to x and continuous with respect to y. Note that the probability distribution of the process $\{Y_i\}$ may change with time, so in this sense the process is not stationary. The notion of Hermite rank (the integer $m \geq 1$ in (2)) was introduced in Taqqu (1975) (see example 1 below), where it was proved that the behavior of the sum of transformations of Gaussian processes with long memory depends essentially on the first nonzero term in the Hermite expansion. The model (1) was introduced in Ghosh et al. (1997), generalizing the model proposed in Taqqu (1975) by letting the unknown function G vary with time. It is a very general model, including nonstationary and nongaussian processes. As a particular case, the classic nonparametric regression model falls within this frame (see example 2 below). We consider two types of covariance structures for the underlying Gaussian process $\{Z_i\}$ in (1).

The short-memory case when the covariances decay fast so that

$$\sum_{l=-\infty}^{\infty} \gamma_Z(l) < \infty.$$

In this case we assume further that

$$\sum_{l=-\infty}^{\infty} |\gamma_Z(l)|^m < \infty. \tag{3}$$

The long-memory case when the covariances decay slower and

$$\sum_{l=-\infty}^{\infty} \gamma_Z(l) = \infty.$$

As showed in Beran (1994) page 42, this slow decrease can be expressed by an asymptotic assumption. Thus we assume that

$$\gamma_Z(l) \sim C_Z |l|^{m(2d-1)} \;\; as \; |l| \to \infty, \;\; 0 < d < \frac{1}{2}, \tag{4}$$

where C_Z is a slowly varying function at infinity (that asymptotically can be approximated by a positive constant). In the frequency domain expression (4) is equivalent to $f_Z(\omega) \sim D_Z |\omega|^{-2md}$ as $|\omega| \to 0$, where $f(\omega) = (2\pi)^{-1} \sum_{l=-\infty}^{\infty} e^{il\omega} \gamma_Z(l)$ is the spectral density of the process $\{Z_i\}$ and $D_Z = \frac{1+2d}{sin()\Gamma(1-2d)} C_Z$.

Example 1 The Taqqu model:

$$Y_i = G(Z_i), \quad i = 1, 2, \ldots. \tag{5}$$

In this case the process $\{Y_i\}$ is stationary and the coefficients c_k in the Hermite expansion (2) do not depend on x. The limiting distribution of sum of transformations of type (5) suitably normalized was obtained by Taqqu (1975) for the case when the underlying Gaussian process $\{Z_i\}$ has long memory. It was proved that this limiting distribution is standard normal only for Hermite rank $m = 1$. When the underlying Gaussian process has short memory, this limiting distribution is standard normal for any $m \geq 1$ provided (3) holds (as proved in Breuer and Major (1983)).

Example 2 Well known and widely used in the past decades are processes of the more general form

$$Y_i = g_1(x_i) + g_2(Z_i), \quad i = 1, 2, \ldots. \tag{6}$$

If g_1 is such that $g_1(x_i) - g_1(x_j)$ depends only on $x_i - x_j$ for any i, j (for example if g_1 is linear), then we are still in the stationary case. In addition, if g_2 is taken to be the identity function and if the Z_i's are assumed independent identically distributed, this model becomes the ordinary nonparametric regression model.

Example 3 A more general example consists in processes of the form

$$Y_i = \mu(x_i) + \sigma(x_i)Z_i, \quad i = 1, 2, \ldots. \tag{7}$$

For such processes the mean and the variance vary with time but the process is still Gaussian, which is a restriction that is overcome by the general model (1). In addition, after trend removal, processes of the form (7) can be reduced to processes of the form (6). The methodology we use in this paper, being done for the general case, does not imply trend removal, thus eliminating an extra error source.

3 Estimation

The problem of interest is to estimate the marginal probability distribution $F_x(y) = P(Y(x) \leq y|x)$ for any $y \in \mathbf{R}$ so that we can obtain the quantile function $Q_\alpha(x) = inf\{y \in \mathbf{R} | F_x(y) \geq \alpha\}$ for any $0 < \alpha < 1$. Because of the assumption that the process $\{Y_i\}$ may change with time, the empirical distribution function would be too rough an estimator for our goal. To overcome this inconvenience we use the convolution of this estimator with a smooth function (see equation (12) below). For any $y \in \mathbf{R}$ let

$$I(x_i, y) = \begin{cases} 1 & \text{if } Y_i \leq y, \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

be the $0-1$ time series based on the observations $\{Y_1, \ldots, Y_n\}$. Clearly $Var\big(I(x,y)\big)$ is finite and, based on the Hermite expansion (2) it can be expressed as

$$\sigma(x,y) := Var\big(I(x,y)\big) = \sum_{k=m}^{\infty} k! c_k^2(x,y). \tag{9}$$

We assume that the changes with time of the process $I(x,y)$ are smooth, so that

$$\frac{\partial^2}{\partial x^2}\big[\sigma(x,y)\big] < \infty, \ \forall x \in [0,1], \ \forall y \in \mathbf{R} \tag{10}$$

and

$$\sum_{k=m}^{\infty} \big[c_k^{(l)}(x,y)\big]^2 k! < \infty, \ l = 1, 2, \quad \forall x \in [0,1], \quad \forall y \in \mathbf{R}, \tag{11}$$

where $c_k^{(l)}(x,y) := \frac{\partial^l}{\partial x^l}\big[c_k(x,y)\big]$. We use the kernel estimator of $F_x(y)$ considered in Ghosh et al. (1997):

$$\hat{F}_x(y) = \frac{1}{nb}\sum_{i=1}^{n} K\Big(\frac{x_i - x}{b}\Big) \cdot I(x_i, y), \quad x \in [0,1], \ y \in \mathbf{R}, \tag{12}$$

where the the kernel K is a symmetric non-negative Lipschitz continuous density function with support $[-1,1]$ that integrates to 1 and $b = b(x,n)$ is a sequence of bandwidths such that, as $n \to \infty$, $b \to 0$ and $nb^3 \to \infty$. Then for any $0 < \alpha < 1$ the quantile estimators are obtained as

$$\hat{Q}_\alpha(x) = inf\{y \in \mathbf{R} \big| \hat{F}_x(y) \geq \alpha\}, \quad x \in [0,1]. \tag{13}$$

The plug-in estimator $\hat{Q}_\alpha$ will inherit all the properties of $\hat{F}_x$ and therefore in the remaining part of this section we will focus on the latter estimator (given by (12)). Ghosh et al. (1997) proved that in the long-memory case, under certain regularity conditions on the true probability distribution function F_x,

$$MSE\Big(\hat{F}_x(y)\Big) = O\Big(max\big(b^4, (nb)^{m(2d-1)}\big)\Big). \tag{13}$$

The following result holds in the short-memory case.

Theorem 1 In the above notations, if $\frac{\partial^2}{\partial x^2}\Big[F_x(y)\Big]$ exist a.e. in $[0,1]$ for any $y \in \mathbf{R}$, if (3), (10) and (11) hold, then for the estimator (12), for any $x \in [0,1]$ and any $y \in \mathbf{R}$, as $n \to \infty,\ b \to 0$ and $nb^3 \to \infty$, we have

$$MSE\Big(\hat{F}_x(y)\Big) = O\Big(max\big(b^4, (nb)^{-1}\big)\Big). \tag{14}$$

Proof of Theorem 1 By standard arguments we get that

$$Bias\Big(\hat{F}_x(y)\Big) = B_n(x,y;b) + O\Big(\frac{1}{nb}\Big), \tag{15}$$

where

$$B_n(x,y;b) = \frac{b^2}{2}\Big[\int_{-1}^{1} u^2 K(u)du\Big] \cdot \frac{\partial^2}{\partial x^2}\Big[F_x(y)\Big]. \tag{16}$$

Note that because of the assumption $nb^3 \to \infty$, the dominant term in (15) is $B_n(x,y;b)$. For the variance we have

$$Var\big(\hat{F}_x(y)\big) = \frac{1}{(nb)^2}\sum_{i,j=1}^{n} K\Big(\frac{x_i - x}{b}\Big)K\Big(\frac{x_j - x}{b}\Big) \cdot cov\big(I(x_i,y), I(x_j,y)\big).$$

By using the Hermite expansion (2), properties of Hermite polynomials and Taylor expansions of $c_k(x_i,y), c_k(x_j,y)$ around x we get that

$$Var\Big(\hat{F}_x(y)\Big) = V_n(x,y;b) + o\big((V_n(x,y;b)\big), \tag{17}$$

where

$$V_n(x,y;b) = \frac{1}{(nb)^2}\sum_{i,j=1}^{n} K\Big(\frac{x_i - x}{b}\Big)K\Big(\frac{x_j - x}{b}\Big) \cdot \sum_{k=m}^{\infty} k! c_k^2(x,y)\gamma_Z^k(|i-j|).$$

(18).

The kernel K is smooth and symmetric, so that $K'(0) = 0$ and by using its Taylor expansion around the origin we get

$$V_n(x, y; b) \sim \frac{K^2(0)\sigma(x, y)}{(nb)^2} \sum_{k=1}^{\infty} \sum_{i,j=n(x-b)}^{n(x+b)} \gamma_Z^k(i-j).$$

By using a change of indices we obtain

$$V_n(x, y; b) \sim \frac{K^2(0)\sigma(x, y)}{(nb)^2} \sum_{k=1}^{\infty} \sum_{l=-nb}^{nb} \big(2nb + 1 - |l|\big) k! \gamma_Z^k(l).$$

The short-memory assumption implies that $\gamma_Z^k(l) \leq \gamma_Z^m(l),\ \forall l,\ \forall k \geq m$. Therefore the first term in the above expansion (which is the dominant term) is bounded by $\frac{K^2(0)\sigma(x,y)}{nb} \cdot \sum_{l=-nb}^{nb} \gamma_Z^m(l)$. The theorem is completely proved by taking into account assumption (3).

Remark 1 In the short-memory case $Var\Big(\hat{F}_x(y)\Big)$ has the same rate of convergence as in the independent case, namely $(nb)^{-1}$ (see for example Simonoff (1996) Chapter 3), whereas in the long-memory case it converges at a slower rate depending on the long-memory parameter d and on the Hermite rank m.

Remark 2 The consistency of the quantile estimator (13) follows immediately from the consistency of $\hat{F}_x(y)$ that is implied by theorem 1.

4 Optimal bandwidth selection

It is well known that for the choice of the optimal bandwidths for kernel curve estimators, two main methods can be used: plug-in and cross-validation (for an overview see for example Hart (1997), Chapter 4). For dependent data the "leave-one-out" principle can lead to inaccurate conclusions by not taking into account correctly the correlation structure of the data, therefore we choose a plug-in method. Remark in (12) that the smoothing parameter b may depend on x but it should not depend on the cut-off value y. This is because we want the estimator $\hat{F}_x$ to preserve the monotonicity property of the true probability distribution function F_x. Remark also that the assumptions on the kernel ensure that $\hat{F}_x(y) \in [0, 1]$. We consider a global measure of accuracy of $\hat{F}_x(y)$ (with respect to y) and choose the optimal smooothing parameter as

$$b_{opt}(x) = \underset{b>0}{argmin} \int_{\mathbf{R}} \Big[MSE\Big(\hat{F}_x(y); b\Big)\Big] dy. \tag{20}$$

Remark 3 In the short-memory case $b_{opt}(x) = O\Big(n^{-\frac{1}{5}}\Big)$ (the proof follows immediately from theorem 1 and definition (20)), same result as in the independent case, whereas in the long-memory case $b_{opt}(x) = O\Big(n^{max\left(-\frac{1}{5}, m(2d-1)-1\right)}\Big)$ as proved in Ghosh et al. (1997), Theorem 1.
In practice, in order to determine the locally optimal smoothing parameters based on (20), we need to estimate the bias and variance of $\hat{F}_x(y)$ so as to produce an estimator of $MSE\Big(\hat{F}_x(y); b\Big)$. Based on (16) we can obtain a plug-in estimator of $Bias\Big(\hat{F}_x(y)\Big)$ based on an estimator of the second derivative $F_2 := \frac{\partial^2}{\partial x^2}\Big[F_x(y)\Big]$ (assuming F sufficiently smooth). Such an estimator was introduced in Ghosh and Draghicescu (2002a) leading to the consistent estimator of $Bias\Big(\hat{F}_x(y)\Big)$

$$\hat{B}_n(x, y; b) = \frac{b^2}{2}\Big[\int_{-1}^{1} u^2 K(u) du\Big] \cdot \frac{1}{nb_1^3} \sum_{i=1}^{n} K_2\Big(\frac{x_i - x}{b_1}\Big) \cdot I(x_i, y), \tag{21}$$

where K_2 is a kernel function possibly different from K. In fact less assumptions are needed for K_2 because the target F_2 does not need to satisfy any boundary and monotonicity conditions. The new smoothing parameter should be such that $b_1 = b_1(b) \geq b$, as in order to obtain the derivative of a function at a given point a sequence of neighboring observations has to be taken into account. Also $b_1 \to 0$ as $n \to \infty$. A similar procedure was used for the estimation of derivatives of the mean function in Brockmann et al. (1993) in the nonparametric regression model with independent identically distributed errors.
Estimation of $Var\Big(\hat{F}_x(y)\Big)$ is a more difficult problem because its analytic expression depends on many unknown parameters that are hard if not impossible to estimate (as for example the Hermite rank m, $\sum_{l=-\infty}^{\infty} |\gamma_Z(l)|^m$ in the short-memory case, C_Z and the dominant coefficient $c_m(x, y)$ in the long-memory case). This inconvenience can be overcome by noting that the

dominant term in $Var\Big(\hat{F}_x(y)\Big)$ is a weighted sum of quantities that depend only on the time lag (see (18)), therefore it seems natural to estimate these quantities by weighting the sample covariances. Thus let us consider the estimator of $Var\Big(\hat{F}_x(y)\Big)$ given by

$$\hat{V}_n(x,y;b) = \frac{1}{(nb)^2}\sum_{i,j=1}^{n} K\Big(\frac{x_i - x}{b}\Big)K\Big(\frac{x_j - x}{b}\Big)\cdot$$

$$\cdot\frac{1}{n_2 - n_1 + 1}\sum_{k=1}^{n_2-n_1+1-l}\big[I(x_k,y) - I(y)\big]\big[I(x_{k+l}, y - I(y)\big], \qquad (22)$$

where $n_1 = [n(x - b_2)] + 1$, $n_2 = min([n(x + b_2)], n)$, $b_2 = b_2(b) \geq b$, $I(y)$ is the sample mean of $I(x_i, y)$ in the respective windows in (22) and $[x]$ is the integer part of x. Remark that to estimate the local covariances we use in fact the box kernel estimator with bandwidth $b_2 = b_2(b)$ such that $b_2 \to 0$ as $n \to \infty$. This new bandwidth should be such that $b_2 \geq b$, otherwise less observations are used in each window and the correlation structure of the data may be altered. We can now define the estimator of $b_{opt}(x)$

$$\hat{b}_{opt}(x) = \underset{b>0}{argmin}\int_{\mathbf{R}}\Big[\hat{B}_n^2(x,y;b) + \hat{V}_n(x,y;b)\Big]dy. \qquad (23)$$

Theorem 2 In the notations and assumptions of theorem 1, if in addition $\frac{\partial^4}{\partial x^4}\Big[F_x(y)\Big]$ exist a.e. in $[0,1]$ for any $y \in \mathbf{R}$, if K_2 is a Lipschitz continuous function with support $[-1,1]$ and such that $\int_{-1}^{1} u^j K_2(u)du = 0$ for $j = 1,2,3$ and $\int_{-1}^{1} u^4 K_2(u)du \neq 0$, then for any $x \in [0,1]$, $\hat{b}_{opt}(x) \longrightarrow_p b_{opt}(x)$.

Proof of Theorem 2 $\hat{B}_n(x,y;b)$ is a consistent estimator of $B_n(x,y;b)$ and the proof follows along the lines of the proof of theorem 1 by noting that the estimator of $\hat{F}_2$ is similar in form with the estimator $\hat{F}$. Then, by standard arguments (see for example Priestley (1981) section 5.3.3) it can be showed that

$$Bias\big(\hat{V}_n(x,y;b)\big) = \frac{2\pi\sigma(x,y)}{n} + O\big((nb)^{-2}\big)$$

and

$$Var\big(\hat{V}_n(x,y;b)\big) = \frac{2\pi}{n}\int_{-1}^{1}\int_{-1}^{1} K(u)K(v)\Big[\int_{-\pi}^{\pi}\big(1+e^{2i\omega|u-v|}\big)h^2(\omega)d\omega\Big]dudv+O\big((nb)^{-2}\big),$$

where $h(\omega) = \sigma(x,y)f(\omega)$. The result follows immediately by using the continuous mapping theorem and Chebysheff's inequality.

Remark 4 $\hat{b}_{opt}(x)$ is consistent for any positive b_1, b_2 (that appear in (21) and (22) respectively). In practice these additional bandwidths can be chosen empirically. For example Ghosh and Draghicescu (2002b) proposed an algorithm for obtaining $\hat{b}_{opt}(x)$ where the above b_1, b_2 are updated at every step of the iterative procedure. We use this procedure for the simulation study in Section 5 and for the applications in Section 6.

Remark 5 An advantage of using local bandwidths is that boundary problems are avoided. Thus no modified kernels are needed for the end points of the time series - for an illustration see Figure 4.

Remark 6 In the long-memory case, by using the analytic expression of $Var\Big(\hat{F}_x(y)\Big)$ (see Ghosh et al. (1997) Theorem 1) we can construct the plug-in estimator

$$\hat{V}_n(x,y;b) = C_Z^m c_m^2(x,y)(nb)^{m(2\hat{d}-1)}m!^2\int_{-1}^{1}\int_{-1}^{1} K(u)K(v)|u-v|^{m(2\hat{d}-1)}dudv \tag{24}$$

based on the maximum likelihood estimator for the long-memory parameter d (for details see Beran (1994) Chapter 5) and obtain consistent estimators of the locally optimal bandwidths by plugging in the above $\hat{V}_n(x,y;b)$ in (23).

5 Simulation study

As mentioned in the last remark, for long-memory processes the variance of $\hat{F}_x(y)$ can be estimated by plugging in the maximum likelihood estimator of the long-memory parameter in the analytic formula of $Var\Big(\hat{F}_x(y)\Big)$ - see (24) - assuming m, C_Z and c_m are known. Consistent estimators of the locally

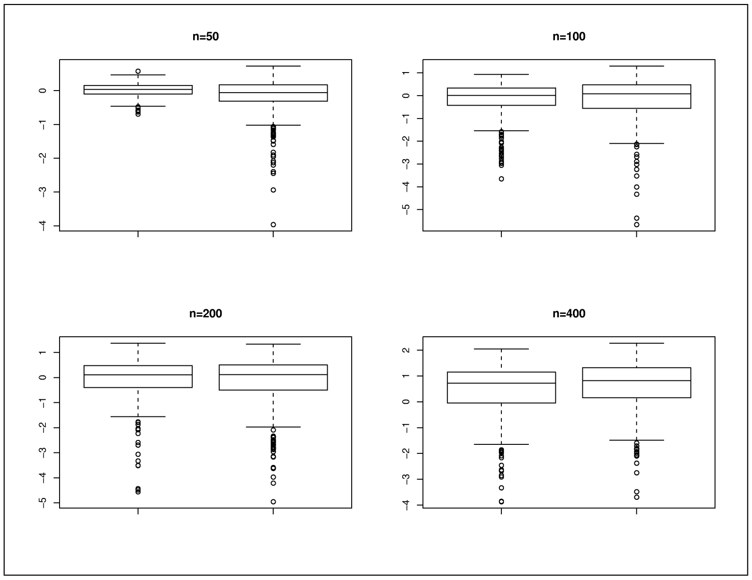

Figure 1: Boxplots of $ln\left(\frac{\hat{b}_{opt}}{b_{opt}}\right)$ computed by using the general method (left) and the long-memory method (right) for $d = 0.1$

optimal bandwidths can then be computed based on this variance estimator and the bias estimator (21). The following simulation study was carried out to compare these two methods. They will be referred to as the general method and the long-memory method.

We simulated Gaussian time series with long memory by taking $m = 1$, $C_Z = 1$ in (4) leading to $c_1(x, y) = 1$ and two values for the long-memory parameter: $d = 0.1$ and $d = 0.4$ respectively. We computed first the true global optimal bandwidths $b_{opt} = \frac{1}{n}\sum_{i=1}^{n} b_{opt}(x_i)$ based on (20) by running 500 simulations at a time. In the next step we computed the estimated global optimal bandwidths based on these two methods, again running 500 simulations for each value of the sample size n and for each value of the long-memory parameter d. The bias, variance and mean squared error of these estimated optimal bandwidths are displayed in Tables 1 and 2 respectively together with the true optimal bandwidths. The corresponding boxplots of $ln\left(\frac{\hat{b}_{opt}}{b_{opt}}\right)$ are displayed in Figures 1 and 2. For this simulation study as well as for the applications in next section we used the truncated Gaussian density kernel with support $[-1, 1]$ as K (for (12)) and its second derivative as K_2 (for (21)).

Table 1: Simulation results: bias, variance and mean squared error of $\hat{b}_{opt}$ for $d = 0.1$

		General method			Long-memory method		
n,	b_{opt}	bias, variance, MSE			bias, variance, MSE		
50	0.390	0.0144	0.0056	0.0058	-0.0161	0.0170	0.0172
100	0.125	0.0025	0.0042	0.0042	0.0157	0.0074	0.0076
200	0.081	0.0176	0.0032	0.0035	0.0168	0.0039	0.0041
400	0.023	0.0310	0.0016	0.0025	0.0321	0.0016	0.0027

Table 2: Simulation results: bias, variance and mean squared error of $\hat{b}_{opt}$ for $d = 0.4$

		General method			Long-memory method		
n,	b_{opt}	bias, variance, MSE			bias, variance, MSE		
50	0.431	-0.0137	0.0079	0.0081	0.0078	0.0233	0.0234
100	0.240	-0.0490	0.0032	0.0056	-0.0408	0.0092	0.0109
200	0.102	0.0167	0.0031	0.0034	0.0022	0.0006	0.0006
400	0.047	0.0113	4.3e-05	0.0001	0.0037	7.3e-06	2.1e-05

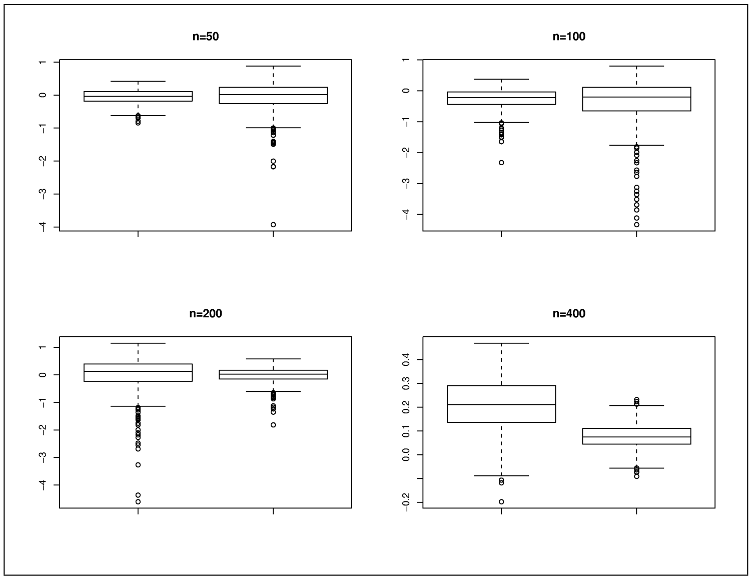

Figure 2: Boxplots of $ln\left(\frac{\hat{b}_{opt}}{b_{opt}}\right)$ computed by using the general method (left) and the long-memory method (right) for $d = 0.4$

It can be seen that in both cases the general method produced better results for smaller sample sizes ($n = 50,\ 100$). This is a natural result as large samples are needed just for detection of long memory. In the case $d = 0.1$, corresponding to weaker long-memory, the behaviors of the estimated optimal bandwidths computed by the two methods are comparable. As expected, for larger sample sizes and for strong long memory ($d = 0.4$), the long-memory method gave better results. This is intuitively clear as this method used the maximum likelihood estimator of the long-memory parameter.

6 Applications

6.1 The short-memory case

In this application we consider six time series of summer maxima of daily precipitation in the South of Switzerland. They were constructed based on records of daily precipitation totals (source: SMA Zürich, Switzerland) between 1901-1999. Summer is defined here as the season between April 1 and September 30. These time series didn't show any indication of long

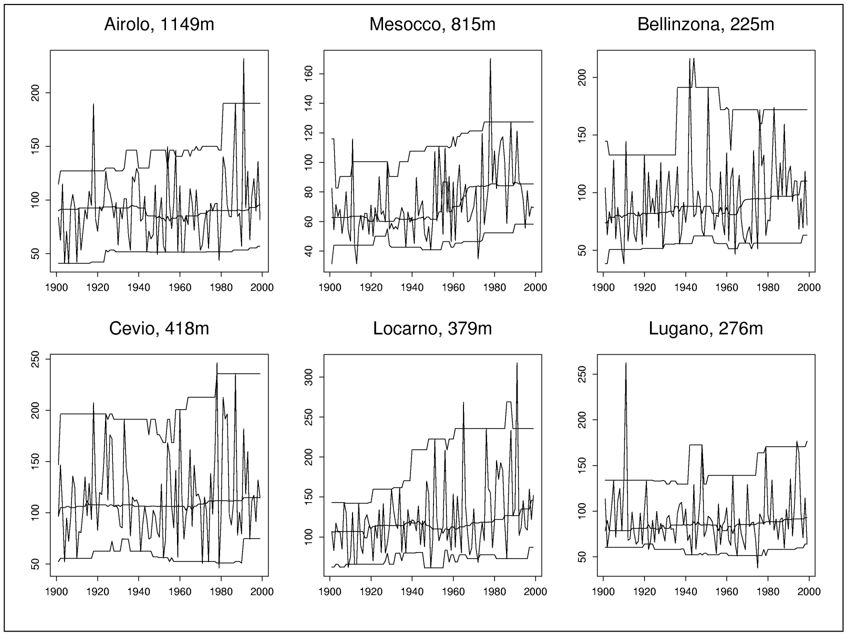

Figure 3: Estimated 0.05 quantile curve, median curve and 0.95 quantile curve for the time series of summer maxima of daily precipitation (in mm) at six meteo stations in the South of Switzerland

memory, but there were indications that their frequency distributions may have changed with time. Figure 3 displays these time series together with the estimated 0.05 quantiles, medians and 0.95 quantiles respectively based on the locally optimal bandwidths showed in Figure 4. For each time series we computed first the locally optimal bandwidths for each year based on the general method described in Section 4 and then we constructed quantile estimators by using (23) and the iterative procedure described in Ghosh and Draghicescu (2002b). By comparing the locally optimal bandwidths with the initial time series some interesting features can be observed. High variability in the data leads to larger values of the optimal bandwidths as can be seen for example in the case of Locarno for the period 1970-1990. This is intuitively clear as when the data fluctuates a lot, more observations are needed in the window around each time point to get accurate results. The opposite situation can be also seen. Thus, for example at Airolo and Lugano in the middle of the century, when the data are pretty stable, the optimal bandwidths have lower values.

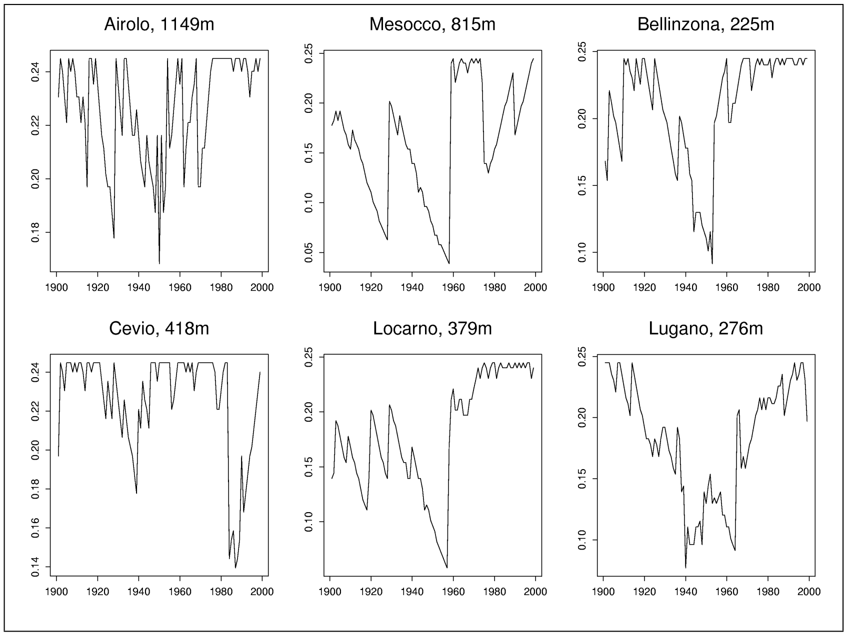

Figure 4: Estimated locally optimal bandwidths corresponding to the time series in Figure 3

6.2 The long-memory case

Figure 5 displays at the left the time series of mean July water surface elevation (in feet) for lake Huron between 1860 - 1986 at the station Harbor Beach Michigan. The series was taken from the Time Series Data Library on the webpage of Monash University, Australia. The log-log plot of the residuals is showed at the right in Figure 5 indicating long-memory. The trend of the initial time series was estimated by using the Splus function *lowess*. Here we computed the global optimal bandwidths $\hat{b}_{opt} = \frac{1}{n}\sum_{i=1}^{n} \hat{b}_{opt}(x_i)$ by using the general method and the long-memory method. In the latter case we obtained $\hat{b}_{opt} = 0.25$ with $\hat{d} = 0.19$. The general method yielded $\hat{b}_{opt} = 0.22$. Figure 6 displays the estimated median curves for this time series computed with the above optimal bandwidths. They are very similar in shape. In particular, they both capture the low levels around 1940.

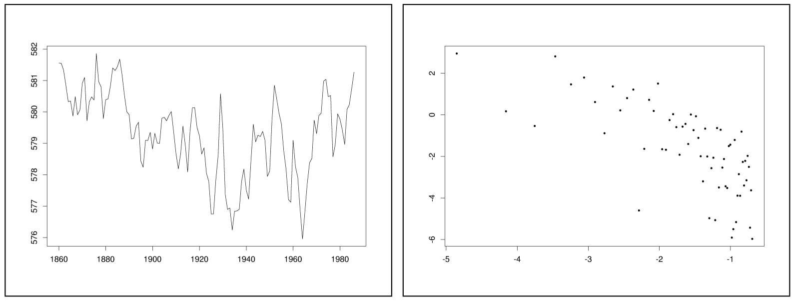

Figure 5: Mean July water surface elevation (in feet) of Lake Huron at Harbor Beach Michigan between 1860-1986 (left) and log-log plot of the periodogram of the residuals after trend removal (right)

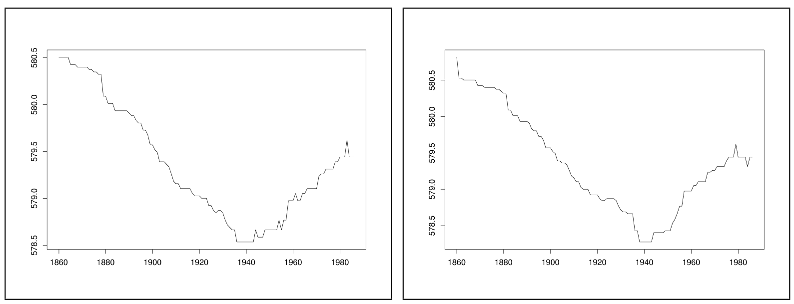

Figure 6: Estimated median curves (in feet) for the time series of mean level of lake Huron by using the long-memory method (left) and the general method (right)

7 Conclusion

The general method for optimal bandwidth selection discussed in this paper proves to be a useful tool for the estimation of probability distribution functions leading to quantile estimators in a very general class of stochastic processes. The method applies to nonstationary and nongaussian time series and does not involve testing, trend removal or applying transformations to the data. Not only the proposed bandwidth estimators have good asymptotic properties, they also behave well for small samples. When the data display strong long-range dependence and the sample size is sufficiently large, the efficiency of the general method can be improved and the computational time can be reduced by making use of the maximum likelihood estimator of the long-memory parameter.

Aknowledgements

I wish to thank Professor Michael L. Stein at the University of Chicago for reading this manuscript and for his valuable comments that helped improve its presentation. This research is based in part on my work for my PhD thesis, which was jointly supported by the Swiss Federal Research Institute WSL, Birmensdorf, Switzerland and the Swiss National Science Foundation grant 21-5541.98. I am grateful to my PhD advisor, Dr. Sucharita Ghosh at WSL, Birmensdorf, Switzerland for her helpful suggestions. The precipitation time series were provided by Dr. Christoph Frei at the Instite for Athmospheric and Climate Science, ETH, Zürich, Switzerland.

References

K. Abberger and S. Heiler, 2002, Nonparametric quantile regression with applications to financial time series, *Statistical Data Analysis Based on the L_1-Norm and related methods,* 149-160.

J. Beran, 1994, Statistics for long-memory processes, *Chapman & Hall.*

P. Breuer and P. Major, 1983, Central limit theorems for non-linear functionals of Gaussian fields, *J. Multiv. Analysis,*13:425:441.

M. Brockmann, T. Gasser and E. Herrmann, 1993, Locally adaptive bandwidth choice for kernel regression estimators, *JASA,* 88:1302-1309.

Z. Cai, 2002, Regression quantiles for time series, *Econometric Theory,* 18:169-192.

D. Draghicescu, 2002, Nonparametric quantile estimation for dependent data, *PhD thesis no. 2592, EPFL.*

D. Draghicescu and S. Ghosh, 2003, Smooth nonparametric quantiles, *Proceedings,* The 2nd International Colloquium for "Mathematics in Engineering and Numerical Physycs",22-27 April 2002, Bucharest, Romania (in press).

T. Gasser and H-G. Müller, 1984, Estimating Regression Functions and their Derivatives by the Kernel Method, *Scandinavian Journal of Statistics,* 11:171-185.

S. Ghosh, J. Beran and J. Innes, 1997, Nonparametric conditional quantile estimation in the presence of long memory, *Student*, Vol 2(2):109-117.

S. Ghosh and D. Draghicescu, 2002a, Predicting the distribution function for long-memory processes, *International Journal of Forecasting*, Vol. 18:283-290.

S. Ghosh and D. Draghicescu, 2002b, An algorithm for optimal bandwidth selection for smooth nonparametric quantile estimation, *Statistical Data Analysis Based on the L_1-Norm and related methods,* 161-168.

P. Hall and J. Marron, 1987, Estimation of integrated squared density derivatives, *Statist. Probab. Lett.,* 6:109-115.

J. Hart, 1996, Some automated methods of smoothing time-dependent data, *Nonparametric Statistics,* 6:115:142.

J. Hart (1997), Nonparametric smoothing and lack-of-fit tests, *Springer.*

M. C. Jones and S. J. Sheather, 1991, Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives, *Statist. Probab. Lett.,* 11:511-514.

R. Koenker and G. Bassett, 1978, Regression quantiles, *Econometrica,* 46:33-50.

M. B. Priestley, 1981, Spectral Analysis and Time Series, *Academic Press.*

S. J. Sheather and J. S. Marron, 1990, Kernel quantile estimators, *JASA,* 85:410-416.

J. S. Simonoff, 1996, Smoothing Methods in Statistics, *Springer.*

M. Taqqu, 1975, Weak convergence to Fractional Brownian Motion and to the Rosenblatt Process. *Prob. Th. Rel. Fields*, 31:287-302.