ELSEVIER Computational Statistics & Data Analysis 19 (1995) 613–630

COMPUTATIONAL STATISTICS & DATA ANALYSIS

A convergent algorithm for quantile regression with smoothing splines

Ronald J. Bosch[a,*], Yinyu Ye[b], George G. Woodworth[b]

[a] *Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA*
[b] *University of Iowa, Iowa City, IA 52242, USA*

Abstract

An important practical problem is that of determining a nonparametric estimate of the conditional quantile of y given x. If we balance fidelity to the data with a smoothness requirement, the resulting quantile function is a smoothing spline. We reformulate this estimation procedure as a quadratic programming problem, with associated optimality conditions. A recently developed interior point algorithm with proven convergence is extended to solve the quadratic program. This solution characterizes the desired nonparametric conditional quantile function. These methods are illustrated in a study of audiologic performance following cochlear implants.

Keywords: Smoothing splines; Nonparametric regression; Conditional quantile; Quadratic programming; Curve estimation.

1. Introduction

Researchers are often interested in how a covariate (x) affects a response variable (y). The regression of y on x is usually defined as the conditional mean of y given x. When the functional form of the conditional expectation is known, one is led to linear or nonlinear regression. A wide variety of techniques are available when the structure of the conditional mean is unknown, including kernel and nearest-neighbor approaches (Altman, 1992) and regression and smoothing splines (Wegman and Wright, 1983).

We focus on the problem of quantile regression. For a fixed $p \in (0, 1)$ we want to estimate the pth quantile of y as a function of x. Compared to mean regression procedures, quantile regression provides a more thorough picture of the relationship

* Corresponding author. E-mail:rbosch@hsph.harvard.edu.

SSDI 0167-9473(94)00018-E

between two variables since the quantiles are modelled individually. Quantile curves may be of primary interest, as in our cochlear application of Section 6, or they may be a useful diagnostic to complement more traditional regression analyses. Assuming symmetric errors, the conditional median ($p = 0.5$) is the analog of the least-squares regression, while the quartiles ($p = 0.25$, $p = 0.75$) provide information on spread similar to the residual standard deviation. Nonparametric quantile curves can indicate nonhomogeneous variance or a skewed conditional distribution. They can also suggest an appropriate transformation or parametric model for use in a standard regression.

In this paper, we estimate conditional quantiles with cubic smoothing splines. This estimator was proposed by Bloomfield and Steiger (1983) and computed using iteratively reweighted least squares. However, IRLS is computationally unstable when any of the estimated residuals approach zero. The interior point algorithm we propose was motivated by the instability of IRLS and allows the estimated residuals to approach zero without computational problems. Cubic spline estimators have been discussed by Jones (1988) for fitting quantile curves, and by Utreras (1981), Cox (1983), Eubank (1988) and Cunningham et al. (1991) in the context of robust smoothing splines. In closely related research, Koenker and Ng (1992) model conditional quantiles with quadratic smoothing splines, which are computed with linear programming techniques.

Other nonparametric estimators of conditional quantiles include the nearest neighbor approach (Stone, 1977) and kernel techniques (Samanta, 1989; Truong, 1989; Antoch and Janssen, 1989). Bhattacharya and Gangopadhyay (1990) study asymptotic properties of kernel and nearest neighbor estimators. White (1992) obtains consistent nonparametric conditional quantile estimators using artificial neural networks.

Parametric models may also be considered. Koenker and d'Orey (1987) solved the quantile regression problem using linear programming when the functional form of the conditional quantile is known to be linear in x. Procházka (1988) investigated the problem for a conditional quantile of a nonlinear functional form. Efron (1991) considers a known linear model for the quantiles and a class of asymmetric loss functions (including the check function presented below), in which the parameters are estimated using IRLS. Applications of quantile regression include salary prediction for statistics professors (Hogg, 1975), infant growth curves (Cole, 1988), lung capacity (Healy et al., 1988) and demand for electricity (Hendricks and Koenker, 1992).

In the next section we show how to cast nonparametric quantile regression as a quadratic programming problem. In Section 3 we develop an interior point algorithm with proven convergence properties, and in Section 4 we apply the algorithm in several situations. An efficient improvement to the algorithm is given in Section 5, and an application of the quantile smoothing spline to cochlear implants is presented in Section 6. The article closes with a section of general concluding remarks.

As with other nonparametric methods such as kernel estimators or the nearest-neighbor approach, our technique requires a smoothing parameter, λ, which balances local fidelity to the data with smoothness of the resulting quantile curve. The choice of λ is a problem of both theoretical and practical interest, but will not be addressed in this paper. A preliminary investigation of cross-validation has shown potential for this technique (Bosch, 1993), and we are currently pursuing this in more depth.

2. The quantile smoothing spline

Suppose that there are $n \geq 3$ observations of (x_i, y_i), where $a < x_1 < x_2 < \cdots < x_n < b$. Bloomfield and Steiger (1983) proposed a nonparametric estimator of the pth conditional quantile function of y given x as the function f that minimizes

$$\lambda \int_a^b [f''(x)]^2 \, \mathrm{d}x + \sum_{i=1}^{n} C_p(y_i - f(x_i)), \tag{2.1}$$

where the check function is defined as

$$C_p(u) = \begin{cases} (1-p)|u| & \text{if } u \leq 0, \\ p|u| & \text{if } u > 0, \end{cases}$$

and the parameter $\lambda > 0$ weights the roughness penalty. The minimization is over W_2, the second-order Sobolev space of functions f on $[a, b]$, with absolutely continuous first derivatives and square integrable second derivatives; i.e., f and f' are absolutely continuous and $\int_a^b [f''(x)]^2$ exists and is finite. We shall call the minimizing function the quantile smoothing spline, and will present an iterative procedure that converges to the correct solution.

The minimizer to the quantile estimation problem (2.1) over $f \in W_2$ is a natural cubic spline with knots $(x_1, \ldots, x_n)$. A natural cubic spline with knots $(x_1, \ldots, x_n)$ is a continuous function that is piecewise cubic in each $[x_{i-1}, x_i]$ and linear over $[a, x_1]$ and $[x_n, b]$, with the additional constraint of continuity of first and second derivatives over $[a, b]$, and a third derivative that is a step function with jumps at $(x_1, \ldots, x_n)$ (Eubank, 1988). The smoothing parameter λ can greatly affect the resulting curve. As $\lambda \to \infty$, the minimizing function becomes a straight line, and as $\lambda \to 0$, the minimizing function becomes the natural cubic spline interpolant of the observations.

Since the interpolating natural cubic spline through $(x_1, f(x_1)), \ldots, (x_n, f(x_n))$ is the smoothest function passing through those points (Eubank, 1988), it easily follows that the minimizing function must be a natural cubic spline. Thus, it is sufficient to minimize (2.1) over the class $NS(\boldsymbol{x})$ of natural cubic splines with knots $(x_1, \ldots, x_n)$. A convenient basis for $NS(\boldsymbol{x})$ is provided by the n δ-splines $\delta_1(x), \ldots, \delta_n(x)$. Each of the δ-splines is a natural cubic spline, but with the property that

$$\delta_i(x_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Since $NS(\boldsymbol{x})$ is a linear n-dimensional space, the n δ-splines provide an n-dimensional basis with the simplifying feature that for $f \in NS(\boldsymbol{x})$, we can write

$$f(x) = \sum_{i=1}^{n} v_i \delta_i(x),$$

where $v_i = f(x_i)$ (Utreras, 1981). That is, we can parametrize the functions in $NS(\boldsymbol{x})$ by their predicted values at the knots $(x_1, \ldots, x_n)$.

The δ-splines are straightforward to compute (Eubank, 1988). Let

$$W(s, t) = \int_a^b (s-u)_+ (t-u)_+ \, \mathrm{d}u,$$

where

$$(u)_+ = \begin{cases} u & \text{if } u > 0, \\ 0 & \text{if } u \leq 0. \end{cases}$$

For $s, t \in [a, b]$, the above integral reduces to

$$W(s, t) = \tfrac{1}{2}(z^* - z_*)(z_* - a)^2 + \tfrac{1}{3}(z_* - a)^3, \tag{2.2}$$

where $z_* = \min(s, t)$ and $z^* = \max(s, t)$. Let W_n be the $n \times n$ matrix with (i, j)th element $W(x_i, x_j)$ and let T be the $n \times 2$ matrix

$$T = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Now, define U' as the $(n-2) \times n$ matrix:

$$U' = \begin{bmatrix} u_1' \\ u_2' \\ \vdots \\ u_{n-2}' \end{bmatrix}, \quad u_i' = (\underbrace{0, \ldots, 0}_{i-1 \text{ 0's}}, \alpha_0^i, \alpha_1^i, \alpha_2^i, \underbrace{0, \ldots, 0}_{n-2-i \text{ 0's}}),$$

where

$$\alpha_0^i = (x_{i+1} - x_i)^{-1},$$
$$\alpha_1^i = -((x_{i+2} - x_{i+1})^{-1} + (x_{i+1} - x_i)^{-1}),$$
$$\alpha_2^i = (x_{i+2} - x_{i+1})^{-1}.$$

The δ-splines are then given by

$$(\delta_1(x), \ldots, \delta_n(x)) = (1, x, W(x, x_1), \ldots, W(x, x_n)) \begin{pmatrix} (T' W_n^{-1} T)^{-1} T' W_n^{-1} \\ U(U' W_n U)^{-1} U' \end{pmatrix}. \tag{2.3}$$

Since the minimization problem (2.1) involves the integrated squared section derivative, we shall need the integrated product of the second-order derivatives of the δ-splines. To be precise, we want the matrix Q, with (i, j)th element

$$Q_{ij} = 2\lambda \int_a^b \delta_i''(x)\,\delta_j''(x)\,\mathrm{d}x. \tag{2.4}$$

From the definition of the δ-splines, and using (2.2), it can be shown that $Q = 2\lambda U(U'W_nU)^{-1}U'$. Then for $f \in NS(\boldsymbol{x})$,

$$\begin{aligned}\lambda \int_a^b [f''(x)]^2\,\mathrm{d}x &= \lambda \int_a^b \left[\sum_{i=1}^n v_i \delta_i''(x)\right]^2 \mathrm{d}x \\ &= \lambda \int_a^b \sum_{i=1}^n \sum_{j=1}^n v_i v_j \delta_i''(x)\,\delta_j''(x)\,\mathrm{d}x \\ &= \tfrac{1}{2}\boldsymbol{v}'Q\boldsymbol{v}.\end{aligned}$$

As in Bloomfield and Steiger (1983), we may now rephrase problem (2.1) as the minimization of

$$\tfrac{1}{2}\boldsymbol{v}'Q\boldsymbol{v} + \sum_{i=1}^n q_i, \tag{2.5}$$

subject to

$$\begin{aligned}&q_i \geq p(y_i - v_i), \\ &q_i \geq (p-1)(y_i - v_i), \quad 1 \leq i \leq n,\end{aligned}$$

over $\boldsymbol{v} \in \Re^n$, $\boldsymbol{q} \in \Re^n_+$. This follows from the fact that for the check function $C_p(u) = \max(pu, (p-1)u)$. The quantile smoothing spline that is the minimizing function to (2.1) can then be expressed as $f(x) = \sum_{i=1}^n v_i \delta_i(x)$, using the $\boldsymbol{v}$ component of the solution to the minimization problem (2.5).

3. A convergent interior point algorithm

In this section, we describe a convergent interior point algorithm that solves the following problem:

(M1) Minimize

$$g(\boldsymbol{q}, \boldsymbol{u}) = \boldsymbol{e}'\boldsymbol{q} + \tfrac{1}{2}\boldsymbol{u}'P\boldsymbol{u} \text{ over } \boldsymbol{q} \in \Re^n_+,\ \boldsymbol{u} \in \Re^t,$$

subject to

$$\begin{aligned}&\boldsymbol{a}_1(\boldsymbol{q}, \boldsymbol{u}) = \boldsymbol{q} + p(M\boldsymbol{u} - \boldsymbol{y}) \geq \boldsymbol{0}, \\ &\boldsymbol{a}_2(\boldsymbol{q}, \boldsymbol{u}) = \boldsymbol{q} + (p-1)(M\boldsymbol{u} - \boldsymbol{y}) \geq \boldsymbol{0},\end{aligned}$$

where $0 < p < 1$ is the desired quantile, $\boldsymbol{y}$ is the observed y vector of length n, M is a fixed $n \times t$ matrix, $M\boldsymbol{u}$ is the vector of predicted values, and the minimizing $\boldsymbol{q}$ is the vector of weighted residuals. The linear term in the objective function $\boldsymbol{e}'\boldsymbol{q}$ is an 'error' penalty, where $\boldsymbol{e}$ is a column vector of ones. The quadratic term in $\boldsymbol{u}$ is a 'smoothness' penalty involving the fixed, positive semi-definite, symmetric matrix P. We require the $(n+t)\times t$ matrix

$$\begin{pmatrix} M \\ P \end{pmatrix}$$

to be of full column rank.

3.1. Optimality conditions

Since we are minimizing a quadratic function subject to linear inequalities, the Kuhn–Tucker optimality conditions determine a system of linear equations. We introduce the Lagrangian multipliers $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ associated with the two sets of linear inequalities $\boldsymbol{a}_1 \geq \boldsymbol{0}$ and $\boldsymbol{a}_2 \geq \boldsymbol{0}$. The optimality conditions, Luenberger (1984), are

$$\nabla_q g(\boldsymbol{q}, \boldsymbol{u}) - \boldsymbol{b}_1' \nabla_q \boldsymbol{a}_1(\boldsymbol{q}, \boldsymbol{u}) - \boldsymbol{b}_2' \nabla_q \boldsymbol{a}_2(\boldsymbol{q}, \boldsymbol{u}) = \boldsymbol{0},$$

$$\nabla_u g(\boldsymbol{q}, \boldsymbol{u}) - \boldsymbol{b}_1' \nabla_u \boldsymbol{a}_1(\boldsymbol{q}, \boldsymbol{u}) - \boldsymbol{b}_2' \nabla_u \boldsymbol{a}_2(\boldsymbol{q}, \boldsymbol{u}) = \boldsymbol{0},$$

$$\boldsymbol{a}'\boldsymbol{b} = 0,$$

$$\boldsymbol{a} \geq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{b} \geq \boldsymbol{0},$$

where $\boldsymbol{a}' = (\boldsymbol{a}_1', \boldsymbol{a}_2')$, $\boldsymbol{b}' = (\boldsymbol{b}_1', \boldsymbol{b}_2')$, and ∇ denotes the gradient operator. Reparametrize $\boldsymbol{q}$ as $\boldsymbol{q} = (1-p)\boldsymbol{a}_1 + p\boldsymbol{a}_2$, where $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are now interpreted as variables in $\Re^n$ subject to the constraint $\boldsymbol{a}_1 - \boldsymbol{a}_2 = M\boldsymbol{u} - \boldsymbol{y}$.

The optimality condition determined by the gradient with respect to $\boldsymbol{q}$ yields

$$\boldsymbol{b}_1 + \boldsymbol{b}_2 = \boldsymbol{e},$$

and the condition determined by the gradient with respect to $\boldsymbol{u}$ yields

$$P\boldsymbol{u} - pM'\boldsymbol{b}_1 - (p-1)M'\boldsymbol{b}_2 = \boldsymbol{0},$$

which upon substitution gives

$$P\boldsymbol{u} - M'\boldsymbol{b}_1 = (p-1)M'\boldsymbol{e}.$$

We have now transformed the original minimization problem (M1) into what is called a general linear complementarity problem:

(M2) Find $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{u}) \in \Re^{4n+t}$ such that $(\boldsymbol{a}, \boldsymbol{b}) \geq 0$, $\boldsymbol{a}'\boldsymbol{b} = 0$ and

$$\begin{pmatrix} 0 & 0 \\ I & -I \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{pmatrix} + \begin{pmatrix} I & I \\ 0 & 0 \\ M' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -M \\ -P \end{pmatrix} \boldsymbol{u} = \begin{pmatrix} \boldsymbol{e} \\ -\boldsymbol{y} \\ (1-p)M'\boldsymbol{e} \end{pmatrix}. \tag{3.1}$$

3.2. Outline of the convergent algorithm

Although the original problem, (M1), was to minimize a convex function over a constrained region, the linear form of the optimality conditions enables us to focus on these directly, (M2). The following iterative algorithm seeks a solution satisfying the optimality conditions (3.1), rather than explicitly minimizing the original objective function. It is to our advantage that the solution to the optimality conditions may be framed in the form of a general linear complementarity problem, for which a convergent algorithm is available.

Kojima et al. (1991) prove the convergence of a 'potential reduction' algorithm, which we extend to solve the problem (M2). Beginning with an interior, feasible solution $(\boldsymbol{a}^0, \boldsymbol{b}^0, \boldsymbol{u}^0)$, the algorithm generates a sequence of interior, feasible solutions $(\boldsymbol{a}^j, \boldsymbol{b}^j, \boldsymbol{u}^j)$ which converge to a solution satisfying the final optimality condition $\boldsymbol{a}'\boldsymbol{b} = 0$. A feasible solution is the one that satisfies (3.1) and an interior point is the one with strictly positive values for $\boldsymbol{a}$ and $\boldsymbol{b}$.

The development of the algorithm and proof of its convergence are based on the Tanabe–Todd–Ye potential function (Tanabe, 1988; Todd and Ye, 1990)

$$\phi(\boldsymbol{a}, \boldsymbol{b}) = (2n + \sqrt{2n})\log(\boldsymbol{a}'\boldsymbol{b}) - \sum_{i=1}^{2n} \log(a_i b_i) - 2n\log(2n). \tag{3.2}$$

The algorithm reduces the potential function by at least 0.2 each iteration (Kojima et al., 1991, Theorem 2.2). To show that it drives $\boldsymbol{a}'\boldsymbol{b}$ to 0, let us rewrite the potential function as

$$\phi(\boldsymbol{a}, \boldsymbol{b}) = \sqrt{2n}\log(\boldsymbol{a}'\boldsymbol{b}) - \sum_{i=1}^{2n} \log\left(\frac{2n a_i b_i}{\boldsymbol{a}'\boldsymbol{b}}\right).$$

Using the fact that the geometric mean of positive values is no greater than the arithmetic mean, we see that the second term is always nonnegative, which leads to

$$\phi(\boldsymbol{a}, \boldsymbol{b}) \geq \sqrt{2n}\log(\boldsymbol{a}'\boldsymbol{b}),$$

or

$$\boldsymbol{a}'\boldsymbol{b} \leq \exp\left(\frac{\phi(\boldsymbol{a}, \boldsymbol{b})}{\sqrt{2n}}\right).$$

From this, we see that an interior, feasible sequence $(\boldsymbol{a}^j, \boldsymbol{b}^j, \boldsymbol{u}^j)$ for which the potential function approaches $-\infty$ is also a sequence such that $a^{j\prime}\boldsymbol{b}^j$ approaches 0.

3.3. Initialization

In order to begin the algorithm, an interior, feasible point $(\boldsymbol{a}^0, \boldsymbol{b}^0, \boldsymbol{u}^0)$ is necessary. Let k^* be defined as

$$k^* = \max_{i=1,n}(\max(|py_i|, |(p-1)y_i|)),$$

and let $\gamma > 1$. (We have chosen $\gamma = 10$.) Then an initial interior, feasible point is

$$\begin{aligned} \boldsymbol{a}_1^0 &= \gamma k^* \boldsymbol{e} - p\boldsymbol{y} > \boldsymbol{0}, \\ \boldsymbol{a}_2^0 &= \gamma k^* \boldsymbol{e} - (p-1)\boldsymbol{y} > \boldsymbol{0}, \\ \boldsymbol{b}_1^0 &= (1-p)\boldsymbol{e} > \boldsymbol{0}, \\ \boldsymbol{b}_2^0 &= p\boldsymbol{e} > \boldsymbol{0}, \\ \boldsymbol{u}^0 &= \boldsymbol{0}. \end{aligned} \tag{3.3}$$

It is clear that this initial point is an interior point, satisfies the feasibility conditions (3.1), but does not satisfy the final optimality condition $\boldsymbol{a}'\boldsymbol{b} = 0$. This initial solution corresponds to predicted values of zero and $\boldsymbol{q} = \gamma k^* \boldsymbol{e}$.

3.4. Feasible descent

Kojima et al. (1991) reduce the potential function by taking a step in the direction of feasible descent. Given the current interior, feasible point $(\boldsymbol{a}^j, \boldsymbol{b}^j, \boldsymbol{u}^j)$, the next point is generated by

$$(\boldsymbol{a}^{j+1}, \boldsymbol{b}^{j+1}, \boldsymbol{u}^{j+1}) = (\boldsymbol{a}^j + \theta\Delta\boldsymbol{a}, \boldsymbol{b}^j + \theta\Delta\boldsymbol{b}, \boldsymbol{u}^j + \theta\Delta\boldsymbol{u}),$$

where $(\Delta\boldsymbol{a}, \Delta\boldsymbol{b}, \Delta\boldsymbol{u})$ is the direction of feasible descent for $\phi(\boldsymbol{a}^j, \boldsymbol{b}^j)$, and θ is the step size. Along the direction of feasible descent, the step size of Kojima et al. (1991) is known to reduce the potential function by 0.2 at each iteration. In the following, for clarity, the current point $(\boldsymbol{a}^j, \boldsymbol{b}^j, \boldsymbol{u}^j)$ will be written as $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{u})$.

The direction $(\Delta\boldsymbol{a}, \Delta\boldsymbol{b}, \Delta\boldsymbol{u})$ is determined by solving the linear system

$$A\Delta\boldsymbol{b} + B\Delta\boldsymbol{a} = (\mu k)^{-1}(\mu\boldsymbol{e} - AB\boldsymbol{e}),$$

where $A = \operatorname{diag}(\boldsymbol{a})$, $B = \operatorname{diag}(\boldsymbol{b})$, $\mu = \boldsymbol{a}'\boldsymbol{b}/(2n + \sqrt{2n})$, and $k = \|(AB)^{-1/2}\boldsymbol{e} - \mu^{-1}(AB)^{1/2}\boldsymbol{e}\|$, along with the system of equations

$$\begin{pmatrix} 0 & 0 \\ I & -I \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{a}_1 \\ \Delta\boldsymbol{a}_2 \end{pmatrix} + \begin{pmatrix} I & I \\ 0 & 0 \\ M' & 0 \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{b}_1 \\ \Delta\boldsymbol{b}_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -M \\ -P \end{pmatrix} \Delta\boldsymbol{u} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}$$

that ensure feasibility of the next point. Therefore, from any given interior, feasible point, the direction of feasible descent is the unique solution to the $(4n + t) \times (4n + t)$ system

$$\begin{pmatrix} I & 0 & A_1B_1^{-1} & 0 & 0 \\ 0 & I & 0 & A_2B_2^{-1} & 0 \\ 0 & 0 & I & I & 0 \\ I & -I & 0 & 0 & -M \\ 0 & 0 & M' & 0 & -P \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{a}_1 \\ \Delta\boldsymbol{a}_2 \\ \Delta\boldsymbol{b}_1 \\ \Delta\boldsymbol{b}_2 \\ \Delta\boldsymbol{u} \end{pmatrix} = \begin{pmatrix} (\mu k)^{-1}(\mu B_1^{-1}\boldsymbol{e} - \boldsymbol{a}_1) \\ (\mu k)^{-1}(\mu B_2^{-1}\boldsymbol{e} - \boldsymbol{a}_2) \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}. \tag{3.4}$$

Uniqueness follows from the fact that the matrix is full rank (Appendix A). The step size θ that ensures a reduction in the potential function is

$$\theta = 0.4 \min_{i=1,2n} (\sqrt{a_i b_i}).$$

The next point is then $\boldsymbol{a}^{j+1} = \boldsymbol{a}^j + \theta\,\Delta\boldsymbol{a}$, $\boldsymbol{b}^{j+1} = \boldsymbol{b}^j + \theta\,\Delta\boldsymbol{b}$, and $\boldsymbol{u}^{j+1} = \boldsymbol{u}^j + \theta\,\Delta\boldsymbol{u}$.

The proof that this direction and step size reduce the potential function by 0.2 each iteration follows Theorem 2.2 in Kojima et al. (1991) with one exception: a different proof is required that $\Delta\boldsymbol{a}'\Delta\boldsymbol{b} \geq 0$. (In Kojima et al., this followed from the relationship $\Delta\boldsymbol{b} = R\,\Delta\boldsymbol{a}$, for positive semi-definite R.) For our feasible descent direction, we have

$$\begin{aligned}\Delta\boldsymbol{a}'\Delta\boldsymbol{b} &= \Delta\boldsymbol{a}_1'\Delta\boldsymbol{b}_1 + \Delta\boldsymbol{a}_2'\Delta\boldsymbol{b}_2\\ &= (\Delta\boldsymbol{a}_1 - \Delta\boldsymbol{a}_2)'\Delta\boldsymbol{b}_1\\ &= (M\,\Delta\boldsymbol{u})'\Delta\boldsymbol{b}_1\\ &= \Delta\boldsymbol{u}'M'\,\Delta\boldsymbol{b}_1\\ &= \Delta\boldsymbol{u}'P\,\Delta\boldsymbol{u}\\ &\geq 0.\end{aligned}$$

We now have extended the result of Kojima et al. (1991) to our general linear complementarity problem.

Theorem 1. *The above step size and feasible descent direction reduce the potential function by* 0.2 *in each iteration.*

Furthermore, this reduction in the potential function at each iteration guarantees that the sequence $(\boldsymbol{a}^j, \boldsymbol{b}^j, \boldsymbol{u}^j)$ converges to a point which satisfies the optimality conditions. Starting from the initial point (3.3), a sequence of feasible, interior points is generated. The algorithm is terminated when $\boldsymbol{a}'\boldsymbol{b}$ becomes sufficiently close to zero (our implementation stops when $\boldsymbol{a}'\boldsymbol{b}/(1 + g(\boldsymbol{q},\boldsymbol{u}))$ falls below 10^{-8}).

4. Quantile function estimation

In this section, we show how the interior point algorithm may be used to solve several minimization problems that arise when estimating conditional quantiles with smoothing splines. We first consider the situation of all distinct x, then extend the problem to allow for tied x values and a partly linear model.

4.1. Distinct x values

Returning to the problem developed in Section 2, we minimize

$$\lambda \int_a^b [f''(x)]^2\,\mathrm{d}x + \sum_{i=1}^{n} C_p(y_i - f(x_i)),$$

over $f \in W_2$, the second-order Sobolev space, by solving the quadratic program:

Minimize

$$g(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{e}'\boldsymbol{q} + \tfrac{1}{2}\boldsymbol{v}'Q\boldsymbol{v} \text{ over } \boldsymbol{q} \in \Re^n_+,\ \boldsymbol{v} \in \Re^n,$$

subject to

$$\boldsymbol{a}_1(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + p(\boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0},$$

$$\boldsymbol{a}_2(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + (p-1)(\boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0}.$$

Letting $M = I$, $P = Q$, $\boldsymbol{u} = \boldsymbol{v}$, and $t = n$ in the interior point algorithm of Section 3, the step direction is obtained by solving the $n \times n$ system

$$[(A_1 B_1^{-1} + A_2 B_2^{-1})Q + I](\Delta\boldsymbol{a}_1 - \Delta\boldsymbol{a}_2)$$
$$= (\mu k)^{-1}[\mu(B_1^{-1} - B_2^{-1})\boldsymbol{e} - (\boldsymbol{a}_1 - \boldsymbol{a}_2)],$$

where $\mu = \boldsymbol{a}'\boldsymbol{b}/(2n + \sqrt{2n})$, $k = \|(AB)^{-1/2}\boldsymbol{e} - \mu^{-1}(AB)^{1/2}\boldsymbol{e}\|$, and A_1, A_2, B_1, and B_2 are diagonal matrices of the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{b}_1$ and $\boldsymbol{b}_2$, respectively. From this solution, the step directions are $\Delta\boldsymbol{b}_1 = Q(\Delta\boldsymbol{a}_1 - \Delta\boldsymbol{a}_2)$, $\Delta\boldsymbol{b}_2 = -\Delta\boldsymbol{b}_1$, $\Delta\boldsymbol{q} = (1-p)(\Delta\boldsymbol{a}_1 - \Delta\boldsymbol{a}_2) - (\mu k)^{-1}\boldsymbol{a}_2 + k^{-1}B_2^{-1}\boldsymbol{e} - A_2 B_2^{-1}\Delta\boldsymbol{b}_2$, $\Delta\boldsymbol{a}_1 = \Delta\boldsymbol{q} + p(\Delta\boldsymbol{a}_1 - \Delta\boldsymbol{a}_2)$ and $\Delta\boldsymbol{a}_2 = \Delta\boldsymbol{q} + (p-1)(\Delta\boldsymbol{a}_1 - \Delta\boldsymbol{a}_2)$. Upon termination of the algorithm, the estimated quantiles are $\boldsymbol{v} = \boldsymbol{y} + \boldsymbol{a}_1 - \boldsymbol{a}_2$, which, via the δ-spline basis functions (2.3), determine the quantile smoothing spline, $f(x) = \sum_{i=1}^n v_i \delta_i(x)$.

4.2. Extensions

One limitation of the algorithm developed above is the requirement of distinct x values. In addition, there appears to be numerical instability in the algorithm when x values are nearly equal, evidently due to the nature of the δ-splines (Bosch, 1993). For these reasons, we are interested in estimating the quantile smoothing spline when not all of the x values are distinct. Suppose that we have observations (x_i, y_i), $1 \leq i \leq n$, with m $(3 \leq m < n)$ distinct x values. To minimize (2.1) over W_2, we solve the quadratic program:

Minimize

$$g(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{e}'\boldsymbol{q} + \tfrac{1}{2}\boldsymbol{v}'Q\boldsymbol{v} \text{ over } \boldsymbol{q} \in \Re^n_+,\ \boldsymbol{v} \in \Re^m,$$

subject to

$$\boldsymbol{a}_1(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + p(N\boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0},$$

$$\boldsymbol{a}_2(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + (p-1)(N\boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0}.$$

Notation follows the previous section, except that now $\boldsymbol{v}$ is a vector of length m, Q is an $m \times m$ positive-semi-definite matrix determined from the m distinct x values and the $n \times m$ matrix N associates the v_j with the appropriate y_i (An example showing the N matrix is given below.) Letting $M = N, P = Q, \boldsymbol{u} = \boldsymbol{v}$, and $t = m$ in the algorithm of Section 3, we obtain the step direction by solving the $2n \times 2n$ system

$$\begin{pmatrix} I & A_1 B_1^{-1} N(N'N)^{-1} Q & A_1 B_1^{-1} N^{\perp} \\ I & -A_2 B_2^{-1} N(N'N)^{-1} Q - N & -A_2 B_2^{-1} N^{\perp} \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{a}_1 \\ \Delta \boldsymbol{v} \\ \Delta \boldsymbol{w} \end{pmatrix}$$
$$= \begin{pmatrix} (\mu k)^{-1} (\mu B_1^{-1} \boldsymbol{e} - \boldsymbol{a}_1) \\ (\mu k)^{-1} (\mu B_2^{-1} \boldsymbol{e} - \boldsymbol{a}_2) \end{pmatrix},$$

where $\Delta \boldsymbol{w}$ is a vector of length $n - m$ and the $n \times (n - m)$ matrix $N^{\perp}$ is the orthogonal complement of N. Now let $\Delta \boldsymbol{a}_2 = \Delta \boldsymbol{a}_1 - N \Delta \boldsymbol{v}$, $\Delta \boldsymbol{b}_1 = N(N'N)^{-1} Q \Delta \boldsymbol{v} + N^{\perp} \Delta \boldsymbol{w}$, and $\Delta \boldsymbol{b}_2 = -\Delta \boldsymbol{b}_1$. Upon termination of the algorithm, the estimated quantiles are $\boldsymbol{v} = (N'N)^{-1} N'(\boldsymbol{y} + \boldsymbol{a}_1 - \boldsymbol{a}_2)$, which determined the quantile smoothing spline $f(x) = \sum_{j=1}^{m} v_j \delta_j(x)$, where the δ-splines are derived from the m distinct x values.

Example. A simple example illustrates the matrices N and $N^{\perp}$. Suppose the x values are (1,2,2,2,2,3,4,4), implying that $n = 8$ and $m = 4$. We then have

$$N = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad N^{\perp} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Heckman (1986) and Wahba (1990) have investigated a partly linear model in conjuction with spline smoothing, modelling the conditional mean as the sum of a linear function and a smooth function. With a partly linear model for the conditional quantile, the minimization of

$$\lambda \int_a^b [f''(x)]^2 \, \mathrm{d}x + \sum_{i=1}^{n} C_p(y_i - z'_{(i)} \beta - f(x_i)),$$

over $\beta \in \mathfrak{R}^r$ and $f \in W_2$, for distinct x, may be cast as the quadratic program:

Minimize

$$g(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{e}'\boldsymbol{q} + \tfrac{1}{2} \boldsymbol{v}' Q \boldsymbol{v} \ \text{ over } \ \boldsymbol{q} \in \mathfrak{R}_+^n, \ \boldsymbol{v} \in \mathfrak{R}^n,$$

subject to

$$\boldsymbol{a}_1(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + p(Z\beta + \boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0},$$

$$\boldsymbol{a}_2(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + (p-1)(Z\beta + \boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0}.$$

where Z is an $n \times r$ regression matrix with rows $z'_{(i)}$. As in Wahba (1990), we need to impose a restriction on the matrix Z to avoid identifiability problems: we assume the $n \times (r+2)$ matrix $(Z : \boldsymbol{e} : \boldsymbol{x})$ to be of full column rank. Applying the algorithm of Section 3, the step direction is obtained from the solution to

$$[(A_1 B_1^{-1} + A_2 B_2^{-1})(Q - QZ(Z'QZ)^{-1}Z'Q) + I]\Delta \boldsymbol{a}_1 - \Delta \boldsymbol{a}_2)$$
$$= (\mu k)^{-1}[\mu(B_1^{-1} - B_2^{-1})\boldsymbol{e} - (\boldsymbol{a}_1 - \boldsymbol{a}_2)].$$

The nonsingularity of $Z'QZ$ follows from the identifiability assumption on Z (see Appendix B). Now let $\Delta \boldsymbol{b}_1 = (Q - QZ(Z'QZ)^{-1}Z'Q)(\Delta \boldsymbol{a}_1 - \Delta \boldsymbol{a}_2)$, $\Delta \boldsymbol{b}_2 = -\Delta \boldsymbol{b}_1$, $\Delta \boldsymbol{q} = (1-p)(\Delta \boldsymbol{a}_1 - \Delta \boldsymbol{a}_2) - (\mu k)^{-1}\boldsymbol{a}_2 + k^{-1}B_2^{-1}\boldsymbol{e} - A_2 B_2^{-1}\Delta \boldsymbol{b}_2$, $\Delta \boldsymbol{a}_1 = \Delta \boldsymbol{q} + p(\Delta \boldsymbol{a}_1 - \Delta \boldsymbol{a}_2)$ and $\Delta \boldsymbol{a}_2 = \Delta \boldsymbol{q} + (p-1)(\Delta \boldsymbol{a}_1 - \Delta \boldsymbol{a}_2)$. Upon termination of the algorithm, calculate $\beta = (Z'QZ)^{-1}Z'Q(\boldsymbol{y} + \boldsymbol{a}_1 - \boldsymbol{a}_2)$ and $\boldsymbol{v} = \boldsymbol{y} + \boldsymbol{a}_1 - \boldsymbol{a}_2 - Z\beta$, which determines the smooth function $f(x) = \sum_{i=1}^{n} v_i \delta_i(x)$.

For a partly linear model with tied x values, we obtain the quadratic program:

Minimize

$$g(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{e}'\boldsymbol{q} + \tfrac{1}{2}\boldsymbol{v}'Q\boldsymbol{v} \quad \text{over} \quad \boldsymbol{q} \in \Re^n_+,\ \boldsymbol{v} \in \Re^m,$$

subject to

$$\boldsymbol{a}_1(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + p(Z\beta + N\boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0},$$

$$\boldsymbol{a}_2(\boldsymbol{q}, \boldsymbol{v}) = \boldsymbol{q} + (p-1)(Z\beta + N\boldsymbol{v} - \boldsymbol{y}) \geq \boldsymbol{0},$$

where the $n \times m$ matrix N is shown in the example above. To avoid identifiability problems, we assume that the matrix Z is such that $(Z : N\boldsymbol{e} : N\boldsymbol{x})$ is full column rank, where $\boldsymbol{x}$ is the vector of distinct x values. This implies full column rank for

$$H = \begin{pmatrix} Z & N \\ 0 & 0 \\ 0 & Q \end{pmatrix}.$$

The step direction for the interior point algorithm is determined by solving

$$\begin{pmatrix} I & A_1 B_1^{-1} H_1^{\perp} \\ I & -(A_2 B_2^{-1} H_1^{\perp} + Z H_2^{\perp} + N H_3^{\perp}) \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{a}_1 \\ \Delta \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} (\mu k)^{-1}(\mu B_1^{-1}\boldsymbol{e} - \boldsymbol{a}_1) \\ (\mu k)^{-1}(\mu B_2^{-1}\boldsymbol{e} - \boldsymbol{a}_2) \end{pmatrix},$$

where $H^{\perp}$ is the orthogonal complement to H. That is, the columns of $H^{\perp}$ span $\{\boldsymbol{c}: H'\boldsymbol{c} = \boldsymbol{0}\}$. Partition $H^{\perp}$ into $H_1^{\perp}$, the top n rows; $H_2^{\perp}$, the next r rows; and $H_3^{\perp}$, the bottom m rows, and calculate $\Delta \boldsymbol{b}_1 = H_1^{\perp}\Delta \boldsymbol{w}$, $\Delta \boldsymbol{b}_2 = -\Delta \boldsymbol{b}_1$, $\Delta\beta = H_2^{\perp}\Delta \boldsymbol{w}$, $\Delta \boldsymbol{v} = H_3^{\perp}\Delta \boldsymbol{w}$, and $\Delta \boldsymbol{a}_2 = \Delta \boldsymbol{a}_1 - Z\Delta\beta - N\Delta \boldsymbol{v}$. Note that $\boldsymbol{w}^0$ is initialized to $\boldsymbol{0}$. Once the iterations have driven $\boldsymbol{a}'\boldsymbol{b}$ to zero, solve for $\beta = H_2^{\perp}\boldsymbol{w}$ and $\boldsymbol{v} = H_3^{\perp}\boldsymbol{w}$, which determines the smooth function $f(x) = \sum_{i=1}^{m} v_j \delta_j(x)$.

5. Computational efficiency

Although the step size of Kojima et al. (1991) is proven theoretically to reduce the potential function by a constant, a heuristic, aggressive step size seems to provide much faster convergence in practice. The aggressive step size that we used moves 90% of the distance to the boundary. That is, let the step size θ be chosen as in Ye (1991):

$$\theta = \beta \sup(\theta : \boldsymbol{a}^k + \theta \Delta \boldsymbol{a} \geq 0, \text{ and } \boldsymbol{b}^k + \theta \Delta \boldsymbol{b} \geq 0),$$

where $\beta \in (0, 1)$. We chose $\beta = 0.9$.

This aggressive step size seems to work well with a slightly different potential function than (3.2). Let

$$\varphi(\boldsymbol{a}, \boldsymbol{b}) = (2n^{1.5})\log(\boldsymbol{a}'\boldsymbol{b}) - \sum_{i=1}^{2n} \log(a_i b_i) - 2n\log(2n). \tag{5.1}$$

The feasible descent direction associated with $\varphi(\boldsymbol{a}^k, \boldsymbol{b}^k)$ of Eq. (5.1) will not be the direction determined above from $\phi(\boldsymbol{a}^k, \boldsymbol{b}^k)$ of Eq. (3.2), and requires the solution of a different system of linear equations. With these modifications, the algorithm is quite feasible in practice, with convergence in less than a minute on a Sun workstation for sample sizes less than 100.

6. The Iowa Cochlear Implant Project

Our research on quantile smoothing splines has been motivated by the third author's work with the Iowa Cochlear Implant Project, a 10 yr clinical trial of a device which can restore partial hearing to totally deaf people. This surgically implanted device directly stimulates the auditory nerve in the inner ear. One goal of the project has been to predict which patients will most benefit from the device. Data were obtained from 48 patients who received multichannel cochlear implants at the University of Iowa Cochlear Implant Center. Prior to implantation, patients were profoundly deaf, with less than 4% recognition of a standard word list. A predictive index that is a linear combination of several preimplant tests was developed as a means of predicting audiologic performance with the implant (Gantz et al., 1993).

Fig. 1 shows the relationship between the predictive index and audiologic performance nine months after connection of the implant as measured by the Iowa Laser Videodisc Sentence Test (Tyler et al., 1983, 1986). This score is the percent of correctly identified words in a test consisting of 100 short sentences presented in random order to the cochlear implant wearer via a television monitor with high-fidelity sound. Subjects both see and hear the sentences being spoken. Estimated 10th, 25th, 50th, 75th, and 90th percentiles were obtained by fitting quantile smoothing splines with $\lambda = 28$. This lambda value was selected by looking at quantile plots for a range of λ values; the selected λ yielded the smoothest curves without crossovers of the estimated quantile curves.

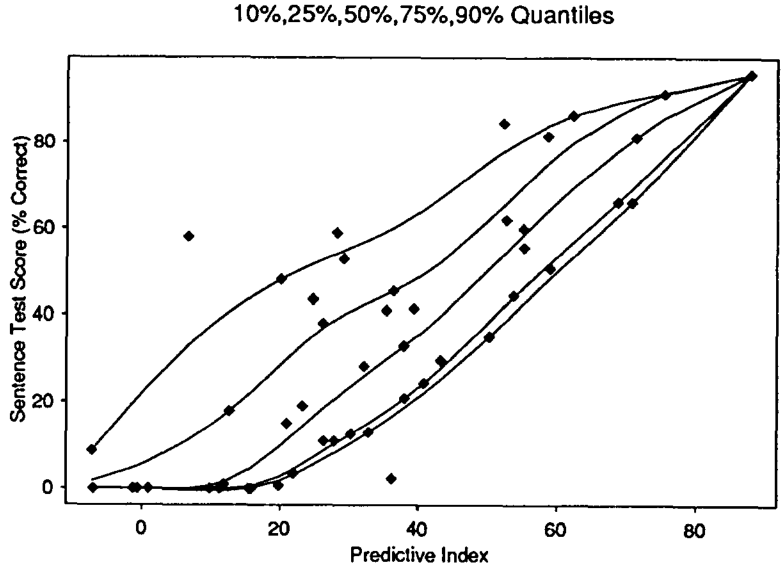

Fig. 1. The Iowa laser videodisc sentence test.

These quantile curves give clinicians the ability to inform the implant candidate about the likelihood of improvement over the performance with a conventional hearing aid and to give a range for the degree of improvement. For example, with a patient having a predictive index of 20, there is a 50% chance of less than 10% recognition. However, with a predictive index of 60, there is an even chance that the patient will recognize two out of three words on the sentence test, and there is a 90% chance that more than half the words will be recognized.

7. Concluding remarks

In this paper we have developed an algorithm for estimating quantile curves with cubic smoothing splines. In addition to theoretical convergence, the algorithm is straightforward to implement and appears to be computationally feasible for moderate sample sizes. However, large samples could be numerically infeasible, due to the size of the system of linear equations that must be solved each iteration.

When the response is a proportion, an alternative to nonparametric quantile regression is to parametrize the underlying success probability in a beta-binomial regression model (see Prentice, 1986, and references therein). Likelihood ratio tests could be used to test hypotheses concerning the regression parameters. Desired conditional quantile curves may then be computed from the resulting maximum likelihood estimates. However, for the cochlear data in our application, there is the added complication that the number of correctly identified words is not distributed binomial, since knowledge of a few words in a sentence provides a useful context for identifying the remaining words.

Some artifacts of the quantile regression model are apparent in Fig. 1. As with other smoothing techniques, there appears to be a pronounced edge effect and care should be exercised for interpretation at the extreme x values. There is also a tendency for these curves to interpolate some of the data points. Although not evident for this choice of λ, it can happen that quantile curves cross. This is a potentially serious problem for quantile regression procedures, since each quantile is estimated separately.

A number of extensions to our work are possible. We have considered the situation where the response is smoothed with respect to a scalar x. An extension of our algorithm could smooth with respect to two or more continuous variables using the thin plate spline (Wahba, 1990). Asymptotics for quantile curves is an area that remains relatively unexplored. Bosch (1993) has shown consistency and asymptotic normality assuming that x has a distribution on a finite number of values. More general results might follow the methods of Cox (1983), with modification to account for the nondifferentiability of the check function. Analogous to the results of Heckman (1986) for squared error, large sample theory may provide asymptotic standard errors for the regression parameters in the partly linear model of Section 4.2. Lastly, we mention that selection of the smoothing parameter λ is an important problem that needs further investigation. Cross-validation appears to be a promising approach, and we are currently exploring this technique for quantile smoothing splines.

Acknowledgements

This research is part of the first author's Ph.D. dissertation, and has been partly supported by the National Institutes of Health research grant ES07142. The work of the third author was supported in part by research grant DC00242 from the National Institutes of Health/NIDCD, grant RR59 from the General Clinical Research Centers Program, Division of Research Resources, NIH, and the Iowa Lions Sight and Hearing Foundation. The authors are grateful to David Wypij and two anonymous referees for detailed comments that have substantially improved the article.

Appendix A

In this appendix, we show full column rank for the square matrix

$$G = \begin{pmatrix} I & 0 & A_1 B_1^{-1} & 0 & 0 \\ 0 & I & 0 & A_2 B_2^{-1} & 0 \\ 0 & 0 & I & I & 0 \\ I & -I & 0 & 0 & -M \\ 0 & 0 & M' & 0 & -P \end{pmatrix},$$

under the following three assumptions: (1) P is positive semi-definite. (2) The matrix

$$\begin{pmatrix} M \\ P \end{pmatrix}$$

is of full column rank. (3) The matrices A_1, A_2, B_1 and B_2 are positive definite.

Assume that

$$\begin{pmatrix} I & 0 & A_1 B_1^{-1} & 0 & 0 \\ 0 & I & 0 & A_2 B_2^{-1} & 0 \\ 0 & 0 & I & I & 0 \\ I & -I & 0 & 0 & -M \\ 0 & 0 & M' & 0 & -P \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \boldsymbol{u} \end{pmatrix} = \boldsymbol{0},$$

and consider this as five sets of equations. The first three sets imply that

$$\boldsymbol{a}_1 - \boldsymbol{a}_2 + (A_1 B_1^{-1} + A_2 B_2^{-1}) \boldsymbol{b}_1 = \boldsymbol{0}.$$

Substituting from the fourth set of equations and premultiplying, we have

$$(A_1 B_1^{-1} + A_2 B_2^{-1})^{-1} M \boldsymbol{u} + \boldsymbol{b}_1 = \boldsymbol{0}.$$

Premultiply by $\boldsymbol{u}'\boldsymbol{M}'$ and substitute to obtain

$$\boldsymbol{u}' \boldsymbol{M}' (A_1 B_1^{-1} + A_2 B_2^{-1})^{-1} M \boldsymbol{u} + \boldsymbol{u}' P \boldsymbol{u} = 0,$$

which implies that $M\boldsymbol{u} = \boldsymbol{0}$ and $P\boldsymbol{u} = \boldsymbol{0}$. Note that $\boldsymbol{u}'P\boldsymbol{u} = 0 \Leftrightarrow P\boldsymbol{u} = 0$ (Seber, 1977). By assumption (2) it follows that $\boldsymbol{u} = 0$, which from above implies that $\boldsymbol{b}_1 = 0$. From the first three sets of equations it now follows that $\boldsymbol{a}_1 = \boldsymbol{a}_2 = \boldsymbol{b}_2 = 0$ and hence, that G is full rank.

Appendix B

This appendix shows the equivalence of two matrix conditions which are involved in our application of a partly linear model (see Section 4.2). We know that Q is a positive semi-definite matrix with rank $n - 2$, and $\boldsymbol{e}'Q\boldsymbol{e} = \boldsymbol{x}'Q\boldsymbol{x} = 0$, where $\boldsymbol{e}$ is a vector of ones and $\boldsymbol{x}$ is the vector of x values.

As in Wahba (1990), we require (1) the $n \times (r + 2)$ matrix $(Z{:}\boldsymbol{e}{:}\boldsymbol{x})$ is full column rank. We will show that this condition is equivalent to (2) the $r \times r$ symmetric matrix $Z'QZ$ is full rank. Moreover, these equivalent conditions imply that $Z'QZ$ is positive definite.

(1) $\Rightarrow$ (2) (not (2) $\Rightarrow$ not (1)). Suppose that $Z'QZ$ is not full rank. Let $\boldsymbol{w} \neq 0$. Then $\boldsymbol{w}'Z'QZ\boldsymbol{w} = (Z\boldsymbol{w})'Q(Z\boldsymbol{w}) \geq 0$, since Q is positive semi-definite. Suppose that for all $\boldsymbol{w} \neq 0$, $\boldsymbol{w}'Z'QZ\boldsymbol{w} > 0$. Since this would imply that $Z'QZ$ is positive definite, this cannot be the case. Therefore, there exists $\boldsymbol{w}_0 \neq 0$, such that $\boldsymbol{w}_0'Z'QZ\boldsymbol{w}_0 = 0$, or

$(Z w_0)' Q(Z w_0) = 0$. This implies that $Z w_0$ is in the column space of $(e : x)$ which in turn implies that $(Z{:}e{:}x)$ is not full column rank.

(2) $\Rightarrow$(1) (not (1) $\Rightarrow$ not (2)). Suppose that $(Z{:}e{:}x)$ is not full column rank. Without loss of generality, this implies that

$$z_1 = \sum_{\alpha=2}^{r} \gamma_\alpha z_\alpha + \gamma_{r+1} e + \gamma_{r+2} x,$$

where z_1 is the first column of Z. In terms of the elements, this implies that

$$Z_{i1} = \sum_{\alpha=2}^{r} \gamma_\alpha Z_{i\alpha} + \gamma_{r+1} + \gamma_{r+2} x_i, \quad i = 1, \ldots, n.$$

Using the fact that if Q is positive semi-definite, then for any conforming matrix R, $R'QR = 0 \Leftrightarrow QR = 0$ (Seber, 1977), we obtain $\sum_{i=1}^{n} Q_{ij} = 0$ and $\sum_{i=1}^{n} x_i Q_{ij} = 0$. Examining the first row of $G = Z'QZ$, we have for $k = 1, \ldots, r$ that

$$\begin{aligned} G_{1k} &= \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{i1} Z_{jk} Q_{ij} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\alpha=2}^{r} \gamma_\alpha Z_{i\alpha} Z_{jk} Q_{ij} \\ &\quad + \gamma_{r+1} \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{jk} Q_{ij} + \gamma_{r+2} \sum_{i=1}^{n} x_i \sum_{j=1}^{n} Z_{jk} Q_{ij} \\ &= \sum_{\alpha=2}^{r} \gamma_\alpha \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{i\alpha} Z_{jk} Q_{ij} \\ &\quad + \gamma_{r+1} \sum_{j=1}^{n} Z_{jk} (\textstyle\sum_{i=1}^{n} Q_{ij}) + \gamma_{r+2} \displaystyle\sum_{j=1}^{n} Z_{jk} (\sum_{i=1}^{n} x_i Q_{ij}) \\ &= \sum_{\alpha=2}^{r} \gamma_\alpha G_{\alpha k}. \end{aligned}$$

Thus, we conclude that $Z'QZ$ is not full rank.

References

Altman, N.S., An introduction to kernel and nearest-neighbor nonparametric regression, *Amer. Statist.*, **46** (1992) 175–185.

Antoch, J. and P. Janssen, Nonparametric regression M-quantiles, *Statist. Probab. Lett.*, **8** (1989) 355–362.

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

Bloomfield, P. and W. S. Steiger, *Least absolute deviations. Theory, applications and algorithms* (Birkhäuser-Verlag, Boston, 1983).

Bosch, R.J., Quantile regression with smoothing splines, Ph.D. Dissertation (Department of Statistics, University of Iowa, Printed by U.M.I., Ann. Arbor, MI, 1993).

Cole, T.J., Fitting smoothed centile curves to reference data (with discussion), *J. Roy. Statist. Soc. Ser. A*, **151** (1988) 385–418.

Cox, D.D., Asymptotics for M-type smoothing splines, *Ann. Statist.*, **11** (1983) 530–551.

Cunningham, J.K., R.L. Eubank and T. Hsing, M-type smoothing splines with auxiliary scale estimation, *Comput. Statist. Data Anal.*, **11** (1991) 43–51.

Efron, B., Regression percentiles using asymmetric squared error loss, *Statist. Sinica*, **1** (1991) 93–125.

Eubank, R.L., *Spline smoothing and nonparametric regression* (Marcel Dekker, New York, 1988).

Gantz, B.J., G.G. Woodworth, P.J. Abbas, J.F. Knutson and R.S. Tyler, Multivariate predictors of audiological success with multichannel cochlear implants, *Ann. Otology Rhinology Laryngology*, **102** (1993) 909–916.

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

Heckman, N.E., Spline smoothing in a partly linear model, *J. Roy. Statist. Soc. Ser. B*, **48** (1986) 244–248.

Hendricks, W. and R. Koenker, Hierarchical spline models for conditional quantiles and the demand for electricity, *J. Amer. Statist. Assoc.*, **87** (1992) 58–68.

Hogg, R. V., Estimates of percentile regression lines using salary data, *J. Amer. Statist. Assoc.* **70** (1975) 56–59.

Jones, M.C., Discussion of Cole, *J. Roy. Statist. Soc. Ser. A*, **151** (1988) 412–413.

Koenker, R. and V. d'Orey, Computing regression quantiles, *Appl. Statist.*, **36** (1987) 383–393.

Koenker, R. and P. Ng, Quantile smoothing splines, in: A.K.Md.E. Saleh (Ed.), *Nonparametric statistics and related topics* (North-Holland, New York, 1992) 205–215.

Kojima, M., S. Mizuno and A. Yoshise, An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems, *Math. Program.*, **50** (1991) 331–342.

Luenberger, D.G., *Linear and nonlinear programming* (Addison-Wesley, Reading, MA, 1984).

Prentice, R.L., Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors, *J. Amer. Statist. Assoc.*, **81** (1986) 321–327.

Procházka, B., Regression quantiles and trimmed least squares estimator in the nonlinear regression model, *Comput. Statist. Data Anal.*, **6** (1988) 385–391.

Samanta, M., Nonparametric estimation of conditional quantiles, *Statist. Probab. Lett.*, **7** (1989) 407–412.

Seber, G.A.F., *Linear regression analysis* (Wiley, New York, 1977).

Stone, C.J., Consistent nonparametric regression (with discussion), *Ann. Statist.*, **5** (1977) 595–645.

Tanabe, K., Centered Newton method for mathematical programming, in: M. Iri and K. Yajima (Eds.), *System modelling and optimization* (Springer, New York, 1988) 197–206.

Todd, M. and Y. Ye, A centered projective algorithm for linear programming, *Math. Oper. Res.*, **15** (1990) 508–529.

Truong, Y., Asymptotic properties of kernel estimators based on local medians, *Ann. Statist.*, **17** (1989) 606–617.

Tyler, R., J. Preece and M. Lowder, *The Iowa cochlear implant tests* (Department of Otolaryngology-Head and Neck Surgery, The University of Iowa, Iowa City, 1983).

Tyler, R., J. Preece and N. Tye-Murray, *The Iowa phoneme and sentence laser videodisc* (Department of Otolaryngology-Head and Neck Surgery, The University of Iowa, Iowa City, 1986).

Utreras, F., On computing robust splines and applications, *SIAM J. Sci. Statist. Computing*, **2** (1981) 153–163.

Wahba, G., *Spline models for observational data* (SIAM, Philadelphia, 1990).

Wegman, E.J. and I.W. Wright, Splines in statistics, *J. Amer. Statist. Assoc.*, **78** (1983) 351–365.

White, H., Nonparametric estimation of conditional quantiles using neural networks, in: C. Page and R. LePage (Eds.), *Proc. 22nd symp. on the interface: computing science and statistics.* (Springer, New York, 1992) 190–199.

Ye, Y., Interior-point algorithms for quadratic programming, in: S. Kumar (Ed.), *Recent developments in mathematical programming* (Gordon and Breach, Philadelphia, 1991) 237–261.