Variable Selection for Nonparametric Quantile Regression via Smoothing Spline AN OVA

Chen-Yen Lin[a], **Howard Bondell**[b,*], **Hao Helen Zhang**[c], and **Hui Zou**[d]

[a] Eli Lilly and Company, IN 46285

[b] Department of Statistics, North Carolina State University, NC 27695-8203.

[c] Department of Mathematics, University of Arizona, AZ 85721-0089

[d] School of Statistics, University of Minnesota, MN 55455

Abstract

Quantile regression provides a more thorough view of the effect of covariates on a response. Nonparametric quantile regression has become a viable alternative to avoid restrictive parametric assumption. The problem of variable selection for quantile regression is challenging, since important variables can influence various quantiles in different ways. We tackle the problem via regularization in the context of smoothing spline ANOVA models. The proposed sparse nonparametric quantile regression (SNQR) can identify important variables and provide flexible estimates for quantiles. Our numerical study suggests the promising performance of the new procedure in variable selection and function estimation. Supplementary materials for this article are available online.

Keywords

Model Selection; COSSO; Reproducing Kernel Hilbert Space; Kernel Quantile Regression

1. Introduction

Quantile regression, as a complement to classical least squares regression, provides a more comprehensive framework to study how covariates influence not only the location but the entire conditional distribution (Koenker, 2005). In quantile regression problems, the primary interest is to establish a regression function to reveal how the $100\,\tau\%$ quantile of the response y depends on a set of covariates $\boldsymbol{x} = \{x^{(1)}, \dots, x^{(d)}\}$. A parametric form of regression function is often assumed for convenience of interpretation and lower computation cost. While a linear regression function is studied in Koenker & Bassett (1978) and numerous follow-up studies, Procházka (1988) explored nonlinear regression; see Koenker & Hallock (2001) and Koenker (2005) for a comprehensive overview.

*bondell@stat.ncsu.edu.

As much as the parametric assumption enjoys a simple model structure and lower implementation cost, it is not flexible enough and hence carries the risk of model misidentifications for complex problems. For a single predictor model, Koenker et al. (1994) pioneered nonparametric quantile regression in spline models, in which the quantile function can be found via solving the minimization problem

$$\min_{f \in \mathscr{F}} \sum_{i=1}^{n} \rho_\tau \left(y_i - f\left(x_i \right) \right) + \lambda V \left(f' \right), \quad (1)$$

where $\rho_\tau(t) = t[\tau - I(t < 0)]$, $\tau \in (0, 1)$, is the so-called "check function" of Koenker & Bassett (1978), λ is a smoothing parameter and $V(f')$ is the total variation of the derivative of f. Koenker et al. (1994) showed that the minimizer is a linear spline with knots at the design points x_i, $i = 1, \ldots, n$, provided that the space $\mathscr{F}$ is an expanded second-order Sobolev space defined as

$$\mathscr{F} = \left\{ f : f\left(x \right) - a_0 + a_1 x + \int_0^1 \left(x - y \right) + d\mu \left(y \right), \quad V\left(\mu \right) < \infty, \ a_i \in \mathbb{R}, \ i = 0, \ 1 \right\}, \quad (2)$$

where μ is a measure with finite total variation. Bloomfield & Steiger (1983) and Nachka et al. (1995) considered a similar problem to (1), but with a different roughness penalty on the function

$$\min_{f \in \mathscr{F}} \sum_{i=1}^{n} \rho_\tau \left(y_i - f\left(x_i \right) \right) + \lambda \int \left[f''\left(x \right) \right]^2 dx. \quad (3)$$

The minimizer of (3) over a second-order Sobolev space is a natural cubic spline with all design points as knots. Bosch et al. (1995) proposed an interior point algorithm which is guaranteed to converge to the minimizer.

For multi-dimensional feature space, He et al. (1998) proposed a bivariate quantile smoothing spline and He & Ng (1999) generalized the idea to multiple covariates using an ANOVA-type decomposition. Li et al. (2007) proposed a more general kernel quantile regression (KQR) method that penalizes the roughness of the function estimator using its squared functional norm in a reproducing kernel Hilbert space (RKHS). More specifically, the KQR solves the regularization problem

$$\min_{f \in \mathscr{H}_K} \sum_{i=1}^{n} \rho_\tau \left(y_i - f\left(\boldsymbol{x}_i \right) \right) + \frac{\lambda}{2} \|f\|_{\mathscr{H}_K}^2, \quad (4)$$

where $\mathscr{H}_K$ is a RKHS and $\| \cdot \|_{\mathscr{H}_K}$ is its functional norm. More recently, Fenske et al. (2011) proposed a boosting method to select and estimate additive quantile functions. Although not primarily designed for variable selection, their method naturally achieves variable selection if the number of boosting iterations is small enough.

Despite several existing nonparametric quantile function estimators, selecting relevant predictors in multi-dimensional data is an important yet challenging topic that has not been addressed in depth. Variable selection in quantile regression is much more difficult than that in the least squares regression. The variable selection is carried at various levels of quantiles, which amounts to identifying variables that are important for the entire distribution, rather than limited to the mean function as in the least squares regression case. This has important applications to handle heteroscedastic data. Several regularization methods were proposed (Zou & Yuan, 2008a,b; Wu & Liu, 2009) for linear quantile regression.

In the presence of multiple predictors, many nonparametric estimation procedures may suffer from the curse of dimensionality. The smoothing spline analysis of variance (SS-ANOVA) model (Wahba, 1990) provides a flexible and effective estimation framework to tackle the problem. Since some of the predictors may not be useful or redundant for prediction, variable selection is important in nonparametric regression. In the context of least squares regression, the COmponent Selection and Shrinkage Operator (COSSO) (Lin & Zhang, 2006) was proposed to perform continuous function shrinkage and estimation by penalizing the sum of RKHS norms of the components. In this paper, we adopt the COSSO-type penalty to develop a new penalized framework for joint quantile estimation and variable selection. In additive model, some nonparametric procedures for joint estimation and selection have been proposed in a basis expansion framework (Meier et al., 2009; Huang et al., 2010). The choice between the reproducing kernel Hilbert space (RKHS) and basis expansions mirrors that between smoothing splines and regression splines in the least squares setup. They are both popular in practice and have their own desired properties. The RKHS approach provides a flexible and rigorous mathematical framework for multivariate function estimation, and it has wide applications in statistics and machine learning. Furthermore, the COSSO and adaptive COSSO methods offer a new regularization framework to nonparametric variable selection and have promising performance in the standard least squares context. It would be interesting to study them in the context of more complex problems such as quantile regression.

The remainder of the article is organized as follows. Section 2 reviews the SS-ANOVA models and introduces the new estimator. An iterative computation algorithm is given in Section 3, along with parameter tuning procedure. Extensive empirical studies, including both the homogeneous and heterogenous errors are given in Section 4. Two real example analysis results are presented in Section 5. We conclude our findings in Section 6.

2. Formulation

2.1. Smoothing Spline ANOVA

In the framework of smoothing spline ANOVA (SS-ANOVA), it is assumed that a multivariate function $f(\boldsymbol{x}) = f(x^{(1)}, \ldots, x^{(d)})$ has the ANOVA decomposition

$$f\left(\boldsymbol{x}\right) = b + \sum_{j=1}^{d} f_j\left(x^{(j)}\right) + \sum_{j<k} f_{j,k}\left(x^{(j)}, x^{(k)}\right) + \cdots, \quad (5)$$

where b is a constant, f_j's are the main effects and $f_{j,k}$'s are the two-way interactions, and so on. We estimate each of the main effects in a RKHS denoted by $\mathscr{H}_j = \{1\} \oplus \bar{\mathscr{H}}_j$ whereas the interactions are estimated in a tensor product spaces of the corresponding univariate function spaces. A popular choice of $\mathscr{H}_j$ is the second-order Sobolev space $\mathscr{S}^2[0, 1] = \{g : g, g'$ are absolutely continuous and $g'' \in \mathscr{L}^2[0, 1]\}$. When endowed with the norm

$$\|g\|^2 = \left\{\int_0^1 g\left(x\right) dx\right\}^2 + \left\{\int_0^1 g'\left(x\right) dx\right\}^2 + \int_0^1 \left\{g''\left(x\right)\right\}^2 dx, \quad (6)$$

the second-order Sobolev space is a RKHS with reproducing kernel

$$R\left(x, y\right) = 1 + k_1\left(x\right) k_1\left(y\right) + k_2\left(x\right) k_2\left(y\right) - k_4\left(|x - y|\right), \quad (7)$$

where $k_1\left(x\right) = x - \frac{1}{2}$, $k_2\left(x\right) = \frac{1}{2}\left[k_1^2\left(x\right) - \frac{1}{12}\right]$ and $k_4\left(x\right) = \frac{1}{24}\left[k_1^4\left(x\right) - \frac{1}{2}k_1^2\left(x\right) + \frac{7}{240}\right]$. When $x^{(j)}$ is a categorical variable that takes only finite distinct values, $\{1, \ldots, L\}$, we define the reproducing kernel as

$$R(s, t) = L \cdot I(s=t) - 1, \quad s, t \in \{1, \dots, L\}. \quad (8)$$

See Wahba (1990) and Gu (2002) for more details. The entire tensor-product space for estimating $f(\boldsymbol{x})$ is given by

$$\mathscr{F} = \overset{d}{\underset{j=1}{\oplus}} \mathscr{H}_j = \{1\} \oplus \sum_{j=1}^{d} \bar{\mathscr{H}}_j \oplus \sum_{j<k} \left(\bar{\mathscr{H}}_j \otimes \bar{\mathscr{H}}_k \right) \oplus \cdots. \quad (9)$$

Note that $\mathscr{F} = \otimes_{j=1}^{d} \mathscr{H}_j$ is also a RKHS, and its reproducing kernel is the sum of the individual reproducing kernels of those component spaces.

In practice, the higher-order interactions in (9) will usually be truncated for convenience in interpretation and to avoid the curse of dimensionality. A general expression for a truncated space can be written as

$$\mathscr{F} = \{1\} \oplus \left\{ \overset{q}{\underset{j=1}{\oplus}} \mathscr{F}_j \right\}, \quad (10)$$

where $\mathscr{F}_1, \dots, \mathscr{F}_q$ are q orthogonal subspaces of $\mathscr{F}$. A special case is the well-known additive model Hastie & Tibshirani (1990) with $q = d$, in which only the main effects are kept in the model, say $f(\boldsymbol{x}) = b + \sum_{j=1}^{d} f_j \left(x^{(j)} \right)$. When both main effects and two-way interaction effects are retained, the truncated space has $q = d(d + 1)/2$. The goal of this paper is variable selection, so we mainly focus on the additive model. However, the propose method can be extended to a general model (5) if one is interested in identifying both important main effects and as well as higher-order interaction terms.

A typical method for estimating nonparametric quantile function is through solving the regularization problem

$$\min_{f \in \mathscr{F}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(y_i - f(\boldsymbol{x}_i) \right) + \lambda J(f), \quad (11)$$

where $J(\cdot)$ is a penalty functional. A smoothing spline estimate uses the penalty $J(f) = \sum_{j=1}^{d} \theta_j^{-1} \| P^j f \|^2$, with θ_j's as smoothing parameters. The estimation in (11) involves multiple tuning parameters $\theta^1, \cdots, \theta_d$, which needs to be selected properly for a good estimation results. The parameter λ is usually included and fixed at some convenient value for computational stability in practice.

2.2 New Methodology: Sparse Nonparametric Quantile Regression (SNQR)

To achieve joint variable selection and function estimation in nonparametric quantile regression, we consider the following regularization problem

$$\min_{f \in \mathscr{F}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(y_i - f(\boldsymbol{x}_i) \right) + \lambda \sum_{j=1}^{d} w_j \| P^j f \|, \quad (12)$$

where P^j's are orthogonal projection operators that project f onto $\mathscr{F}_j$ respectively and w_j's are known weights. We will refer to the solution to (12) as the sparse nonparametric quantile regression (SNQR) estimator.

The problem in (12) provides a flexible modeling framework that includes existing methods as special cases. For instance, it reduces to the L_1-norm quantile regression (Li & Zhu, 2008) in linear models. Let $f(\boldsymbol{x}) = b + \sum_{j=1}^{d} \beta_j x^{(j)}$ and consider a linear function space $\mathscr{F} = \{1\} \oplus \left\{ \oplus_{j=1}^{d} \left\{ x^{(j)} - 1/2 \right\} \right\}$ with inner product $\langle f, g \rangle = \int f g$, then the RKHS norm penalty $\|P^j f\|$ is proportional to $|\beta_j|$. Moreover, we allow each functional component to be penalized differently depending on its associated weight w_j. In principle, smaller weights are assigned to important function components while larger weights are assigned to less important components. This is in the same spirit of the adaptive LASSO (Zou, 2006) and adaptive COSSO (Storlie et al., 2011). We propose to construct the weights from the data adaptively. For each component f_j, its L_2-norm $\|f_j(x^{(j)})\|_{L_2} = \{[f_j(x^{(j)})]^2 dF_{X^{(j)}}(x)\}^{1/2}$, $F_{X^{(j)}}(\cdot)$ is the distribution function of $x^{(j)}$, is a natural measure to quantify the importance of the j-th functional component. In practice, given a reasonable initial estimator, $\tilde{f}$, we propose to construct the weights by its inverse empirical L_2-norm, i.e.

$$w_j^{-1} = \|P^j \tilde{f}\|_{n, L_2} = \left\{ n^{-1} \sum_{i=1}^{n} \left[P^j \tilde{f} \left(x_i^{(j)} \right) \right]^2 \right\}^{1/2}, \quad j = 1, \ldots, d. \quad (13)$$

We recommend use the solution of the KQR as an initial estimate $\tilde{f}$.

Due to the fact that both the check loss and the penalty functional $J(f)$ are continuous and convex in f, the existence of the minimizer of (12) is guaranteed as stated in the following Theorem.

Theorem 1

Let $\mathscr{F}$ be an RKHS of functions with the decomposition (10), *then there exists a minimizer to* (12) *in $\mathscr{F}$.*

Directly minimizing (12) can be a daunting task as searching over the infinite dimensional space $\mathscr{F}$ for a minimizer is practically infeasible. Analogous to the smoothing spline models, the following Theorem shows that the minimizer to (12) lies in a finite dimensional space. This important result assures the feasibility of computation.

Theorem 2

Representer Theorem: Let the minimizer of (12) *be* $\hat{f} = \hat{b} + \sum_{j=1}^{d} \hat{f}_j$ *with* $\hat{f}_j \in \mathscr{F}_j$, *then* $\hat{f}_j \in span \left\{ R_{\mathscr{F}_j} \left(x_i^{(j)}, \cdot \right), i = 1, \ldots, n \right\}$ *where* $R_{\mathscr{F}_j}(\cdot, \cdot)$ *is the reproducing kernel of $\mathscr{F}_j$.*

Due to the space restriction, we relegate the technical proofs of both Theorems to the Appendix, which are given in online supplementary materials.

3. Computational Algorithm

To further facilitate the computation, we present an equivalent formulation of (12). By introducing non-negative slack variables θ_j, $j = 1, \ldots, d$, and using Lemma 2 in Lin &

Zhang (2006), it is easy to show that minimizing (12) is equivalent to solving the following optimization problem

$$\min \frac{1}{n}\sum_{i=1}^{n}\rho_{\tau}\left(y_i - f\left(\boldsymbol{x}_i\right)\right) + \lambda_0 \sum_{j=1}^{d} w_j^2 \theta_j^{-1} \|P^j f\|^2, \quad \text{s.t.} \sum_{j=1}^{d}\theta_j \leq M, \quad \theta_j \geq 0, \forall j, \quad (14)$$

where λ_0 and M are both smoothing parameters. The roles of the slack variables θ_j's are different from those in the standard smoothing spline model. The slack variables θ_j's allow us to recover the sparse structure since $\theta_j = 0$ if and only if $\|P^j f\| = 0$ (Lin & Zhang, 2006). Moreover, when θ_j's are given, the penalty function in (14) reduces to that in the traditional smoothing spline and thus by the representer theorem of Kimeldorf & Wahba (1971), the minimizer of (14) has a finite representation form

$$f\left(\boldsymbol{x}\right) = b + \sum_{i=1}^{n}\left(\sum_{j=1}^{d}\theta_j w_j^{-2} R_{\mathscr{F}_j}\left(x_i^{(j)}\right), x^{(j)}\right) c_i, \quad (15)$$

where $\boldsymbol{c} = (c_1, \ldots, c_n) \in \mathbb{R}_n$ and $b \in \mathbb{R}$.

Let $\boldsymbol{R}_j = \left\{ R_{\mathscr{F}_j}\left(x_i^{(j)}, x_{i'}^{(j)}\right) \right\}_{i,i'=1}^{n}$ be an $n \times n$ matrix and 1_n be a column vector of n ones. When evaluated the minimizer at the design points, we write $\boldsymbol{f} = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n))^T$ as $\boldsymbol{f} = b 1_n + \left(\sum_{j=1}^{d} w_j^{-2}\theta_j \boldsymbol{R}_j\right) \boldsymbol{c}$ and define $\|\boldsymbol{v}\|_{C_\tau} = n^{-1}\sum_{i=1}^{n}\rho_{\tau}\left(v_i\right)$ for any vector $\boldsymbol{v}$ of length n. With these notations, the objective function in (14) can be expressed as

$$\min \left\| \boldsymbol{y} - b 1_n - \sum_{j=1}^{d}\theta_j w_j^{-2}\boldsymbol{R}_j \boldsymbol{c} \right\|_{C_\tau} + \lambda_0 \boldsymbol{c}^T \left(\sum_{j=1}^{d}\theta_j w_j^{-2}\boldsymbol{R}_j\right) \boldsymbol{c}, \text{ s.t.} \sum_{j=1}^{d}\theta_j \leq M, \quad \theta_j \geq 0, \quad \forall j. \quad (16)$$

3.1. Iterative Optimization Algorithm

It is possible to simultaneously minimize the objective function in (16) with respect to all unknown parameters $(b, \boldsymbol{c}^T; \boldsymbol{\theta}^T)^T$, but the programming effort can be daunting. Alternatively, we can separate the parameters into two parts, $\boldsymbol{\theta}$ and $(b, \boldsymbol{c}^T)^T$, and then iteratively solve two optimization problems in turn, with respect to $\boldsymbol{\theta}$ and $(b, \boldsymbol{c}^T)^T$. Consequently, we suggest the following iterative algorithm:

1. Fix $\boldsymbol{\theta}$ solve $(b, \boldsymbol{c}^T)^T$

$$\min_{b, \boldsymbol{c}} \left\| \boldsymbol{y} - b 1_n - \sum_{j=1}^{d}\theta_j w_j^{-2}\boldsymbol{R}_j \boldsymbol{c} \right\|_{C_\tau} + \lambda_0 \boldsymbol{c}^T \left(\sum_{j=1}^{d}\theta_j w_j^{-2}\boldsymbol{R}_j\right) \boldsymbol{c}. \quad (17)$$

2. Fix $(b, \boldsymbol{c}^T)^T$, solve $\boldsymbol{\theta}$

$$\min \|\boldsymbol{y}^* - \boldsymbol{G}\theta\|_{C_\tau} + \lambda_0 \boldsymbol{c}^T \boldsymbol{G}\theta, \quad \text{s.t.} \sum_{j=1}^{d}\theta_j \leq M, \theta_j \geq 0, \quad \forall j, \quad (18)$$

where $\boldsymbol{y}^* = \boldsymbol{y} - b 1_n$ and $\boldsymbol{G} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_d)$ is an $n \times d$ matrix with columns $\boldsymbol{g}_j = w_j^{-2}\boldsymbol{R}_j\boldsymbol{c}$.

The optimization problems in (17) and (18) can be cast into quadratic and linear programming problems, respectively. We defer all the derivations to the Appendix available from online supplementary materials. Based on our experience, the algorithm converges quickly in a few steps. We have noted that the one-step solution often provides a satisfactory approximation to the solution. As a result, we advocate the use of one-step update in practice.

An important connection between the SNQR and the KQR can be unraveled by realizing that the objective function in (17) is exactly the same as that in the KQR. This connection suggests that the SNQR method reduces to the KQR when θ is known. The optimization problem for estimating θ essentially imposes the non-negative garrote (Breiman, 1995) type shrinkage on θ's, and hence achieves variable selection by shrinking some of θ_j's to zero.

3.2. Parameter Tuning

Like any other penalized regression procedure, the performance of the SNQR critically depends on properly-tuned smoothing parameters. A reasonable parameter is usually the one that minimizes some generalized error or information criterion. In the quantile regression literature, one commonly used criterion is the Schwarz information criterion (SIC) (Schwarz, 1978; Koenker et al., 1994)

$$\log\left(\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_i - \hat{f}(\boldsymbol{x}_i)\right)\right) + \frac{\log n}{2n}df, \quad (19)$$

where df is a measure of complexity of the fitted model. Various authors (Koenker, 2005; Yuan, 2006; Li et al., 2007) have suggested using the number of zero residuals as an estimate of effective degrees of freedom (EDF). In our numerical experiments, we observed that the estimated EDF fluctuates a lot when smoothing parameter changes, which may make the tuning process unstable. Instead of using the number of zero residuals as an EDF estimate, we propose a bootstrap method to estimate the EDF and defer the technical detail to the appendix.

In addition to the SIC, another popular criterion to choose the smoothing parameter is k-fold cross validation (CV), which has been widely applied to various regression and classification problems and usually gives competitive performance. Based on our numerous experiments, the tuning with CV overall produces better prediction performance than that with SIC for the proposed nonparametric quantile estimation. Therefore, we recommend the use of k-fold CV for parameter tuning in practice. In this paper, $k = 5$ is used in both simulated and real data analysis.

In the following, we summarize the complete algorithm for the SNQR method, including both model fitting and parameter tuning steps.

Step 1. Initialization. Set $\theta_j = 1, \forall j$.

Step 2. For each of the grid points of λ_0, solve (17) for $(b, \boldsymbol{c}^T)^T$, and record the SIC score or CV error. Choose the best λ_0 that minimizes the SIC or CV error, then fix it in later steps.

Step 3. For each of the grid points of M, solve (16) for $(\boldsymbol{\theta}^T, b, \boldsymbol{c}^T)^T$ using the aforementioned iterative optimization algorithm. Record the SIC score or CV error at each grid point and choose the best M that minimizes either SIC score or CV error.

Step 4. Solve (16) using the chosen λ_0 and M pair, on the full data. Note that this is already done if tuning was based on SIC.

Since the tuning procedure described above does not cover all possible pairs of (λ_0, M), we suggest to do the following to enhance the tuning. After Step 3, say, we obtain the optimal pair (λ_0^*, M^*). Then we focus on a narrowed and more focused region, the neighborhood of (λ_0^*, M^*) and apply Step2 and 3 again. The optimal parameters determined at this refined step, say, (λ_0^{**}, M^{**}) will be used as the final selection. Our simulation study also confirms that this refined tuning procedure can improve the prediction and selection performance substantially.

4. Numerical Results

In this section we present the empirical performance of the SNQR procedure using simulated data. For the experiment design, we use the following functions as building

blocks: $g_1(t) = t$; $g_2(t) = (2t - 1)^2$; $g_3(t) = \dfrac{\sin(2\pi t)}{2 - \sin(2\pi t)}$ and $g_4(t) = 0.1 \sin(2\pi t) + 0.2 \cos(2\pi t) + 0.3 \sin^2(2\pi t) + 0.4 \cos^3(2\pi t) + 0.5 \sin^3(2\pi t)$.

We evaluate the performance from two aspects, prediction accuracy and model selection. The integrated absolute error (IAE), defined as $\mathrm{IAE} = E|\hat{f}(\mathbf{X}) - f(\mathbf{X})|$, is used to assess prediction accuracy, where the expectation is evaluated by a monte carlo integration with 10,000 test points generated from the same distribution as the training data. In terms of model selection, we first denote $\widehat{\mathcal{M}} = \left\{ j : \hat{\theta}_j \neq 0 \right\}$ and $\mathcal{M}_0 = \left\{ j : \|P^j f\| > 0 \right\}$ as the selected model and true model, respectively, and $|\mathcal{M}|$ as the cardinality of the set $\mathcal{M}$. Then we compute four statistics for assessing selection accuracy: correct selection, $I\left(\widehat{\mathcal{M}} = \mathcal{M}_0\right)$, type I error rate, $\dfrac{|\widehat{\mathcal{M}} \cap \mathcal{M}_0^c|}{d - |\mathcal{M}_0|}$, power, $\dfrac{|\widehat{\mathcal{M}} \cap \mathcal{M}_0|}{|\mathcal{M}_0|}$, and model size, $|\widehat{\mathcal{M}}|$. For the purpose of comparison, we also include the solution of the KQR fitted with only relevant predictors based on 5-fold cross validation tuning. This method will later be referred to as the Oracle estimator. The Oracle estimator provides a benchmark for the best possible estimation risk if the important variables were known. We also include the KQR and boosting QR (Fenske et al., 2011) for comparison, to illustrate the superior performance of our method to the existing methods.

Another property that we would like to study is the role of the adaptive weights in the performance of the SNQR procedure. Without any a priori knowledge on the importance of each predictor, we can set all $w_j = 1$ in (16) and proceed to solve the objective function. This method will be referred to as an unweighted method. For the weighted procedure, we use KQR as an initial estimate, $\tilde{f}$, to produce an adaptive weight.

Two different quantile values $\tau = \{0.2, 0.5\}$, are used throughout the simulation. For each of the following examples, we repeat 100 times and report the average summary statistics and their associated standard errors.

4.1. Computation Cost

Before comparing different methods, we first study the computation cost for the SNQR method in this section. In particular, we assess the computation cost using the average elapsed CPU times for solving (16) with a fixed (λ_0, M) over 200 replicates. We first generate the predictors $x^{(j)}, j = 1, \ldots, d$, independently from $U(0, 1)$ and then take n observations from the model

$$y_i = 5g_1\left(x_i^{(1)}\right) + 3g_2\left(x_i^{(2)}\right) + 4g_3\left(x_i^{(7)}\right) + 6g_4\left(x_i^{(10)}\right) + \varepsilon_i, \quad i=1,\ldots,n, \quad (20)$$

where ε_i are independently drawn from $t(3)$. We consider multiple combinations of sample sizes n and dimension d. All the computations are done on a desktop PC with an Intel i7-2600K CPU and 12GB of memory. The average CPU time is presented in Table 1. Table 1 shows that the proposed algorithm is pretty fast in general. For example, when $n = 200$; $d = 40$, it takes in average about 0.3 second to solve the optimization problem with a fixed tuning parameter. The sample size n is one major factor influencing the computation speed. We observe that the computing time varies little when p increases from 10 to 40 for a fixed n, but increases substantially when n increases from $n = 100$ to $n = 300$ for a fixed d.

4.2. Homoskedastic Error Model

We first consider a model with the response generated from a location family in (20) and with the homoskedastic error. The $100\tau\%$ quantile function is then given by

$$Q_y\left(\tau|\boldsymbol{x}\right) = 5g_1\left(x^{(1)}\right) + 3g_2\left(x^{(2)}\right) + 4g_3\left(x^{(7)}\right) + 6g_4\left(x^{(10)}\right) + F_\varepsilon^{-1}\left(\tau\right), \quad (21)$$

where $F_\varepsilon(\cdot)$ is the distribution function of ε. The covariates $x^{(j)}$, $j = 1, \ldots, 40$, follow $U(0, 1)$ distribution and there is autoregressive (AR) type of dependency among them, i.e., with the pairwise correlation cor $(x^{(j)}, x^{(k)}) = \rho^{|j-k|}$, $\forall j \quad k$. We consider two levels of dependency with $\rho = \{0, 0.7\}$. The sample size $n = 200$. Tables 2 summarizes the performance of five competing procedures: KQR, boosting QR, SNQR, adaptive SNQR, and the Oracle estimator. Another desired property of the SNQR procedures is their robustness property. Although the least squares regression and quantile regression are not generally comparable, the conditional median and mean functions coincide in this example, making the comparison between them justifiable. Therefore we also incorporate two sparse least squares nonparametric regression models, COSSO (Lin & Zhang, 2006) and adaptive COSSO (Storlie et al., 2011), to estimate the conditional mean function and compare with the corresponding SNQR estimate.

From Tables 2, we observe that the adaptive SNQR outperforms existing procedures under both independent and dependent cases. In terms of prediction accuracy, the adaptive SNQR has the smallest IAE, which is also quite close to that of the Oracle, and it is followed by SNQR, boosting QR, and the KQR is the worst. Furthermore, the tuning with 5-fold CV produces better results than the tuning with SIC most of the times. It is clear that the KQR suffers considerably from the presence noise variables. Although the boosting QR has a model size consistently greater than 20, during the boosting process, the four important predictors are selected to update the solution most frequently, whereas the noise variables are selected with a much smaller frequency. Fenske et al. (2011) found similar results and recommended using the selection frequency as a guidance for variable selection. With regard to variable selection, the SNQR procedures effectively identifies important variables, particularly when using SIC as a tuning procedure. Throughout all simulation scenarios, the SNQR have a well-controlled Type I error, and the power is no less than 90%. With regard to estimation performance, relative to the Oracle estimator, the proposed estimator is more than 90% and about 70-80% as efficient in terms of IAE for the independent and dependent cases, respectively.

The conditional mean function estimators, COSSO and adaptive COSSO, give comparable model selection. However, their estimation performance is seriously affected by the heavy tail of the error distribution. Even with adaptive weight, COSSO still performs worse than the SNQR method. In addition, the standard errors are almost 10 times larger than the other

median estimators, implying our SNQR method enjoys the robust property for median estimation.

Figure 1 gives a graphical illustration for the pointwise confidence bands for the fitted curves using $\tau = 0.2$. For comparison, the estimated functions by the Oracle are also depicted. Since we apply each procedure to 100 simulated datasets, the pointwise confidence band is constructed by using the 5% and 95% percentiles. From Figure 1, we see that the SNQR estimate the true function very well, and it performs very close to the Oracle estimator.

4.3. Heteroskedastic Error Model

To further examine the finite sample performance of the new methods, we consider generating response from a location-scale family

$$y_i = 5g_1\left(x_i^{(1)}\right) + 3g_2\left(x_i^{(2)}\right) + 4g_3\left(x_i^{(7)}\right) + 6g_4\left(x_i^{(10)}\right) + \exp\left[2g_3\left(x_i^{(12)}\right)\right]\varepsilon_i, \quad i = 1, \ldots, n, \quad (22)$$

where $\varepsilon_i \overset{i.i.d.}{\sim} \mathscr{N}(0, 1)$. In the heteroskedastic model, the $100\tau\%$ quantile function is given by

$$Q_y(\tau|\boldsymbol{x}) = 5g_1\left(x^{(1)}\right) + 3g_2\left(x^{(2)}\right) + 4g_3\left(x^{(7)}\right) + 6g_4\left(x^{(10)}\right) + \exp\left[2g_3\left(x^{(12)}\right)\right]\Phi^{-1}(\tau), \quad (23)$$

where $\varphi(\cdot)$ is the distribution function of standard normal. The covariates are generated in the same fashion as that in the previous example and we use sample size $n = 300$ in this case.

From this example, we aim to evaluate the performance of SNQR when some covariates can only be influential on a certain range of quantiles. More specifically, like the homoskedastic example, the median function in this example depends on the 1, 2, 7 and 10th predictors. However, other than the median function, $x^{(12)}$ will not only be influential but its effect becomes larger and larger toward the tails.

Tables 3 summarizes the performance of all the competing methods under comparison. We first notice that, when $\tau = 0.5$, the average model size is close to 4 as expected. However, when τ is away from 0.5, the SNQR procedures successfully identifies the additional informative predictor, $x^{(12)}$, suggesting that the new method's capability to identify all the relevant predictors that influence the distribution of the response. As for the function estimation, again, we observe that the adaptive SNQR performs the best and the KQR is the worst. However, the boosting QR provides as efficient or sometimes even more efficient estimation than the SNQR. For variable selection, the tuning with SIC usually selects a model of a smaller size and identifies the correct model with high frequency.

5. Real Data Analysis

We apply the SNQR method to two real datasets: prostate cancer data and ozone data. The prostate data is from Stamey et al. (1989), consisting of 97 patients who were about to receive a radical prostatectomy. This data was used by Tibshirani (1996) to model the mean function of the level of prostate-specific antigen on 8 clinical covariates and select relevant variables. The ozone data contains 330 observations collected in Los Angeles in 1976, and the purpose of the study is to model the relationship between the daily ozone concentration and 8 meteorological covariates. The data has been used in various studies (Buja et al., 1989; Breiman, 1995; Lin & Zhang, 2006). These two data are publicly available in R packages ElemStatLearn and cosso, respectively.

We assess the model performance using the prediction risk, $E_{\rho_\tau}(Y - f(X))$, whereas the expectation is evaluated by randomly reserving 10% of the data as testing set. We select the smoothing parameters and estimate the function using only the training set. The estimated function will then be applied on the testing set and the compute the prediction risk. The entire procedure is repeated 100 times and averaged.

Based on the result summarized in Table 4, the adaptive weights is not always beneficial in real applications. The advantage of adaptive weight is clear for the prostate data. But, the difference between the weighted and the unweighted methods are usually within reasonable error margin. Overall, we still observe that the SNQR method provides competitive and sometimes superior prediction than the existing methods.

Apart from comparing prediction error, we also apply the SNQR to the complete prostate data and summarize variable selection. An interesting comparison is that in the study of mean function, Tibshirani (1996) selected three prominent predictors, log-cancer volume, log-weight and seminal vesicle invasion. These three predictors are also selected by the SNQR when we consider the median. However, in the 20% quantile, gleason score shows up as an additional predictor. Meanwhile, in the 80% quantile, only log-cancer volume and seminal vesicle invasion are chosen.

6. Conclusions

We propose a new regularization method that simultaneously selects important predictors and estimate the conditional quantile function nonparametrically. Our method is implemented in the R package cosso, which is available online supplementary materials and the Comprehensive R Archive Network. The SNQR method conquers the limitation of selecting only predictors that influence the conditional mean in least squares regression, facilitating the analysis of the full conditional distribution. The proposed method also includes the L_1-norm quantile regression and the KQR as special cases. In a simulation study and real data analysis, our method provides satisfactory model fitting and great potential for selecting important predictors.

The sample size and the number of predictors we consider in both simulation and real data are moderate. As demonstrated in Section 4.1, the computation bottleneck in our algorithm is the sample size, which affects both linear programming and quadratic programming. In our experience, the computation effort becomes formidable when the sample size is greater than 500. On the other hand, the number of predictors only affects the linear programming, and hence has relatively minor impact on the computation difficulty. In addition, the linear programming can be scaled up to a much higher dimension if using efficient numerical optimizers. In ultra-high dimensional feature space, Fan et al. (2011) recently proposed a screening procedure for nonparametric regression model. Further study can work toward incorporating a suitable screening procedure as a first step and then apply our proposed method at the second in a ultra-high dimensional feature space.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Bloomfield, P.; Steiger, W. Least absolute deviations: Theory, Applications and Algorithms. Birkh user-Verlag; Boston: 1983.

Bosch R, Ye Y, Woodworth G. A convergent algorithm for qunatile regression with smoothing splines. Computational Statistics and Data Analysis. 1995; 19:613–630.

Breiman L. Better subset selection using the nonnegative garrote. Technometrics. 1995; 37:373–384.

Buja A, Hastie T, Tibshirani R. Linear smoothers and additive models (with discussion). Annals of Statistics. 1989; 17:453–555. doi:10.1214/aos/1176347115.

Fan J, Feng Y, Song R. Nonparametric independence screening in sparse ultra-high dimensional additive model. Journal of the American Statistical Association. 2011; 106:544–557. doi:10.1198/jasa.2011.tm09779. [PubMed: 22279246]

Fenske N, Kneib T, Hothorn T. Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. Journal of the American Statistical Association. 2011; 106:494–510. doi:10.1198/jasa. 2011.ap09272.

Gu, C. Smoothing Spline ANOVA Models. Springler-Verlag; New York: 2002.

Hastie, T.; Tibshirani, R. Generalized Additive Models. Chapman and Hall; London: 1990.

He X, Ng P. Quantile splines with several covariates. Journal of Statistical Planning and Inference. 1999; 75:343–352. doi: 10.1111/1467-9868.00138.

He X, Ng P, Portnoy S. Bivariate quantile smoothing splines. Journal of the Royal Statistical Society, Ser. B. 1998; 60:537–550.

Huang J, Horowitz JL, Wei F. Variable selection in nonparametric additive models. Annals of Statistics. 2010; 38:2282–2313. doi:10.1214/09-AC)S781. [PubMed: 21127739]

Kimeldorf G, Wahba G. Some results on tchebycheffian spline functions. Journal of Mathematical Analysis and Applications. 1971; 33:82–95.

Koenker, R. Quantile Regression. Cambridge University Press; New York: 2005.

Koenker R, Bassett G. Regression quantiles. Econometrica. 1978; 46:33–50.

Koenker R, Hallock K. Quantile regression. Journal of Economic Perspectives. 2001; 15:143–156.

Koenker R, Ng P, Portnoy S. Quantile smoothing splines. Biometrika. 1994; 81:673–680. doi: 10.2307/2337070.

Li Y, Liu Y, Zhu J. Quantile regression in reproducing kernel Hilbert spaces. Journal of the American Statistical Association. 2007; 477:255–267. doi:10.1198/016214506000000979.

Li Y, Zhu J. L_1-norm quantile regression. Journal of Computational and Graphical Statistics. 2008; 17:163–185. doi: 10.1198/106186008X289155.

Lin Y, Zhang HH. Component selection and smoothing in multivariate nonparametric regression. Annals of Statistics. 2006; 34:2272–2297. doi: 10.1214/009053606000000722.

Meier L, Van De Geer S, Bühlmann P. High-dimensional additive modeling. Annals of Statistics. 2009; 37:3779–3821. doi:10.1214/09-AC)S692.

Nachka D, Gray G, Haaland P, Martin D, O'Connell M. A nonparametric regression approach to syringe grading for quality improvement. Journal of the American Statistical Association. 1995; 90:1171–1178. doi: 10.2307/2291509.

Procházka B. Regression quantiles and trimmed least squares estimator in the nonlinear regression model. Computational Statistics and Data Analysis. 1988; 6:385–391. doi: 10.1016/0167-9473(88)90078-3.

Schwarz G. Estimating the dimension of a model. Annals of Statistics. 1978; 6:1135–1151. doi: 10.1214/aos/1176344136.

Stamey T, Kabalin J, McNeal J, Johnstone I, Freiha F, Redwine E, Yang N. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate ii radical prostatectomy treated patients. Journal of Urology. 1989; 16:1076–1083. [PubMed: 2468795]

Storlie C, Bondell H, Reich B, Zhang HH. The adaptive cosso for nonparametric surface estimation and model selection. Statistica Sinica. 2011; 21:679–705. [PubMed: 21603586]

Tibshirani R. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Ser. B. 1996; 58:267–288.

Wahba, G. Spline Models for Observational Data. Philadelphia: SIAM: 1990.

Wu Y, Liu Y. Variable selection in quantile regression. Statistics Sinica. 2009; 19:801–817.

Yuan M. GACV for quantile smoothing splines. Computational Statistics and Data Analysis. 2006; 5:813–829. doi:10.1016/j.csda.2004.10.008.

Zou H. The adaptive Lasso and its oracle properties. Journal of the American Statistical Association. 2006; 101:1418–1429. doi: 10.1198/016214506000000735.

Zou H, Yuan M. Composite quantile regression and the oracle model selection theory. Annals of Statistics. 2008a; 36:1108–1126. doi: 10.1214/07-AOS507.

Zou H, Yuan M. Regularized simultaneous model selection in multiple quantiles regression. Computational Statistics and Data Analysis. 2008b; 52:5296–5304. doi:10.1016/j.csda. 2008.05.013.
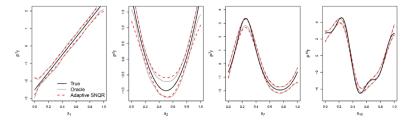
Figure 1.
The fitted function components and the associated pointwise confidence band on homoskedastic example with independent features. The dark solid line is for the true function component, the light solid line is for the Oracle estimator and the broken line is for the adaptive SNQR estimator. The online version of this plot is color.

Table 1

Elapsed CPU time (in seconds) for solving SNQR optimization problem.

	(n, d)					
τ	(100,10)	(100,40)	(200,10)	(200,40)	(300,10)	(300,40)
0.2	0.041	0.045	0.300	0.329	0.998	1.154
0.5	0.038	0.044	0.278	0.299	0.914	1.048

Table 2

Selection and estimation results for the homoskedasitc example. The standard errors are given in the parentheses.

τ	Method	Correct	Type 1 Error	Power	Model Size	IAE
			$\rho = 0.0$			
	KQR	-	-	-	-	2.223 (0.019)
	Boosting QR	0.00 (0.00)	0.67 (0.02)	1.00 (0.00)	28.13 (0.60)	1.098 (0.016)
	SNQR-5CV	0.70 (0.05)	0.01 (0.00)	0.98 (0.01)	4.23 (0.09)	0.949 (0.021)
	SNQR-SIC	0.81 (0.04)	0.02 (0.01)	0.99 (0.01)	4.58 (0.18)	0.983 (0.021)
	Adaptive SNQR-5CV	0.73 (0.05)	0.02 (0.01)	0.99 (0.01)	4.69 (0.20)	0.645 (0.016)
	Adaptive SNQR-SIC	0.82 (0.04)	0.02 (0.00)	1.00 (0.00)	4.53 (0.15)	0.667 (0.016)
	Oracle	-	-	-	-	0.634 (0.011)
0.2			$\rho = 0.7$			
	KQR	-	-	-	-	1.743 (0.015)
	Boosting QR	0.00 (0.00)	0.54 (0.02)	1.00 (0.00)	23.50 (0.73)	0.992 (0.020)
	SNQR-5CV	0.22 (0.04)	0.03 (0.00)	0.87 (0.01)	4.39 (0.15)	0.935 (0.017)
	SNQR-SIC	0.18 (0.04)	0.02 (0.01)	0.84 (0.01)	4.22 (0.21)	0.978 (0.018)
	Adaptive SNQR-5CV	0.23 (0.04)	0.03 (0.01)	0.88 (0.01)	4.52 (0.23)	0.690 (0.014)
	Adaptive SNQR-SIC	0.23 (0.04)	0.03 (0.01)	0.87 (0.01)	4.67 (0.28)	0.710 (0.014)
	Oracle	-	-	-	-	0.609 (0.011)
			$\rho = 0.0$			
	KQR	-	-	-	-	1.921 (0.017)
	Boosting QR	0.00 (0.00)	0.76 (0.02)	1.00 (0.00)	31.18 (0.60)	0.781 (0.009)
	SNQR-5CV	0.84 (0.04)	0.01 (0.00)	1.00 (0.00)	4.36 (0.15)	0.612 (0.015)
	SNQR-SIC	0.92 (0.03)	0.00 (0.00)	0.99 (0.01)	4.08 (0.13)	0.638 (0.017)
	Adaptive SNQR-5CV	0.82 (0.04)	0.01 (0.00)	1.00 (0.00)	4.39 (0.11)	0.461 (0.008)
	Adaptive SNQR-SIC	0.93 (0.03)	0.00 (0.00)	0.99 (0.00)	4.07 (0.06)	0.505 (0.012)
	COSSO	0.83 (0.04)	0.01 (0.00)	0.99 (0.01)	4.15 (0.05)	0.824 (0.025)
	ACOSSO	0.76 (0.04)	0.01 (0.00)	0.99 (0.01)	4.30 (0.08)	0.616 (0.017)
	Oracle	-	-	-	-	0.489 (0.007)
0.5			$\rho = 0.7$			
	KQR	-	-	-	-	1.512 (0.012)
	Boosting QR	0.00 (0.00)	0.50 (0.02)	1.00 (0.00)	21.97 (0.76)	0.700 (0.010)
	SNQR-5CV	0.18 (0.04)	0.04 (0.01)	0.90 (0.01)	4.93 (0.23)	0.718 (0.014)
	SNQR-SIC	0.27 (0.05)	0.03 (0.01)	0.90 (0.01)	4.83 (0.22)	0.711 (0.015)
	Adaptive SNQR-5CV	0.38 (0.05)	0.04 (0.01)	0.96 (0.01)	5.23 (0.23)	0.488 (0.011)
	Adaptive SNQR-SIC	0.60 (0.05)	0.02 (0.00)	0.95 (0.01)	4.51 (0.16)	0.481 (0.012)
	COSSO	0.35 (0.05)	0.01 (0.00)	0.88 (0.01)	3.90 (0.09)	1.436 (0.103)
	ACOSSO	0.39 (0.05)	0.01 (0.00)	0.91 (0.01)	4.13 (0.10)	0.780 (0.088)
	Oracle	-	-	-	-	0.459 (0.007)

Table 3

Selection and estimation results for the heteroskedasitc example with independent features. The standard errors are given in the parentheses.

τ	Method	Correct	Type 1 Error	Power	Model Size	IAE
				$\rho = 0.0$		
	KQR	-	-	-	-	2.422 (0.019)
	Boosting QR	0.00 (0.00)	0.75 (0.01)	1.00 (0.00)	30.72 (0.56)	1.289 (0.017)
	SNQR-5CV	0.34 (0.01)	0.01 (0.00)	0.91 (0.01)	5.04 (0.11)	1.419 (0.028)
	SNQR-SIC	0.43 (0.01)	0.01 (0.00)	0.92 (0.01)	4.92 (0.11)	1.474 (0.026)
	Adaptive SNQR-5CV	0.57 (0.02)	0.02 (0.00)	0.98 (0.01)	5.65 (0.13)	0.976 (0.026)
	Adaptive SNQR-SIC	0.65 (0.00)	0.00 (0.00)	0.94 (0.01)	4.84 (0.10)	1.154 (0.027)
	Oracle	-	-	-	-	0.663 (0.013)
0.2				$\rho = 0.7$		
	KQR	-	-	-	-	1.735 (0.018)
	Boosting QR	0.00 (0.00)	0.78 (0.02)	1.00 (0.00)	32.06 (0.55)	1.267 (0.012)
	SNQR-5CV	0.11 (0.03)	0.07 (0.01)	0.91 (0.01)	7.02 (0.28)	1.155 (0.023)
	SNQR-SIC	0.19 (0.04)	0.05 (0.01)	0.90 (0.01)	6.35 (0.27)	1.188 (0.022)
	Adaptive SNQR-5CV	0.21 (0.04)	0.08 (0.01)	0.97 (0.01)	7.46 (0.31)	0.822 (0.020)
	Adaptive SNQR-SIC	0.38 (0.05)	0.03 (0.01)	0.94 (0.01)	5.84 (0.23)	0.891 (0.022)
	Oracle	-	-	-	-	0.613 (0.013)
				$\rho = 0.0$		
	KQR	-	-	-	-	1.718 (0.015)
	Boosting QR	0.00 (0.00)	0.74 (0.01)	1.00 (0.00)	30.63 (0.41)	0.681 (0.009)
	SNQR-5CV	0.93 (0.03)	0.00 (0.00)	1.00 (0.00)	4.09 (0.05)	0.538 (0.013)
	SNQR-SIC	0.97 (0.02)	0.00 (0.00)	1.00 (0.00)	3.97 (0.02)	0.569 (0.016)
	Adaptive SNQR-5CV	0.84 (0.04)	0.01 (0.00)	1.00 (0.00)	4.26 (0.07)	0.371 (0.007)
	Adaptive SNQR-SIC	0.95 (0.02)	0.00 (0.00)	1.00 (0.00)	4.11 (0.07)	0.390 (0.010)
0.5	Oracle	-	-	-	-	0.391 (0.005)
				$\rho = 0.7$		
	KQR	-	-	-	-	1.343 (0.012)
	Boosting QR	0.00 (0.00)	0.73 (0.00)	1.00 (0.00)	30.38 (0.15)	0.562 (0.004)
	SNQR-5CV	0.41 (0.05)	0.04 (0.01)	0.95 (0.00)	5.11 (0.18)	0.653 (0.016)
	SNQR-SIC	0.51 (0.05)	0.03 (0.01)	0.95 (0.00)	4.87 (0.24)	0.656 (0.014)
	Adaptive SNQR-5CV	0.45 (0.05)	0.05 (0.01)	0.99 (0.00)	5.68 (0.27)	0.413 (0.012)
	Adaptive SNQR-SIC	0.70 (0.05)	0.03 (0.01)	0.99 (0.00)	4.86 (0.26)	0.405 (0.011)

Table 4

Estimated prediction risk for real datasets. The standard error is given in the parentheses.

		Methods			
τ	**Data**	**KQR**	**Boosting QR**	**SNQR**	**Adaptive SNQR**
0.2	Prostate	0.246 (0.008)	0.236 (0.019)	0.230 (0.006)	0.226 (0.007)
	Ozone	1.130 (0.016)	1.146 (0.017)	1.100 (0.016)	1.117 (0.017)
0.5	Prostate	0.301 (0.007)	0.313 (0.019)	0.318 (0.007)	0.310 (0.008)
	Ozone	1.656 (0.021)	1.690 (0.026)	1.657 (0.020)	1.669 (0.021)
0.8	Prostate	0.240 (0.007)	0.235 (0.018)	0.208 (0.005)	0.193 (0.006)
	Ozone	1.175 (0.017)	1.167 (0.018)	1.169 (0.017)	1.188 (0.016)