Fast Nonparametric Quantile Regression with Arbitrary Smoothing Methods

Hee-Seok Oh[*]   Thomas C. M. Lee[†]   Douglas W. Nychka[‡]

April 15, 2010; revised: August 13, 2010

Abstract

The calculation of nonparametric quantile regression curve estimates is often computational intensive, as typically an expensive nonlinear optimization problem is involved. This paper proposes a fast and easy–to–implement method for computing such estimates. The main idea is to approximate the costly nonlinear optimization by a sequence of well–studied penalized least–squares type nonparametric mean regression estimation problems. The new method can be paired with different nonparametric smoothing methods and can also be applied to higher dimensional settings. Therefore, it provides a unified framework for computing different types of nonparametric quantile regression estimates, and it also greatly broadens the scope of the applicability of quantile regression methodology. This wide-applicability and the practical performance of the proposed method are illustrated with smoothing spline and wavelet curve estimators, for both uni– and bivariate settings. Results from numerical experiments suggest that estimates obtained from the proposed method are superior to many competitors.

Keywords: bivariate quantile regression, nonparametric regression, pseudo data, regression quantile, wavelets.

Abbreviated Title: Fast Quantile Regression

[*]Associate Professor, Department of Statistics, Seoul National University, Seoul 151-747, Korea. Email: `heeseok@stats.snu.ac.kr`

[†]Professor, Department of Statistics, University of California at Davis, CA 95616. Email: `tcmlee@ucdavis.edu`

[‡]Director, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307. Email: `nychka@ucar.edu`

1 Introduction

Suppose that we have a set of independent measurements of random variables (X, Y). The goal of quantile regression analysis is to estimate the conditional τth quantile of the response Y given the (possibly vector–valued) covariate X, where $0 < \tau < 1$. When compared to the traditional conditional mean regression analysis, quantile regression is capable of providing more information about the conditional distribution $f_{Y|X}$ of Y given X. It is because the primary target of conditional mean regression is the conditional expectation $E(Y|X)$ (i.e., the "center" of $f_{Y|X}$), while by changing the values of τ, quantile regression analysis can be applied to explore the different behaviors of $f_{Y|X}$ at its center, lower and upper tails. Therefore, many useful methods for conducting quantile regression analysis have been proposed. For examples, Chaudhuri (1991a, 1991b) and Yu and Jones (1998) studied quantile estimates obtained from local polynomial regression, Lejeune and Sarda (1988) proposed the use of moving parabolic fitting, and Bhattacharya and Gangopadhyay (1990), Truong (1989) and Yu (1999) investigated kernel–type methods. In addition, univariate spline methods were considered by Koenker and Bassett (1978), Hendricks and Koenker (1992) and Koenker, Ng and Portnoy (1994), bivariate spline methods were studied by He, Ng and Portnoy (1998), while regression trees were used by Chaudhuri and Loh (2002). More recently, Li, Liu and Zhu (2007) studied the problem using the reproducing kernel Hilbert space methodology. Excellent comprehensive reviews of quantile regression were given by, for examples, Koenker (2005) and Yu, Lu and Stander (2003).

However, most of the above methods are computationally expensive, as quite often such methods involve a costly minimization of some nonlinear objective function. Also, most of these methods are tailored to a single nonparametric smoothing technique (e.g., smoothing splines or kernel methods) and are generally not straightforward to implement. The goal of this paper is to propose a general methodology for computing nonparametric quantile regression that do not suffer from these restrictions. As to be demonstrated below, this goal is achieved via the introduction of two new theoretical constructs, *pseudo quantile* and *pseudo data*.

A major contribution of this paper is a fast and easy–to–implement algorithm for com-

puting quantile regression curve estimates. This algorithm can be coupled with different nonparametric smoothing methods to compute the estimates, and it can also be applied to higher dimensional problems. Thus, the proposed algorithm permits the nonparametric quantile regression methodology to be extended to many different settings, including those that no previous nonparametric quantile regression estimators have been developed before. The main idea behind is to approximate the costly nonlinear optimization problem by a sequence of fast and well–studied penalized least–squares type conditional mean estimation problems.

The empirical properties of the proposed algorithm have been tested extensively through simulations. In below it is paired with smoothing spline and nonlinear wavelet curve estimators to obtain nonparametric quantile curve estimates for both uni– and bivariate problems. In addition, theoretical properties of the proposed methodology are also investigated. Lastly we note that the proposed algorithm was inspired by the robust smoothing method of Oh, Nychka and Lee (2007).

The rest of this article is organized as follows. Section 2 provides some background material and introduces the concept of pseudo quantile. Section 3 contains the main results of this article, namely, the proposed algorithm for computing nonparametric quantile estimates together with some theoretical development. In Sections 4 and 5, this algorithm is paired with, respectively, univariate and bivariate quantile smoothing splines to obtain quantile estimates. The wide-applicability of the algorithm is further illustrated in Section 6, in which it is coupled with nonlinear wavelet shrinkage methods. Finally concluding remarks are given in Section 7 while technical details are provided in the appendix.

2   Pseudo Quantile Estimation

2.1   Quantile and Pseudo Quantile

Suppose that $y \in \mathbb{R}$ is a random variable which is characterized by its distribution function $F(y) = P(Y \leq y)$. For any $\tau \in (0, 1)$, the τth quantile of y is defined as

$$F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}.$$

Similarly, the conditional quantile $F^{-1}(\tau|x)$ for a pair of random variables (x, y) is defined as

$$F^{-1}(\tau|x) = \inf\{y : F(y|x) \geq \tau\}.$$

It is known that the τth quantile can be also defined as the minimizer of an expected loss:

$$F^{-1}(\tau) = \arg\min_{\alpha} E\rho_\tau(y - \alpha), \tag{1}$$

where the so-called check function ρ_τ is

$$\rho_\tau(u) = \begin{cases} (\tau - 1)u & \text{if } u < 0 \\ \tau u & \text{if } u \geq 0. \end{cases}$$

Notice that the function ρ_τ is not differentiable at zero. This poses some technical issues in the study of quantile estimation. For example, the optimization of (1) is not trivial nor it is easy to study the theoretical properties of the corresponding minimizer. Moreover, extensions to higher dimensional settings are not straightforward.

To overcome these issues, we slightly modify ρ_τ so that it is differentiable at zero. The idea is to round its corner with a quadratic function, and we define *pseudo quantile* as

$$\arg\min_{\alpha} E\rho_{\tau,c}(y - \alpha), \tag{2}$$

where the modified check function is

$$\rho_{\tau,c}(u) = \begin{cases} (\tau - 1)(u + 0.5c) & \text{for } u < -c \\ 0.5(1 - \tau)u^2/c & \text{for } -c \leq u < 0 \\ 0.5\tau u^2/c & \text{for } 0 \leq u < c \\ \tau(u - 0.5c) & \text{for } c \leq u. \end{cases}$$

Observe that $\rho_{\tau,c}$ converges to ρ_τ as $c \to 0$. In our calculations, c is chosen to be effectively zero when comparing to the magnitude of the data values.

The following proposition states that the pseudo quantile induced by a small c is almost identical to the true quantile. Its proof is given in the online Appendix.

Proposition 1. *Let α_0 be the τth quantile defined by (1) and α^* be the τth pseudo quantile defined by (2). Denote respectively the density and distribution function of the random variable Y as f and F. Then, we have $|F(\alpha^*) - F(\alpha_0)| = |F(\alpha^*) - \tau| \leq \frac{c}{2}\sup|f|$.*

4

2.2 Pseudo Sample Quantile

Given sample observations $\{y_1, y_2, \ldots, y_n\}$, the τth sample quantile estimate can be obtained by sorting and ordering the sample observations. Alternatively, the τth sample quantile can be also obtained as

$$\arg\min_{\alpha} \int \rho_\tau(y - \alpha) dF_n(y) = \arg\min_{\alpha} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha),$$

where F_n is the empirical distribution function of $\{y_1, \ldots, y_n\}$. Therefore, the problem of finding the τth sample quantile can be expressed as

$$\min_{\alpha} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha). \tag{3}$$

As a typical example, minimizing the l_1-loss function for a location estimator gives the median. That is, minimizing $\sum_{i=1}^{n} |y_i - \alpha|$ with respect to α is equivalent to that an equal number of terms $y_i - \alpha$ lie on either side of zero in order for the derivative with respect to α to vanish.

Due to the non-differentiability of ρ_τ at zero, optimization of (3) is not always trivial. Therefore, as in the previous subsection, we replace ρ_τ with the modified check function $\rho_{\tau,c}$. That is, we approximate the solution to (3) with the solution of

$$\min_{\alpha} \sum_{i=1}^{n} \rho_{\tau,c}(y_i - \alpha), \tag{4}$$

which we shall call the τth pseudo sample quantile.

Proposition 2. *Let $y_1, y_2, \ldots, y_n$ be n random observations and $\tau \in (0, 1)$. The solution of (4), denoted as α_τ, satisfies:*

1. the number of terms, n_-, with $y_i < \alpha_\tau$ is bounded from above by $\tau n + \delta$,

2. the number of terms, n_+, with $y_i > \alpha_\tau$ is bounded from above by $(1 - \tau)n + \delta'$, and

3. as $n \to \infty$, $n_-/n \to \tau + \delta''$ provided that $P(y)$ does not contain discrete components,

where δ, δ' and δ'' are negligible when c is chosen to be effectively zero relative to the magnitude of the data values.

Proposition 2 states that the minimizer of (4) with a small c adequately approximates the classical τth sample quantile defined by (3). The proof is given in the online Appendix.

2.3 Nonparametric Quantile Regression

We now extend our discussion from quantile estimation to nonparametric quantile regression. Suppose n pairs of independent measurements $\{(x_i, y_i)\}_{i=1}^n$ are observed. For any $\tau \in (0, 1)$, the conditional τth quantile of y_i given x_i is defined as

$$q_\tau(y_i|x_i) = \inf\{y_i : P(Y \le y_i|x_i) \ge \tau\}.$$

To simplify notation, we shall write $f_\tau(x_i) = q_\tau(y_i|x_i)$. The goal is to, given the measurements $\{(x_i, y_i)\}_{i=1}^n$ and a specific value of τ, estimate $f_\tau(x)$.

Given a nonnegative penalty function $p(\cdot)$ and a smoothing parameter λ, one possible method for estimating f_τ is to define the estimate as the minimizer of the following criterion (e.g., see Koenker et al. 1994 and He et al. 1998):

$$\sum_{i=1}^n \rho_\tau\{y_i - f(x_i)\} + \lambda p(f). \tag{5}$$

A typical example for $p(\cdot)$ is $p(f) = \int (f'')^2 dx$, which can be considered as an quantile extension of the classical smoothing splines (Nychka et al., 1995).

As similar to above, we shall consider approximating the above minimizer with the *pseudo quantile regression estimate* $\hat{f}_\tau$, defined as

$$\arg\min_f \left[\sum_{i=1}^n \rho_{\tau,c}\{y_i - f(x_i)\} + \lambda p(f)\right]. \tag{6}$$

3 Pseudo Nonparametric Quantile Regression

This section presents the main results of this paper: it examines the theoretical properties of $\hat{f}_\tau$ defined by (6), and develops a fast algorithm for computing it.

3.1 Pseudo Quantile Regression and Pseudo Data

Let $f_i = f(x_i)$ and $\boldsymbol{f} = (f_1, \ldots, f_n)$. A discretized version of (6) is

$$\hat{\boldsymbol{f}}_\tau = \arg\min_{\boldsymbol{f}} \sum_{i=1}^n \rho_{\tau,c}(y_i - f_i) + \lambda p(\boldsymbol{f}), \tag{7}$$

where $\hat{\boldsymbol{f}}_\tau$ is the discretized estimate of f_τ. A necessary condition for $\hat{\boldsymbol{f}}_\tau$ is that it solves

$$-\psi_{\tau,c}(y_i - f_i) + \lambda \frac{\partial p(\boldsymbol{f})}{\partial f_i} = 0, \qquad i = 1, \ldots, n, \tag{8}$$

where $\psi_{\tau,c} = \rho'_{\tau,c}$. However, due to the nonlinear nature of $\rho_{\tau,c}$, it is not trivial to find the solution.

Now we define the theoretical construct *pseudo data* $\tilde{y}_i$ that facilities the theoretical and algorithmic development of our work:

$$\tilde{y}_i = f_i + \frac{\psi_{\tau,c}\{y_i - f_i\}}{2}, \quad i = 1, \ldots, n.$$

Of course in practice these pseudo data $\tilde{y}_i$'s are unknown, but for the moment let us suppose that they are available, and define $\tilde{\boldsymbol{f}}_\tau$ as

$$\tilde{\boldsymbol{f}}_\tau = \arg\min_{\boldsymbol{f}} \sum_{i=1}^{n} (\tilde{y}_i - f_i)^2 + \lambda p(\boldsymbol{f}). \tag{9}$$

As similar to above, it is necessary for $\tilde{\boldsymbol{f}}_\tau$ to solve, for all i,

$$-2(\tilde{y}_i - f_i) + \lambda \frac{\partial p(\boldsymbol{f})}{\partial f_i} = -2\left[f_i + \frac{\psi_{\tau,c}\{y_i - f_i\}}{2} - f_i \right] + \lambda \frac{\partial p(\boldsymbol{f})}{\partial f_i},$$

which is identical to (8). This means that both $\hat{\boldsymbol{f}}_\tau$ and $\tilde{\boldsymbol{f}}_\tau$ solve the same equation, namely, (8). It raises two important questions: are $\hat{\boldsymbol{f}}_\tau$ and $\tilde{\boldsymbol{f}}_\tau$ the same? And if yes, what do we gain if we had $\tilde{y}_i$?

Next subsection shows $\hat{\boldsymbol{f}}_\tau$ and $\tilde{\boldsymbol{f}}_\tau$ are asymptotically same. Given this and if the $\tilde{y}_i$'s were known, we could compute $\hat{\boldsymbol{f}}_\tau$, which is the same as $\tilde{\boldsymbol{f}}_\tau$, by minimizing the penalized least-squares criterion in (9) instead of the nonlinear functional in (7). In other words, the nonlinear optimization problem posed by (7) can now be solved much quicker as a least–squares type linear optimization problem. Although in practice the pseudo data $\tilde{y}_i$'s are unknown, but this suggests a fast algorithm for approximating $\hat{\boldsymbol{f}}_\tau$, as to be described in Section 3.3. We note that this concept of pseudo data has been adopted by previous authors to derive asymptotic properties of robust smoothing splines, and to develop automatic regularization parameter selection methods (Cox, 1984; Cantoni and Ronchetti, 2001; and Oh et al., 2007).

3.2 Equivalence Result

Here we establish the asymptotic equivalence of $\hat{\boldsymbol{f}}_\tau$ and $\tilde{\boldsymbol{f}}_\tau$. Our results are similar to those provided by Theorem 1 of Oh et al. (2007). We first highlight some important assumptions.

The modified check function $\rho_{\tau,c}$ needs be altered to satisfy several assumptions, including the function $\psi_{\tau,c} = \rho'_\tau$ has a continuous second derivative and satisfies $\sup |\psi''_{\tau,c}(x)| < \infty$, and $\text{Var}(\psi_{\tau,c}(x)) < \infty$ and $\text{Var}(\psi'_{\tau,c}(x)) < \infty$. It should also be strictly convex and be qualitatively similar to $\rho_{\tau,c}$. One possible choice is $\tilde{\rho}_{\tau,c}(x) = \log[2c^2\{e^{\tau x/c} + e^{-(1-\tau)x/c}\}]$. Note that c is the parameter to control the amount of (continuously) cutoff. As long as $c \to 0$, this function is almost identical to the check function.

It is required to assume the existence of a smoothing matrix G_λ which, when applied to the pseudo data, produces the penalized least squares estimates in (9). It is also assumed that there exist a small $c < \delta$ such that the condition number of the operator G_λ is bounded for all $\lambda \in [\lambda_n, \Lambda_n]$ and all τ. This assumption is related to the following. If c is very small but not exactly zero, the design points x_i are distinct under $\tilde{\rho}_{\tau,c}$. Then G_λ^{-1} can be well defined.

Theorem 1. *Let* $\boldsymbol{f}_\tau = (f_\tau(x_1), \ldots, f_\tau(x_n))^T$. *Under the assumptions of Theorem 1 of Oh et al. (2007), for any $\delta > 0$, there is an n_0 such that for all $n \geq n_0$ and $\lambda \in [\lambda_n, \Lambda_n]$,*

$$P\left[\|\hat{\boldsymbol{f}}_\tau - \tilde{\boldsymbol{f}}_\tau\|_n \leq \delta\sqrt{E\|\tilde{\boldsymbol{f}}_\tau - \boldsymbol{f}_\tau\|_n^2}\right] > 1 - \delta,$$

where $\|x\|_n^2 = \sum_{i=1}^n x_i^2/n$.

Theorem 1 states that with high probability, $\hat{\boldsymbol{f}}_\tau$ and $\tilde{\boldsymbol{f}}_\tau$ are much closer than $\tilde{\boldsymbol{f}}_\tau$ and $\boldsymbol{f}_\tau$. In other words, $\|\tilde{\boldsymbol{f}}_\tau - \hat{\boldsymbol{f}}_\tau\|_n^2 / E\|\tilde{\boldsymbol{f}}_\tau - \boldsymbol{f}_\tau\|_n^2$ converges to zero in probability as $n \to \infty$.

The proof of Theorem 1 follows the main steps of the proof of Oh et al. (2007), and can be obtained from the authors.

3.3 The ES-Algorithm

We have established the closeness between $\hat{\boldsymbol{f}}_\tau$ and $\tilde{\boldsymbol{f}}_\tau$, but in practice the pseudo data $\tilde{y}_i$'s are not available. Therefore, $\hat{\boldsymbol{f}}_\tau$ cannot be computed quickly as the minimizer of (9). However, the above discussion suggests the following heuristics for computing $\hat{\boldsymbol{f}}_\tau$: given a current

estimate of $\boldsymbol{f}_\tau$, first calculate the corresponding $\tilde{y}_i$'s, then plug in these newly calculated $\tilde{y}_i$'s in (9) to obtain the next iterative estimate for $\boldsymbol{f}_\tau$. This motivates the following algorithm, termed the ES-algorithm, for computing $\hat{\boldsymbol{f}}_\tau$:

1. Obtain an initial τth quantile estimate $\hat{\boldsymbol{f}}_\tau^{(0)} = (\hat{f}_\tau^{(0)}(x_1), \ldots, \hat{f}_\tau^{(0)}(x_n))^T$.

2. Iterate, until convergence, the following two steps for $l = 0, 1, \ldots,$

E-Step: (**E**valuation of empirical pseudo data): for $i = 1, \ldots, n$, compute the following *empirical pseudo data*

$$s_i^{(l)} = \hat{f}_\tau^{(l)}(x_i) + \frac{\psi_{\tau,c}\{y_i - \hat{f}_\tau^{(l)}(x_i)\}}{2}.$$

S-Step: (**S**moothing of empirical pseudo data): obtain the next iterative estimate $\hat{\boldsymbol{f}}_\tau^{(l+1)}$ by minimizing the penalized least–squares criterion (9) with $\tilde{y}_i = s_i^{(l)}$ for all i.

3. Take the converged estimate as the final τth quantile estimate for $\boldsymbol{f}_\tau$.

We have the following remarks.

1. The ES-algorithm essentially performs a sequence of penalized least-squares minimization (S-Step). Many fast, stable and well-studied methods are available for executing this task. Therefore, it is relatively fast and can be easily implemented.

2. The S-Step states the ES-algorithm can be coupled with any nonparametric smoothing methods, as long as they can be formulated as a penalized least–squares problem. In fact, in practice one could apply other nonparametric methods to smooth the empirical pseudo data $s_i^{(l)}$; e.g., nonlinear wavelet shrinkage as demonstrated in Section 6 below. However, we note that these other methods may not be covered by Theorem 1.

3. The above description of the ES-algorithm does not depend on the dimension of the covariates x_i's. Thus, it can also be applied to higher dimensional quantile smoothing problems. Examples of bivariate quantile smoothing are given in Sections 5 and 6.

4. An initial estimate $\hat{f}_\tau^{(0)}$ is required to start the algorithm. Our numerical experience suggests that, as long as $\hat{f}_\tau^{(0)}$ is of reasonable quality, it does not have a huge effect on

9

the final estimate. In our numerical work, we use the standard k-NN (nearest neighbor) technique to obtain a quick $\hat{f}_\tau^{(0)}$ (e.g., Bhattacharya and Gangopadhyay, 1990 and Yu, 1999). The idea of k-NN is to treat the k observations $(x'_1, y'_1), \ldots, (x'_k, y'_k)$ that are closest to x_0 as independent replicates and estimate $f_\tau(x_0)$ by the τth quantile of the empirical distribution of $y'_1, \ldots, y'_k$. For the univariate setting, the simple method of Healy, Rasbash and Yang (1988) can be adopted to obtain improved k-NN estimates.

5. In all our numerical work we have never encountered a case that the ES-algorithm did not converge. In fact, for most cases the algorithm converged within 5 iterations. In particular, for the setting described in Section 4.1 below, for a data set with $n = 2000$, the ES-algorithm on average took less than 0.3 second to finish one fitting on a MacBook Pro machine with a 2.53 GHz Intel Core 2 Duo processor.

3.4 Smoothing Parameter Selection

In the computation of $\hat{\boldsymbol{f}}_\tau$, an important issue that needs to be addressed is the choice of the smoothing parameter λ. Although this is not a major focus of the present paper, this subsection provides a brief discussion about this issue.

3.4.1 Quantile Cross-Validation

In the context of nonparametric regression *mean* estimation, a common smoothing parameter selection method is cross-validation (CV). If $\hat{f}_\lambda^{-i}(x_i)$ denotes the usual leave-one-out estimate for $f(x_i)$ computed with λ, then CV chooses λ as the minimizer of

$$\mathrm{CV}(\lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{f}_\lambda^{-i}(x_i)\}^2.$$

This CV criterion can be generalized to nonparametric quantile regression. A natural candidate is the following quantile cross-validation (QCV) criterion:

$$\mathrm{QCV}(\lambda) = \frac{1}{n} \sum_{i=1}^n \rho_\tau \{y_i - \hat{f}_{\tau,\lambda}^{-i}(x_i)\}, \tag{10}$$

where $\hat{f}_{\tau,\lambda}^{-i}(x_i)$ is the estimate of $f_\tau(x_i)$ computed with smoothing parameter λ using all but the ith observation (x_i, y_i). This QCV criterion has been considered by previous authors

10

(e.g., Nychka et al., 1995 and Yuan, 2006), but due to the heavy computations involved, these previous authors only provide approximated calculations of $\text{QCV}(\lambda)$. Now with the above ES-algorithm, the term $\hat{f}_{\tau,\lambda}^{-i}(x_i)$ can be computed relatively quick and hence $\text{QCV}(\lambda)$ can be calculated directly from its definition, as least when the number of data points n is not too large. To see this, let N_I be the number of iterations needed for the ES-algorithm to converge, and N_λ be the number of λ values from which the search of the minimum of $\text{QCV}(\lambda)$ is limited to. Then, the minimization of $\text{QCV}(\lambda)$ requires $n N_I N_\lambda$ calculations of the term $\hat{f}_{\tau,\lambda}^{-i}(x_i)$, which is roughly equal to executing $n N_I N_\lambda$ times the L_2 minimization inside the S-Step. In Section 4, this QCV method is used for choosing λ for univariate quantile smoothing splines.

3.4.2 A Faster Alternative

If the number of data points n is large, then the above direct calculation of $\text{QCV}(\lambda)$ may not be practically feasible. To overcome this issue, we suggest a simple and yet effective alternative for choosing the amount of smoothing. This alternative was motivated by the observation that every S-Step can be seen as a penalized least–squares type nonparametric *mean* regression problem. Thus, at each execution of the S-Step, one can apply a reliable and well-studied smoothing parameter selection method designed for nonparametric *mean* regression estimation to choose λ. Similar ideas have been adopted for example by Gu (1992). Our extensive experience suggests that this alternative works really well in practice, and it can be shown that for many cases, this alternative is roughly n times faster than the QCV approach discussed before. We use this method to choose the amount of smoothing for bivariate quantile smoothing splines (Section 5) and wavelet quantile regression (Section 6).

4 Univariate Quantile Smoothing Splines

This section presents our first application of the ES-algorithm: univariate quantile smoothing spline fitting with penalty term $p(f) = \int \{f''(x)\}^2 dx$ (e.g., see Nychka et al., 1995 and

11

references given therein). That is, $\hat{f}_\tau$ is defined as

$$\hat{f}_\tau = \arg\min_f \left[\sum_{i=1}^{n} \rho_\tau\{y_i - f(x_i)\} + \lambda \int \{f''(x)\}^2 dx \right]. \tag{11}$$

The above criterion can be minimized by applying the ES-algorithm in following manner. In the S-Step of the algorithm, apply least–squares smoothing spline to smooth the empirical pseudo data $s_i^{(l)}$'s to obtain the next iterative estimate $\hat{\boldsymbol{f}}_\tau^{(l+1)}$ (e.g., see Green and Silverman, 1994). For the choice of λ, we use $\mathrm{QCV}(\lambda)$ as defined in (10).

4.1 Simulation Results

A simulation study was conducted to evaluate the practical performance of the ES-algorithm with univariate smoothing splines. Two different noise types were considered: increasing–variance and asymmetric. For $x_i = (i-1)/2000$, $i = 1, \ldots, 2000$, the increasing–variance and the asymmetric noisy data sets were generated, respectively, from

$$y_i = \sin(10x_i) + \frac{(x_i + 0.25)}{0.1}\epsilon_i, \quad \epsilon_i \sim \text{ iid } N(0, 0.07^2)$$

and

$$y_i = \sin(10x_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{ iid gamma random variables with } E(\varepsilon_i) = \text{Var}(\varepsilon_i) = 3.$$

For each noise type, 100 data sets were generated, and for each generated data set, the following three nonparametric quantile regression estimators were applied to estimate f_τ for $\tau = (0.1, 0.5, 0.9)$:

- `ES-SS`: The quantile smoothing spline (11) approximated by the ES-algorithm. The initial fit was obtained by the k-NN technique with the modification of Healy et al. (1988). The value of k was 100.

- `Two-Step`: The two-step method proposed by Yu (1999). This method combines the use of k-NN and local linear smoothing.

- `QSS`: The quantile smoothing spline method by Koenker et al. (1994). In this method, instead of L_2, a L_1 penalty $\int |f''(x)|dx$ is imposed on $\hat{f}_\tau$, and the final estimate is computed by linear programming.
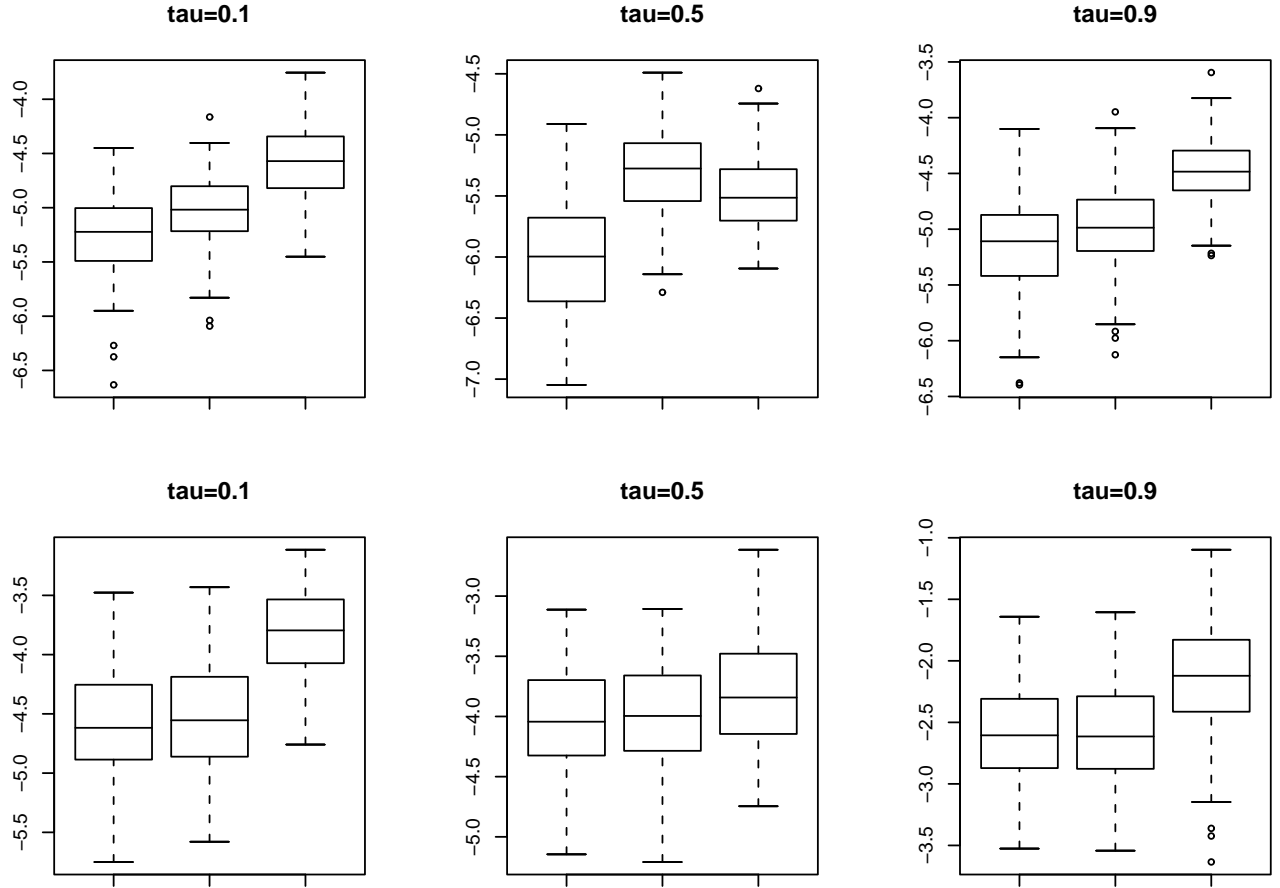
12

Figure 1: Boxplots of the log of the MSE values for the simulations in Section 4.1. The top row is for increasing–variance noise while the bottom row is for asymmetric noise. In each panel the boxplots, from left to right, correspond respectively to `ES-SS`, `Two-Step` and `QSS`.

For each estimate $\hat{f}_\tau$, the following mean-squared-error (MSE) was calculated:

$$\mathrm{MSE}(\hat{f}_\tau) = \frac{1}{n} \sum_{i=1}^{n} \left\{ f_\tau(x_i) - \hat{f}_\tau(x_i) \right\}^2 .$$

Boxplots of the log of these calculated MSEs are given in Figure 1. These plots seem to suggest that `ES-SS` is the preferred method. It is because it never performed worse than the remaining two methods, while in some situations it performed better. Results from pairwise t-tests also support this claim. To visually evaluate the quality of various curve estimates, we randomly selected one generated data set from each noise type and plotted the true and estimated quantile curves in Figure 2. Qualitatively the estimated curves produced from `ES-SS` and `Two-Step` are similar and better than those from `QSS`.
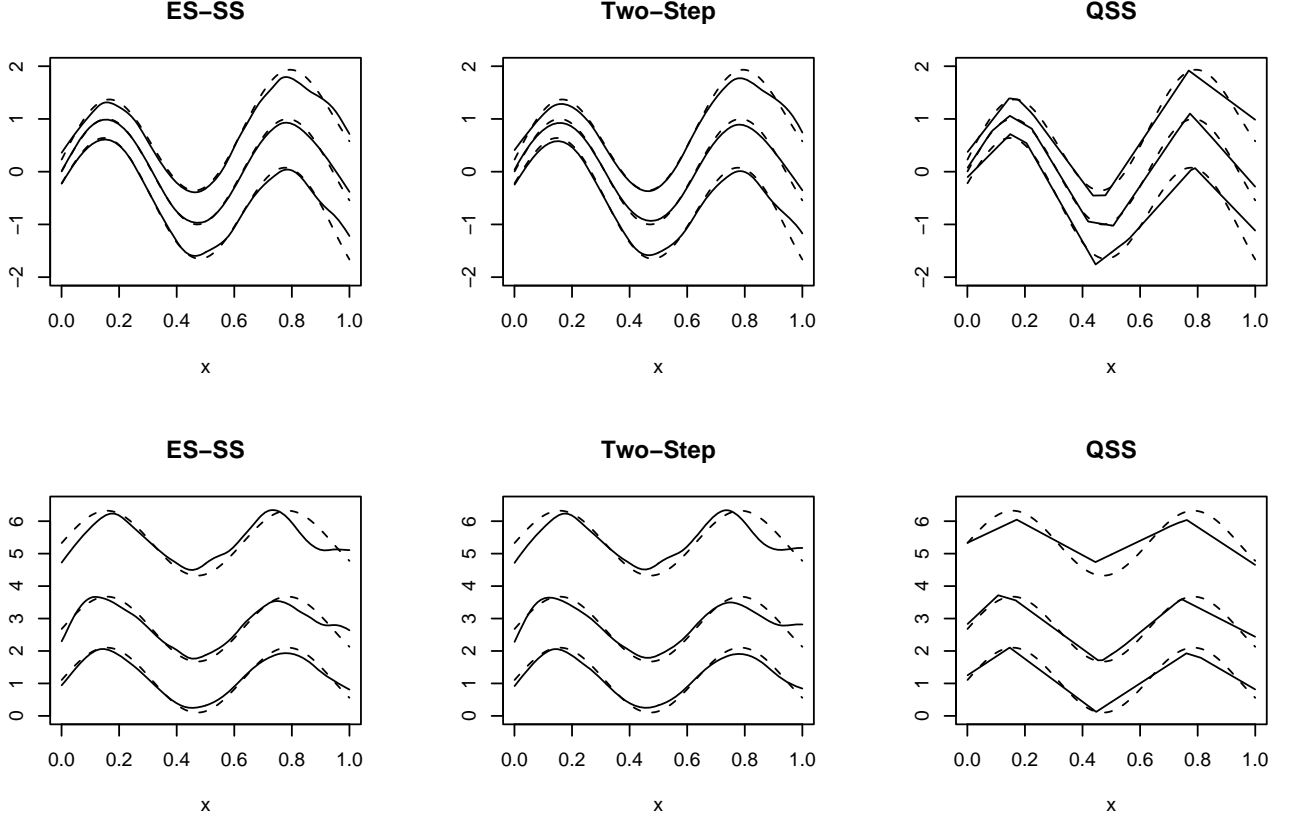
Figure 2: Quantile curve estimates obtained by the methods compared in Section 4.1. The top row is for increasing–variance noise while the bottom row is for asymmetric noise. In each panel the solid lines, from bottom to top, correspond to the quantile curve estimates computed for $\tau = 0.1, 0.5$ and 0.9, while the broken lines correspond to the true quantile functions for the same τ.

5 Bivariate Quantile Smoothing Splines

Next we consider bivariate nonparametric quantile regression. Given response variables z_i observed at (x_i, y_i), $i = 1, \ldots, n$, our bivariate smoothing spline estimate $\hat{f}_\tau(x, y)$ of the quantile regression function $f_\tau(x, y)$ is defined as

$$\hat{f}_\tau(x, y) = \arg\min_f \left[\sum_{i=1}^n \rho_\tau\{z_i - f(x_i, y_i)\} + \lambda p(f) \right],$$

where the penalty is defined as

$$p(f) = \int \int_{R^2} \left(\frac{\partial^2 f}{\partial x^2} + 2\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2} \right) dx dy.$$

In the classical bivariate penalized least–squares regression setting for iid Gaussian noise, the use of the above penalty is known as thin–plate spline smoothing (e.g., see Green and Silverman, 1994). Thus, the ES-algorithm can be applied to obtain $\hat{f}_\tau(x, y)$ by applying a thin–plate spline smoothing operation in the S-Step. In below the k-NN technique was used to obtain the initial fit with $k = 3^2$. For selecting the amount of smoothing, we adopted the approach described in Section 3.4.2 and used generalized cross-validation to choose λ in each S-Step. We call the resulting bivariate quantile regression estimate `ES-TPS`.

5.1 Simulation Results

Numerical experiments were conducted for assessing the performance of `ES-TPS`. The setup is the same as in He et al. (1998). The data were generated from the following model

$$z_i = \sin(3\pi x_i)\cos(\pi y_i) + \epsilon_i/3 \qquad i = 1, \ldots, 100,$$

where the (x_i, y_i)'s form a regularly-spaced grid of size 10×10 in the squares $[0.1, 1] \times [0.1, 1]$. The following four types of iid noise ϵ_i's are considered:

- Normal: standard normal distribution $\Phi(x)$,

- CN(0.05): a 5% contaminated normal mixture distribution $0.95\Phi(x) + 0.05\Phi(x/5)$,

- CN(0.10): a 10% contaminated normal mixture distribution $0.9\Phi(x) + 0.1\Phi(x/5)$, and

- T(3): t-distribution with three degrees of freedom.

Altogether 1000 data samples for each type of noise were generated. For each generated sample, the `ES-TPS` procedure was applied to compute $\hat{f}_\tau(x, y)$ with $\tau = 0.5$, and the corresponding MSE was also computed:

$$\text{MSE} = \frac{1}{100} \sum_{i=1}^{100} \left\{ f(x_i, y_i) - \hat{f}_{0.5}(x_i, y_i) \right\}^2.$$

The average and the standard error of the computed MSE values are listed in Table 1. Also listed in Table 1 are the corresponding MSE averages and standard errors obtained from the following four bivariate smoothing methods:

Table 1: Averages and standard errors (in parentheses) of MSE values computed from estimates of methods compared in Section 5.1.

Method	noise			
	Normal	CN(0.05)	CN(0.10)	T(3)
ES-TPS	0.0522	0.0568	0.0616	0.0665
	(0.00038)	(0.00039)	(0.00043)	(0.00051)
BMSS	0.0529	0.0577	0.0637	0.0690
	(0.00038)	(0.00054)	(0.00066)	(0.00070)
LOESS	0.0424	0.0634	0.0915	0.0800
	(0.00028)	(0.00086)	(0.00129)	(0.00176)
TPS	0.0381	0.0647	0.0922	0.0827
	(0.00042)	(0.00101)	(0.00158)	(0.00206)
MARS	0.0447	0.0918	0.1400	0.1150
	(0.00044)	(0.00154)	(0.00221)	(0.00289)

- BMSS: the bivariate median smoothing spline of He et al. (1998),

- LOESS: the least-squares based LOESS,

- TPS: least–squares thin-plate spline fitting, and

- MARS: MARS of Friedman (1991).

The MSE values for these four methods are copied from Table 1 of He et al. (1998). Note that the last three methods are targeting the conditional mean of the regression function, not the median.

The following empirical conclusions can be made from Table 1: (i) the mean smoothing estimators (LOESS, TPS and MARS) are more efficient for normal errors; (ii) both quantile smoothing procedures (ES-TPS and BMSS) outperform the mean smoothing procedures when the errors are T(3), CN(0.05) and CN(0.10); and (iii) ES-TPS seems to dominate BMSS.
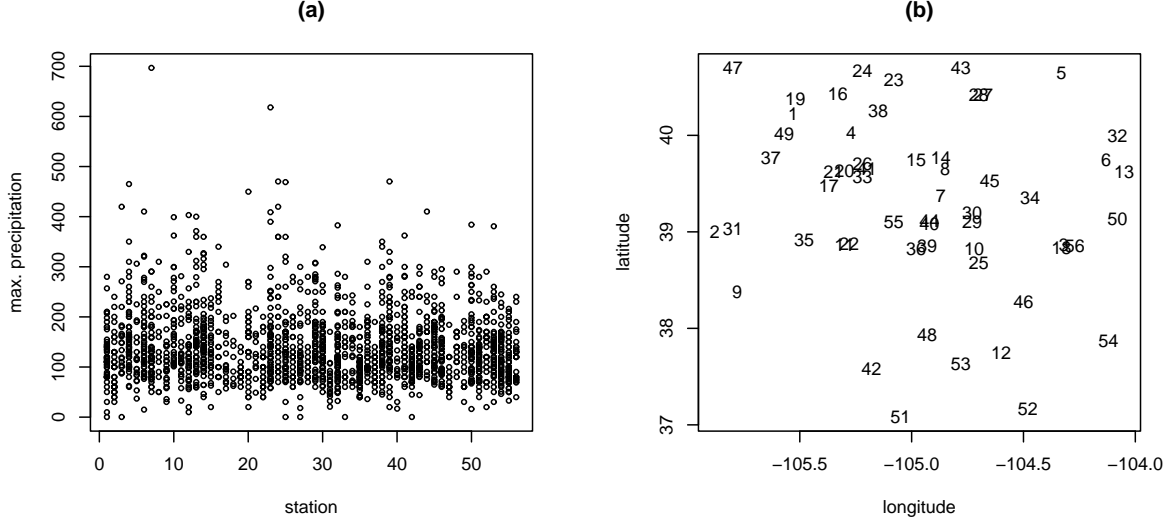
Figure 3: (a) Annual daily maximum precipitation according to 56 stations of Colorado's Front Range. (b) The locations of 56 stations.

5.2   Precipitation Data

We apply the proposed method to the precipitation on the Front Range of Colorado previously analyzed by Cooley et at. (2005). The original data contains daily precipitation from a network of 56 stations of the Front Range that lies between 104.0 and 105.9 W longitude and 37.08 and 41.70 N latitude (Figure 3). Some of the stations have 50 year records and others have shorter ones. For our analysis, we take the annual daily maximum precipitation from each station which is shown in Figure 3(a). Figure 3(b) displays the locations of 56 stations. The goal of analyzing the precipitation data is to estimate fields of precipitation return levels. In hydrology, the return level is a common measure of extreme events. The a-year return level is the quantile that has probability $1/a$ of being exceeded in a particular year.

The ES-algorithm coupled with thin-plate splines is applied to the annual daily maximum precipitation in Figure 3. Then we obtain maps of 2-year and 10-year return levels, that is, 50% and 90% quantiles of the precipitation of the Front Range, which has the interpretation that on the average 1 out of 2 and 1 out of 10 years will exceed these values. The results are displayed in Figure 4. These estimated surfaces can be used to aid designing new drainage
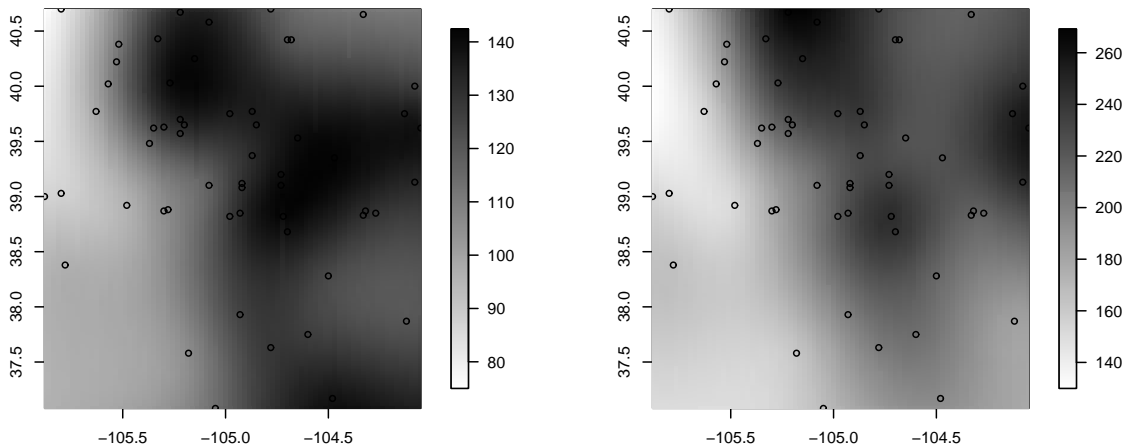
17

Figure 4: (a) The estimate of the 50% quantile surface of the annual daily maximum precipitation by the proposed method. (b) Similar to (a) but for the 90% quantile surface.

systems in the area.

6   Uni- and Bivariate Wavelet Quantile Regression

It has been shown that for the standard nonparametric mean regression setting, wavelet techniques are particularly good at capturing discontinuities and abrupt structures in the regression functions. This section demonstrates that the ES-algorithm can be coupled with wavelet shrinkage methods for performing nonparametric quantile regression. To the best of our knowledge, no wavelet quantile regression methods have been proposed in the literature.

It is known that wavelet shrinkage for regression mean estimation can be posed as a penalized least–squares problem (e.g., Antoniadis and Fan, 2001). In other words, the wavelet shrinkage estimator can be expressed in the form of (5). Thus, the ES-algorithm can be applied straightforwardly with wavelet methods: simply perform a wavelet shrinkage operation in the S-Step of the algorithm. The wavelet shrinkage operation can be either univariate or bivariate. As before, we use the k-NN technique to obtain initial fits.

We illustrate the idea with two examples. In the first univariate example, the noisy data with $n = 4096$ were generated by superimposing independent $N(0, 0.1^2)$ noise to a testing
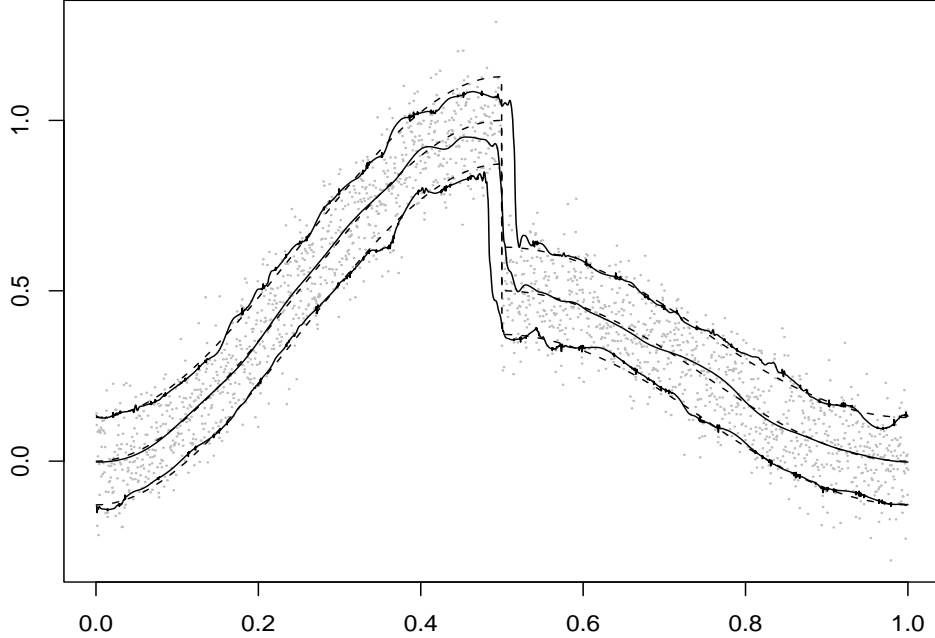
18

Figure 5: Solid lines are the estimated wavelet quantile functions for $\tau = (0.1, 0.5, 0.9)$ and dashed lines are the corresponding true quantile functions.

regression function. The testing regression function can be obtained from the `WaveThresh3` package of Nason (1998), and is plotted in Figure 5. The initial fit was computed using the k-NN technique with $k = 42$, and the empirical Bayes thresholding method of Johnstone and Silverman (2005) was used in the S-Step. In Figure 5, solid lines denote the estimated wavelet quantile functions for $\tau = (0.1, 0.5, 0.9)$ and dashed lines are the corresponding true quantile functions. One can see the estimated quantile curves follow quite closely to the true quantile curves.

Bivariate wavelet image quantile estimation can also be achieved by simply replacing the univariate wavelet shrinkage operation in the S-Step with a bivariate wavelet shrinkage operation. The image, of size 256×256, displayed in Figure 6 is the data `lennon` which is also available from `WaveThresh3`. Artificial additive independent Gaussian noise was added with a signal-to-noise ratio 3. The bivariate wavelet shrinkage method used is the false discovery rate rule of Abramovich and Benjamin (1995). We employed a two-dimensional version of the k-NN method to obtain the initial fit with $k = 4^2$. The noisy image and the τth quantile
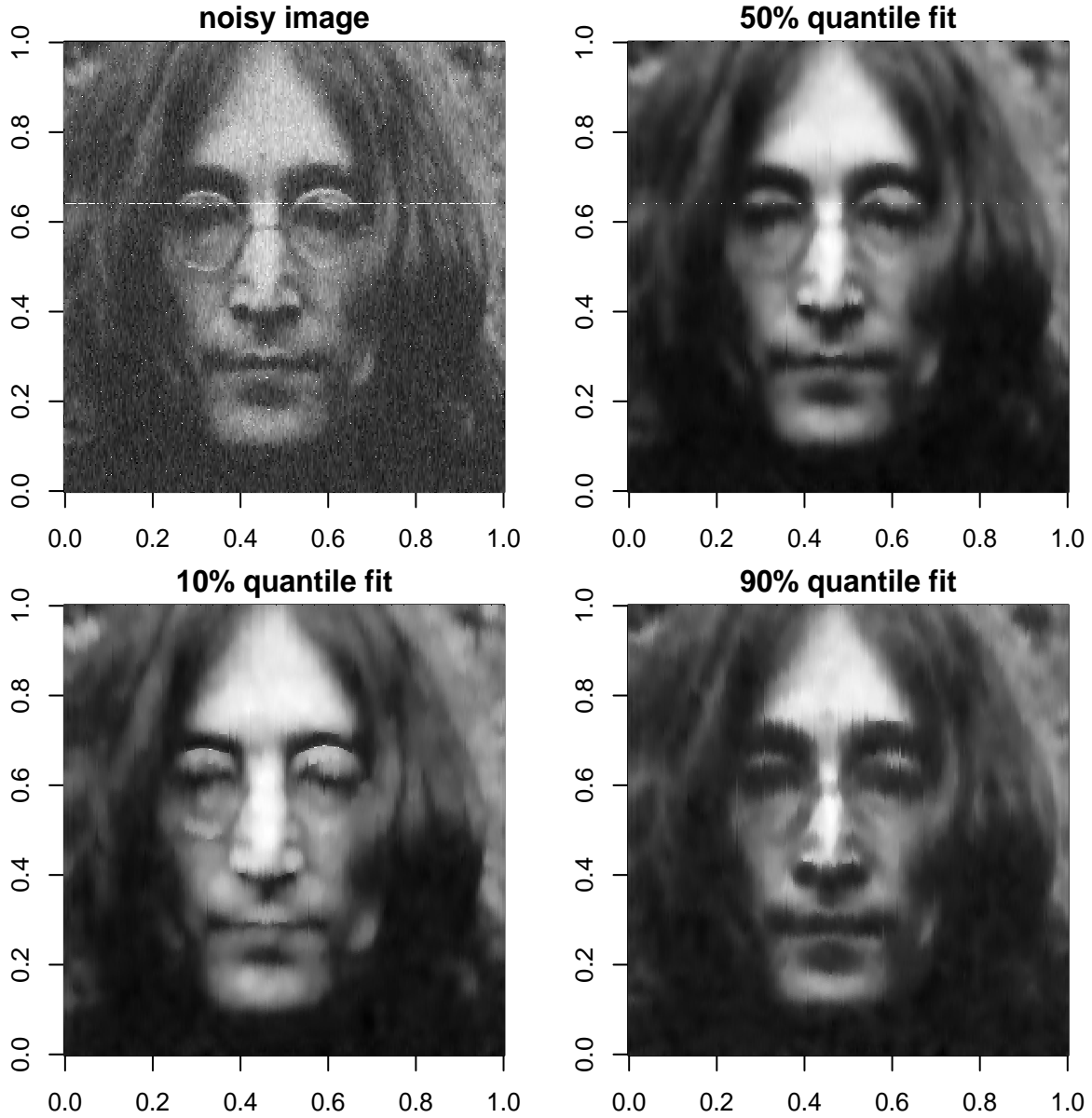
19

Figure 6: Wavelet quantile smoothing for noisy Lennon image.

estimates with $\tau = (0.1, 0.5, 0.9)$ are displayed in Figure 6.

7   Concluding Remarks

A vital component of our work is the concept of pseudo data, which links nonlinear quantile regression calculations with simple and fast least-squares type computations. It also nat-

urally suggests an algorithm, the ES-algorithm, that provides a fast and unified approach for computing nonparametric quantile regression estimates. This ES-algorithm switches between two relatively straightforward steps: the E-Step evaluates the so-called empirical pseudo data while the S-Step smoothes them. It also allows a convenient way for choosing the smoothing parameter. It has been applied successfully to obtain quantile regression estimates with both univariate and bivariate smoothing splines and wavelet estimators. Given its simplicity, speed and promising empirical performance, the ES-algorithm is a viable tool for many practitioners who need to perform nonparametric quantile regression.

Supplementary Materials

The following two items are in the supplementary materials archive of the journal website:

- **appendix** that contains technical details and proofs (appendix.pdf), and

- **R codes** that implement the proposed methodology (Rcodes.zip).

Acknowledgement

References

Abramovich, F. and Benjamin, Y. (1995) Thresholding of wavelet coefficient as multiple hypotheses testing procedure. *Wavelets and Statistics*, Springer, New York, Lecture notes in Statistics **103**, 5–14.

Antoniadis, A. and Fan, J. (2001) Regularization of wavelet approximations. *Journal of the American Statistical Association* **96**, 939–962.

Bhattacharya, P.K. and Gangopadhyay, A.K. (1990) Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics* **18**, 1400–1415.

Cantoni, E. and Ronchetti, E. (2001) Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, **11**, 141–146.

Chaudhuri, P. (1991a) Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis* **39**, 246-269.

Chaudhuri, P. (1991b) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* **19**, 760-777.

Chaudhuri, P. and Loh, W. Y. (2002) Nonparametric estimation of conditional quantiles using quantile regression trees. *Bernoulli*, **8**, 561-576.

Cooley, D., Nychka, D. and Naveau, P. (2005) Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**, 824–840.

Cox, D.D. (1983) Asymptotics for M-type smoothing splines. *The Annals of Statistics*, **11**, 530–551.

Friedman, J.H. (1991) Multivariate adaptive regression splines. *Annals of Statistics* **19**, 1–67.

Green, P.J. and Silverman, B.W. (1994) *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, London.

Gu, C. (1992) Cross-Validating Non-Gaussian Data. *Journal of Computational and Graphical Statistics* **2**, 169–179.

He, X., Ng, P. and Portnoy, S. (1999) Bivariate quantile smoothing splines. *Journal of the Royal Statistical Society Series* B **60**, 537–550.

Healy, M.J.R., Rasbash, J. and Yang, M. (1988) Distribution-free estimation of age-related centiles. *Annals of Human Biology* **15**, 17-22.

Hendricks, W. and Koenker, R. (1992) Hierarchical spline model for conditional quantiles and the demand for electricity. *Journal of the American Statistical Society* **87**, 58–68.

Johnstone, I.M. and Silverman, B.W. (2005) Empirical Bayes selection of wavelet thresholds. *Annals of Statistics* **33**, 1700-1752.

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

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

Koenker, R. and D'Orey, V. (1987) Computing regression quantiles. *Applied Statistics* **36**, 383–393.

Koenker, R., Ng, P. and Portnoy, S. (1994) Quantile smoothing splines. *Biometrika* **81**, 673–680.

Lejeune, M.G. and Sarda, P. (1988) Quantile regression: a nonparametric approach. *Computational Statistics and Data Analysis* **6**, 229–239.

Li, Y., Liu, Y. and Zhu, J. (2007) Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association* **102**, 255–268.

Nason, G.P. (1998) *WaveThresh3 Software.* Department of Mathematics, University of Bristol, Bristol, UK.

Oh, H.-S., Nychka, D.W. and Lee, T.C.M. (2007) The Role of Pseudo Data for Robust Smoothing with Application to Wavelet Regression. *Biometrika* **94**, 893–904.

Nychka, D.W., Gray, G., Haaland, P., Martin, D. and O'Connell, M. (1995) A nonparametric regression approach to syringe grading for quality improvement. *Journal of the American Statistical Association* **90**, 1171–1178.

Truong, Y.K.N. (1989) Asymptotic properties of kernel estimators based on local medians. *Annals of Statistics* **17**, 606–617.

Yu, K. (1999) Smoothing regression quantile by combining k-NN with local linear fitting. *Statistica Sinica* **9**, 759–771.

Yu, K. and Jones, M.C. (1998) Local linear regression quantile estimation. *Journal of the American Statistical Association* **93**, 228–238.

Yu, K., Lu, Z. and Stander, J. (2003) Quantile regression: application and current research areas. *The Statistician* **52**, 331–350.

Yuan, M (2006) GACV for Quantile Smoothing Splines. *Computational Statistics and Data Analysis* **50**, 813–829.