Bivariate Quantile Smoothing Splines

Xuming He Pin Ng Stephen Portnoy *

Abstract

It has long been recognized that the mean provides an inadequate summary while the set of quantiles can supply a more complete description of a sample. We introduce bivariate quantile smoothing splines, which belong to the space of bi-linear tensor-product splines, as nonparametric estimators for the conditional quantile functions in a two dimensional design space. The estimators can be computed using standard linear programming techniques and can further be used as building blocks for conditional quantile estimations in higher dimensions. For moderately large data sets, we recommend using penalized bivariate B-splines as approximate solutions. We use real and simulated data to illustrate the proposed methodology.

KEY WORDS: Conditional quantile; Linear program; Nonparametric regression; Robust regression; Schwarz information criterion; Tensor-product spline.

---

*Xuming He is Associate Professor and Stephen Portnoy is Professor, Department of Statistics, University of Illinois, 725 S Wright, Champaign, Illinois 61820. Pin Ng is Lecturer, Department of Economics, University of University, Champaign, Illinois 61820. Xuming He's research was partially supported by NSA grant MDA904-96-1-0011 and NSF Grant SBR 9617278. Stephen Portnoy's research was partially supported by NSF grant SES 9121776. The authors would like to thank Roger Koenker for very helpful discussions. The comments from the Associate Editor and two referees are also appreciated.

1 INTRODUCTION

Smoothing splines play an important role in nonparametric function estimations. Their successful development from a mathematical framework of variational problems in Hilbert spaces to useful statistical techniques is described in two recent monographs by Wahba (1990), and Green and Silverman (1994). Like most other smoothing techniques , smoothing splines have primarily been used to estimate the conditional mean function of a response variable given one or several independent or design variables. Averaging (including local averaging) is well known to be sensitive to the presence of outliers. Härdle (1989, Chap. 6), and more recently Wang and Scott (1994), provided several alternatives for robust smoothing.

The special case of median smoothing considered in this paper provides an important robust alternative to traditional mean smoothing. However, we are concerned mainly with estimation of *conditional quantile functions,* which have been recognized increasingly as an integral part of modern routine data analyses. Efron (1991) provided some interesting examples of conditional quantile estimations for linear models. In a study on electricity demand, Hendricks and Koenker (1992) estimated conditional quantile functions and showed that heavy users of electricity exhibited much larger daily periodicity during the summer than moderate users did due to air conditioner usage. In many (if not all) regression examples, one would expect a different structural relationship for the higher (or lower) responses than the average responses. In such cases, both mean and median analyses would overlook important features that could be uncovered only by a more general conditional quantile analysis.

Some recent work on univariate nonparametric estimation of conditional quantile functions can be found in Antoch and Janssen (1989), Bhattacharya and Gangopadhyay (1989), White (1990), Zelterman (1990), Chaudhuri (1991), He and Shi (1994), Welsh (1996), and He (1997). Koenker *et al.* (1994) suggested the following direct approach using *quantile smoothing splines.*

For a univariate design variable x_i with observed responses z_i, the τ-th quantile smoothing spline minimizes (over functions $g(x)$):

$$\sum_{i=1}^{n} \rho_\tau \left(z_i - g\left(x_i\right)\right) + \lambda V\left(g'\right) \tag{1}$$

for $\tau \in (0,1)$, where $\rho_\tau(r) = \tau \max\{r, 0\} + (1-\tau)\max\{-r, 0\}$, and $V(h) = \sup \sum_{i=1}^{K} |h(t_i) - h(t_{i-1})|$ denotes the total variation of the function h with the 'sup' being taken over all finite partitions $t_0 < t_1 < \cdots < t_K$ of the support of h. If h is differentiable, it is easy to see that $V(h) = \int |h'(x)| dx$. The optimal solution, $\hat{g}_\tau(x)$, estimates the τ-th conditional quantile function $g_\tau(x)$ which satisfies

$$P\left\{Z < g_\tau(x) \mid X = x\right\} = \tau$$

for all x. The problem of quantile smoothing in (1) can be viewed as an analogy to the more extensively studied mean smoothing of minimizing

$$\sum_{i=1}^{n} \left(z_i - g\left(x_i\right)\right)^2 + \lambda \int \left(g''\left(x\right)\right)^2 dx. \tag{2}$$

The solution to (2) is a natural cubic smoothing spline with knots at the observed design points. Its computation is rather efficient as it simply amounts to solving a linear system. The solution to (1) is a linear smoothing spline with possible breaks in the derivative at the design points, and the computation can be performed by modern linear programming methods such as Ng (1996)'s adaptation of the non-simplex active-set algorithm of Bartels and Conn (1980). Consistency results of mean smoothing splines require the existence of $g\left(x\right) = E\left[Z|X = x\right]$ but no restriction on the distribution of (X, Z) is necessary for quantile smoothing splines. Recently, Shen (1997) obtained the global convergence rates of the quantile smoothing splines and Portnoy (1997) presented some local asymptotic properties.

When the response Z depends on two variables X and Y, we first assume, for the sake of convenience, that z_{ij} is observed at each (x_i, y_j), $(i = 1, \cdots, m; j = 1, \cdots, n)$, over the partition, $x_1 < x_2 < \cdots < x_m$, and $y_1 < y_2 < \cdots < y_n$. The methodology proposed here, however, applies to more general designs as will become apparent later.

Generalization of smoothing splines to bi- or multi-variate cases is not always straightforward. The form of the solution often depends on the roughness penalty used in the optimization process. For conditional mean estimation, a natural generalization of (2) is to solve the following:

$$\min_g \sum_{i=1}^m \sum_{j=1}^n \left(z_{ij} - g\left(x_i, y_j\right)\right)^2 + \lambda J\left(g\right) \tag{3}$$

where the penalty is defined as

$$J(g) = \int\int_{R^2} \left\{g_{xx}^2 + 2g_{xy}^2 + g_{yy}^2\right\} dxdy. \tag{4}$$

The solution is a natural thin-plate spline. See Green and Silverman (1994, p. 144), and Wahba (1990, p. 30) for more details.

For conditional quantile estimation, we propose the *bivariate quantile smoothing spline*, $\hat{g}_\tau\left(x, y\right)$, which solves

$$\min_g \sum_{i=1}^m \sum_{j=1}^n \rho_\tau\left(z_{ij} - g\left(x_i, y_j\right)\right) + R\left(g\right) \tag{5}$$

for the roughness penalty

$$R(g) = \lambda_1 \sum_{i=1}^m V_y\left(g_y\left(x_i, \cdot\right)\right) + \lambda_2 \sum_{j=1}^n V_x\left(g_x\left(\cdot, y_j\right)\right) \tag{6}$$

where λ_1 and λ_2 are smoothing parameters, $V_y\left(h\right)$ is the total variation of $h\left(x, \cdot\right)$ along the y direction and $V_x\left(h\right)$ is the total variation of $h\left(\cdot, y\right)$ along the x direction. We will show in Section 2 that the optimal solution is a bi-linear tensor-product spline.

The natural thin-plate splines are rotation invariant. The bi-linear tensor-product splines are, on the other hand, scale invariant. Optimality of the natural thin-plate spline in (3) holds only if the penalty is integrated over the complete two-dimensional Euclidean space. In practical problems where the design variables (X, Y) usually lie in a bounded region, the thin-plate spline is nowhere near optimal. For example, consider interpolating the four points $(0,0,0),(0,1,0),(1,1,1),(1,0,0)$. Direct calculations show that when integrated over the unit square, the natural thin-plate spline has $J(g) = 2\pi/\log 2 > 9$ as compared to $J(g^*) = 2$ for the bi-linear function $g^*(x,y) = xy$.

The L_1 nature of the problem in (5) allows us to also compute the bivariate quantile smoothing splines using linear programming techniques. The possibility of using linear programming methods for data smoothing was introduced by Schuette (1978) in a remarkable paper in the actuarial literature. Recently Esther Portnoy (1994 & 1995) extended it to bivariate problems for smoothing actuarial tables.

In Section 2, we study optimality properties of bivariate quantile smoothing splines. Their computational aspects are discussed in Section 3. *Penalized (bivariate) quantile B-splines* are proposed as approximate solutions for moderately large data sets. We also discuss the choice of smoothing parameters through a Schwarz-type information criterion. An example and a Monte Carlo simulation are given in Section 4 to illustrate the applications of our methodology, and some concluding remarks proffered in Section 5.

2 OPTIMALITY PROPERTY

We consider the special case where the covariates are observed on a rectangular grid $\mathcal{G} = (x_1, x_2, \cdots, x_m) \times (y_1, y_2, \cdots, y_n)$, say on $[0,1]^2$, for convenience and without much loss of generality. The grid provides a natural $(m-1)$ by $(n-1)$ partition of the unit square. Grid designs often appear in controlled experiments or when data are rounded or grouped. We must emphasize that all the results in this paper remain valid under more general designs where observations are not necessary available at every single grid point. In such situations, the rectangular grid $\mathcal{G}$ will be the tensor-product mesh formed by the unique values of both covariates.

To be specific, we assume

C1. The τ-th bivariate conditional quantile function, g_τ, belongs to the class $\mathcal{U}$ of functions, g, that are continuous on $[0,1]^2$, twice continuously differentiable on each sub-rectangle $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$ and the partial derivatives g_x and g_y have bounded total variations in x and y respectively.

Analogous to (4), we could measure roughness by

$$K(g) = \int_0^1 \{V_x(g_x) + V_x(g_y)\}\, dy + \int_0^1 \{V_y(g_x) + V_y(g_y)\}\, dx. \tag{7}$$

If g is twice differentiable on $[0,1]^2$, $K(g) = \int_0^1 \int_0^1 \{|g_{xx}| + 2|g_{xy}| + |g_{yy}|\}\, dxdy$. Unfortunately, the smoothest interpolant for (7) in a bivariate design is not in the space of bi-linear tensor-product splines with knots located at the design points. We illustrate this by constructing an example of two bi-linear tensor-product splines, $g(x, y)$ and $\tilde{g}(x, y)$, interpolating the same values

$$\begin{pmatrix} 2 & 0 & -1 & -1 \\ 3 & 1 & 1 & 3 \\ -1 & -1 & 0 & 2 \end{pmatrix}$$

on the 3×4 grid $\mathcal{G} = (0,1,2,3) \times (0,1,2)$ over the domain $\mathcal{D}= [0,3] \times [0,2]$. In particular, g is the bi-linear tensor-product spline with knots at each point in $\mathcal{G}$ while $\tilde{g}$ is the bi-linear tensor-product spline with knots in $\mathcal{H} \equiv (0, 1.5, 3) \times (0,1,2)$ defined over $\mathcal{D}$ by

$$\tilde{g}(i, j) = \begin{pmatrix} 3 & -1 & -1 \\ 3 & 0 & 3 \\ -1 & -1 & 2 \end{pmatrix} \qquad (i, j) \in \mathcal{H}$$

Straightforward calculations (using peicewise linearity) show

$$\int V_x(g_x)\, dy = \int V_x(\tilde{g}_x) dy, \quad \int V_x(g_y) dy = \int V_x(\tilde{g}_y) dy, \quad \int V_y(g_x) dx = \int V_y(\tilde{g}_x) dx$$

but $\int V_y(g_y) dx > \int V_y(\tilde{g}_y) dx$. Therefore, $\tilde{g}$ is strictly better than g in terms of roughness based on any linear combination of the integrals in (7) (giving positive weight to the last integral) though g and $\tilde{g}$ interpolate the same values on $\mathcal{G}$.

To overcome the deficiency that (7) does *not* lead to a simple characterization, we decompose the roughness of a surface over $\mathcal{G}$ into roughness inside each sub-rectangular block plus roughness along block boundaries. Specifically, we introduce

$$\tilde{K}(g) = \lambda_1 \sum_{i=1}^{m} V_y\left(g_y\left(x_i, .\right)\right) + \lambda_2 \sum_{j=1}^{n} V_x\left(g_x\left(., y_j\right)\right) + \sum_{i=1}^{m} \sum_{j=1}^{n} \delta_{ij} K\left(g_{ij}\right) \tag{8}$$

where g_{ij} is the function g restricted to $(x_i, x_{i+1}) \times (y_j, y_{j+1})$, and λ_1, λ_2 and δ_{ij} are nonnegative constants. We now have

Theorem 1: Under condition **C1,** the function g that solves

$$\min_{g \in \mathcal{U}} \sum_i \sum_j \rho_\tau\left(z_{ij} - g\left(x_i, y_j\right)\right) + \tilde{K}\left(g\right)$$

is a bi-linear tensor-product spline with knots located at the grid points.

Proof: We first show that on each sub-rectangle, there exists a unique bi-linear function that minimizes $K(g_{ij})$ among all twice differentiable functions that interpolate the same values at the four vertices.

There always exists a bi-linear function g^* matching g at the four vertices of $(a, b) \times (c, d)$. Since $g^*_{xx} = g^*_{yy} = 0$ and $g^*_{xy} = c^*$, a constant over the rectangle, it then holds, for every twice differentiable function f interpolating the same values at the four vertices, that $\int\int |f_{xy}| \geq |\int\int f_{xy}| = |f\left(a, c\right) + f\left(b, d\right) - f\left(b, c\right) - f\left(a, d\right)| = \int\int \left|g^*_{xy}\right|$.

We need to show next that if f minimizes $S(g)$, the bi-linear tensor-product spline g^* that has the same fitted values at (x_i, y_j) satisfies $\tilde{K}\left(g^*\right) \leq \tilde{K}\left(f\right)$. This holds because, by the same arguments used in Koenker *et al.* (1994, p. 675), $V_x\left(g^*_x\left(., y_j\right)\right) \leq V_x\left(f_x\left(., y_j\right)\right)$ for each j and $V_y\left(g^*_y\left(x_i, .\right)\right) \leq V_y\left(f_y\left(x_i, .\right)\right)$ for each i.

Finally, we note that the space of bi-linear tensor-product splines with knots at the grid points is compact as each spline is determined by mn values. The minimum of $S\left(g\right)$ is attainable within the class. The proof of Theorem 1 is then complete. $\diamondsuit$

Again we can view $\tilde{K}(g)$ as an L_1 version of the penalty

$$\begin{aligned}\tilde{J}(g) &= \lambda_1 \sum_{i=1}^{m} \int_0^1 \left(g_{yy}(x_i, y)\right)^2 dy + \lambda_2 \sum_{j=1}^{n} \int_0^1 \left(g_{xx}(x, y_j)\right)^2 dx \\ &\quad + \lambda_3 \int_0^1 \int_0^1 \left(g_{xxyy}(x, y)\right)^2 dx dy \end{aligned} \tag{9}$$

used in Hu and Schumaker (1985) and Schumaker and Utreras (1990). The smoothest interpolant for $\tilde{J}(g)$ is in the class of bi-cubic tensor-product splines. In our case, optimality of bi-linear tensor-product splines is less sensitive to the specification of the roughness penalty on g_{ij}. We can replace $K(g_{ij})$ in (8) by $J(g_{ij})$ or any linear combination of $\int\int g_{xx}^2$, $\int\int g_{xy}^2$ and $\int\int g_{yy}^2$ in (4) without altering the form of the solution. Define

$$\tilde{R}(g) = \lambda_1 \sum_{i=1}^{m} V_y\left(g_y(x_i, .)\right) + \lambda_2 \sum_{j=1}^{n} V_x\left(g_x(., y_j)\right) + \sum_{i=1}^{m} \sum_{j=1}^{n} \delta_{ij} J(g_{ij}), \tag{10}$$

and we have

Theorem 2: Under condition **C1**, the function g that solves

$$\min_{g \in \mathcal{U}} \sum_i \sum_j \rho_\tau\left(z_{ij} - g(x_i, y_j)\right) + \tilde{R}(g)$$

is a bi-linear tensor-product spline with knots located at the grid points.

Proof: Similar to the proof of *Theorem 1* using the approach of Green and Silverman (1994, p.144).

Since the optimal solution is a bi-linear tensor-product spline that can be completely defined by its values at the boundaries $x = x_i$ and $y = y_j$, we can simplify the roughness measure in both Theorem 1 and Theorem 2 even further. The inclusion of $K(g_{ij})$ in (8) and $J(g_{ij})$ in (10) favors block-wise additivity and linearity, which is generally not desirable in surface fitting. If the true quantile surface is in fact additive and linear, the roughness measured along the block boundaries should automatically favor the desired structure without even using $K(g_{ij})$ or $J(g_{ij})$. This is why we suggest using (6) as the roughness penalty.

3 IMPLEMENTATION

A major virtue of the L_1 roughness penalty of (6) is that we can compute the bivariate quantile smoothing spline as a linear program. Given a general set of knots, $0 = u_0 = u_1 < \cdots < u_{k_m} = u_{k_m+1} = 1$ and $0 = v_0 = v_1 < \cdots < v_{k_n} = v_{k_n+1} = 1$, let $\Delta_x = \{u_k\}_{k=1}^{k_m}$ and $\Delta_y = \{v_l\}_{l=1}^{k_n}$ be partitions, and $\tilde{\Delta}_x = \{u_k\}_{k=0}^{k_m+1}$ and $\tilde{\Delta}_y = \{v_l\}_{l=0}^{k_n+1}$ be extended partitions on the domains of x and y respectively. Also let $S_{\tilde{\Delta}_x} = \left\{s|s(x) = \sum_{k=0}^{k_m+1} \gamma_k B_k(x)\right\}$ and $S_{\tilde{\Delta}_y} = \left\{s|s(y) = \sum_{l=0}^{k_n+1} \gamma_l B_l(y)\right\}$ be the spaces of linear B-splines where B_k and B_l are the basis splines as defined in Schumaker (1981). Denote $\mathbf{\Delta} = \Delta_x \otimes \Delta_y$ as the tensor-product mesh and $\tilde{\mathbf{\Delta}} = \tilde{\Delta}_x \otimes \tilde{\Delta}_y$ as the extended tensor-product mesh. The space of bi-linear tensor-product B-splines for $(x, y) \in [0,1]^2$ is then defined as

$$\mathbf{S}_{\tilde{\mathbf{\Delta}}} = S_{\tilde{\Delta}_x} \otimes S_{\tilde{\Delta}_y} = \left\{ s|s(x,y) = \sum_{k=0}^{k_m+1} \sum_{l=0}^{k_n+1} \gamma_{kl} B_k(x) B_l(y) \right\}.$$

When the tensor-product mesh is equivalent to the design grid, i.e. $\mathbf{\Delta} \equiv \mathcal{G}$, $\mathbf{S}_{\tilde{\mathbf{\Delta}}}$ provides a B-splines representation for our bivariate quantile smoothing splines.

Now we can express (5) as

$$\begin{aligned} \min_{s \in \mathbf{S}_{\tilde{\mathbf{\Delta}}}} \quad & \sum_{i=1}^{m} \sum_{j=1}^{n} \rho_\tau(z_{ij} - s(x_i, y_j)) + \lambda_1 \left(\sum_{i=1}^{m} \sum_{j=2}^{n} |s_y(x_i, y_{j+1}) - s_y(x_i, y_j)| \right) \\ & + \lambda_2 \left(\sum_{j=1}^{n} \sum_{i=2}^{m} |s_x(x_{i+1}, y_j) - s_x(x_i, y_j)| \right). \end{aligned} \tag{11}$$

Writing the $(m+2)(n+2)$ vector of parameters for the bi-linear tensor-product B-splines as $\gamma = (\gamma_{0,0}, \cdots, \gamma_{0,m+1}, \cdots, \gamma_{n+1,0}, \cdots, \gamma_{n+1,m+1})^T$, we can rewrite (11) into the following compact L_1 formulation:

$$\min_{\gamma \in R^{(n+2)(m+2)}} \sum_{i=1}^{3mn-m-n} \omega_i |\tilde{y}_i - \tilde{x}_i \gamma| \tag{12}$$

where $\tilde{y} = (z_{11}, \cdots, z_{1n}, \cdots, z_{m1}, \cdots, z_{mn}, \mathbf{0})^T$ is a $(3mn - m - n)$ *pseudo-response* vector, ω_i is a weight assigned to the ith observation and $\tilde{x}_i$ is the ith row of a $(3mn - m - n) \times$

$(n+2)(m+2)$ *pseudo-design* matrix $\tilde{X}$, the details of which are explicitly given in the Appendix.

It is well known that (12) can be solved as a linear program. FORTRAN subroutines utilizing the algorithm described in Ng (1996) and an S(plus) interface are available from the authors upon request.

It is obvious from the pseudo-design matrix $\tilde{X}$ of (13) that computation in (12) requires $O\left((mn)^2\right)$ operations. This can create storage and memory problems even for modestly big data sets. Our asymptotic analysis, confirmed by real data experience, indicates that the effective dimension (measured by the number of interpolated points) as well as the required number of B-spline knots in each covariate for an "optimal" fit is typically small. We, therefore, suggest approximating the solution using the penalized B-spline approach with a smaller number of knots in the order of $k_m = O\left(m^{1/3}\right)$ and $k_n = O\left(n^{1/3}\right)$. A simple but usually effective method is to use no more than ten equally-spaced knots either in Euclidean distance or in percentile ranks of each covariate unless the function is extremely wiggly.

An important issue in the implementation of bivariate quantile smoothing splines is the choice of the smoothing parameters λ_1 and λ_2. As in the univariate case, these values determine the effective dimension of the fit. We can view the choice of smoothing parameters as a model selection problem where one needs to balance between goodness-of-fit and complexity of the model. Denote the total number of observations as N. Following the successful use of the Schwarz's type information criterion in Koenker et al. (1994) for the univariate quantile smoothing splines, we choose $\lambda_N = \lambda_1 = \lambda_2$ to minimize

$$SIC\left(\lambda_N\right) = \log\left(\sum_j \sum_i \rho_\tau\left(z_{ij} - \hat{g}_\tau\left(x_i, y_j\right)\right)\right) + (1/2)\, p_{\lambda_N} \log\left(N\right)/N,$$

where the effective dimension, p_{λ_N}, which plays the role of model dimension in parametric regression, is the number of points interpolated by $\hat{g}_\tau$. The resulting choice of λ_N is scale invariant.

In the case of $\tau = .5$, the information criterion we use may be motivated as the Gaussian likelihood based information criterion of Schwarz (1978) where the root mean square error is replaced by a robust alternative of the average of the absolute residuals as a measure of fidelity to the observed data. Similar ideas can be applied to other quantiles.

Generalized cross-validation (GCV) is commonly used in the literature for the least squares based smoothing splines. Asymptotically, GCV is equivalent to the Akaike information criterion (AIC), which is similar to SIC for modest sample sizes. It is not, however, as direct to motivate the projection based GCV for the L_1-type objective function we use.

Asymptotically one needs to choose λ_N to minimize SIC over a reasonable range of the smoothing parameter to ensure consistency of the resulting quantile estimates. In practice, we usually do so without explicit bounds. Instead, we suggest a rather subjective approach: if the spline with a λ_N chosen by SIC appears far too smooth, then one should examine the SIC more carefully to see if a smaller λ_N corresponding to a local minimum might yield a better fit. On the other hand, the obvious minimum of SIC at $\lambda_N = 0$ should also be avoided.

We also note that SIC is a step function in λ_N. Sensitivity analysis (parametric programming, see Gal, 1979) can then be used to obtain the whole spectrum of distinct bivariate quantile smoothing splines corresponding to all the *finitely many* λ_N. More details may be found in Ng (1996).

4 EXAMPLES

We use some real and simulated data sets to demonstrate bivariate quantile smoothing splines in action. In all cases, we let $\lambda_N = \lambda_1 = \lambda_2$ as discussed in Section 3. The data in the first example do not fall on a regular grid. The second example with simulated data is used to evaluate the performance of the median smoothing spline in estimating the true regression function.

Example 1: Consider the annual salary (in thousands of dollars) of baseball players as a function of performance and seniority. The data, available from *lib.stat.cmu.edu*, consist of 263 North American Major League players for the 1986 season. We use the number of home runs (HR) in the latest year to measure performance, and the number of the years (YEAR) as seniority variable. Figure 1 contains the scatter plots of the data. There is a rather low signal-to-noise ratio and severe non-additivity in the data. Furthermore, observations are available only on part of the grid as in most observational data.

One could use a grid formed by the unique observed values in both covariates. There are 36 and 21 unique values in HR and YEAR respectively. This gives rise to a linear program with a larger number (874) of parameters than observations (263). So we chose to compute the penalized B-spline approximation with 10 equally spaced knots in each variable.

The approximate $\tau = (.25, .5, .75)$-th penalized bivariate quantile B-splines are plotted in Figures 2 – 4. Only the fitted surfaces above the convex hull of the data are plotted. The smoothing parameters chosen by the SIC criterion are $\lambda_N = (4.3, 1.7, 1.2)$ respectively. The approximate median smoothing spline indicates that salaries increase with YEAR, peak around the tenth year, and then begin to decline. The rates of increase and decrease are higher for better players. Since there is only one player with a career of longer than twenty years, the fitted surface for $YEAR > 20$ should not be taken too seriously. The third quartile shows a similar pattern except that salaries of average players peak at around the fifteenth year and better players enjoy a slower rate of salary decline than those who are paid a median salary. The first quartile fit is quite uneventful as expected.

For the purpose of comparison, we also computed Cleveland and Grosse's (1991) robust *loess* fit with span = .45 and Gu and Wahba's (1993) thin plate spline with smoothing parameter chosen by the generalized maximum likelihood (GML) criterion. The robust *loess* is quite similar to our median spline, so the fitted function is not shown. The least squares based thin plate spline, shown in Figure 5, displays the usual sensitivity of least squares based smoothers to the outliers that are singled out in the scatter plots.

Example 2: To assess the accuracy of our method, we present a small-scale simulation experiment. Since alternative algorithms for bivariate quantile estimation with automatic choices of smoothing parameters are not readily available, we restrict the performance comparison to the median.

We draw random samples of size N=100 from the following model

$$z_{ij} = sin\,(3\pi x_i)\,cos\,(\pi y_j) + e_{ij}/3, \quad i = 1, \cdots 10, \; j = 1, \cdots 10$$

where $x_i = y_i = i/10$ and e_{ij}'s are independent and identically distributed errors from (i) standard normal, Φ; (ii) 5% contaminated normal, $CN_{.05} = .95\Phi(x) + .05\Phi(x/5)$; (iii) 10% contaminated normal, $CN_{.10} = .9\Phi(x) + .1\Phi(x/5)$ and (iv) Student's t with three degrees of freedom, t_3.

For regression models with symmetric errors, both conditional mean and median functions represent the true regression function in the usual sense. We compared the mean squared errors,

$$MSE = (1/N) \sum_{i,j} \left(\hat{g}_\tau\,(x_i, y_j) - g\,(x_i, y_j)\right)^2,$$

of our bivariate median smoothing spline (*bmss*), the least squares based *loess*, thin-plate splines and Friedman's (1991) *MARS*. A total of 1000 samples were included in the study. The means and standard errors (in parentheses) of the MSE are reported in Table 1. For *bmss*, we used the design grid as the tensor-product mesh. The smoothing parameter (span) for *loess* was set to 0.20 for the normal distribution, 0.30 for t_3 and the two contaminated normal distributions to make *loess* as competitive as possible. We adopted all the default

Smoothers	Distributions			
	Φ	$CN_{.05}$	$CN_{.10}$	t_3
bmss	0.0529	0.0577	0.0637	0.0690
	(0.00038)	(0.00054)	(0.00066)	(0.00070)
loess	0.0424	0.0634	0.0915	0.0800
	(0.00028)	(0.00086)	(0.00129)	(0.00176)
thin plate	0.0381	0.0647	0.0922	0.0827
(GCV)	(0.00042)	(0.00101)	(0.00158)	(0.00206)
MARS	0.0447	0.0918	0.1400	0.1150
	(0.00044)	(0.00154)	(0.00221)	(0.00289)

Table 1: Means and Standard Errors of MSE of Various Bivariate Smoothers

settings in *MARS* except the degrees-of-freedom (*df*) charged for knot optimization was set to 1.0 to obtain the best possible performance from $MARS$.

It is no surprise that the mean regression estimators (*loess*, *thin plate* splines and *MARS*) are more efficient for the normal error distribution. The bivariate median smoothing spline is clearly more robust than the mean estimators if the e_{ij}'s are drawn from any of the thicker-tailed distributions.

5 CONCLUDING REMARKS

Motivated by the optimality property of bi-linear tensor-product splines, we propose the bivariate quantile smoothing splines as estimators for bivariate conditional quantile functions. For large data sets, a smaller number of equally-spaced knots in either Euclidean distance or percentile ranks of each covariate may be used to obtain a penalized bivariate B-spline approximation. Efficient computation by linear programs and the choice of the smoothing parameters using a Schwarz-type information criterion are also demonstrated. Our experience with real and simulated data has been positive and encouraging.

If the sample is drawn from a regression model $z = g_\tau(x, y) + e$ where the error distribution e has zero as the τ-th quantile, then the bivariate quantile smoothing spline defined by (5) and (6) is a consistent estimator of g_τ in the area where the design points are asymptotically dense, provided that the smoothing parameters are so chosen that $\lambda_1 = o(m^{1/3})$ and $\lambda_2 = o(n^{1/3})$. Details on the asymptotic properties of the estimate will appear elsewhere. We should also point out that our experience with linear regression quantiles suggests that the quantile smoothing splines are consistent for a broader class of models than the regression model indicated here with identically distributed errors.

To see how the bivariate smoothing splines would behave under non-rectangular designs, we performed an almost identical simulation to that of Example 2 with the only exception that the (x_i, y_i) pairs were drawn from a uniform distribution on a right-angle triangle. The results are similar to those presented in Table 1. However, if the design points fall in regions of very different shapes, one may prefer to perform a transformation of variables to map the domain of the function into a rectangle before using tensor-splines.

Finally, we wish to point out an alternative method of estimating conditional quantile functions. Stone (1991) considered the approach of estimating the entire set of conditional density functions and using this for simultaneous estimation of the conditional quantiles. This approach will automatically preserve the natural ordering of the quantile functions. The method we choose in the present paper, which is computationally less demanding, estimates the quantile functions directly without having to assume that conditional densities are smooth.

APPENDIX: MORE DETAILS FOR SECTION 3

The pseudo-design matrix $\tilde{X}$ in Section 3 has $(3mn - m - n)$ rows and $(n+2)(m+2)$ columns in the form of

$$\tilde{X} = \begin{bmatrix} \mathbf{B} \\ \mathbf{V}^y \\ \mathbf{V}^x \end{bmatrix} \tag{13}$$

in which

$$\mathbf{B} = \left[\begin{array}{ccccccc} \mathbf{B}_{11}^T & \cdots & \mathbf{B}_{1n}^T & \cdots & \mathbf{B}_{m1}^T & \cdots & \mathbf{B}_{mn}^T \end{array}\right]^T$$

is an $(mn) \times (n+2)(m+2)$ matrix with rows

$$\begin{aligned} \mathbf{B}_{ij} &= (B_0(x_i) B_0(y_j), \cdots, B_0(x_i) B_{n+1}(y_j), \cdots, B_{m+1}(x_i) B_0(y_j), \cdots, \\ &\quad B_{m+1}(x_i) B_{n+1}(y_j)), \end{aligned}$$

$$\mathbf{V}^y = \left[\begin{array}{ccccccc} \mathbf{V}_{11}^{yT} & \cdots & \mathbf{V}_{1(n-1)}^{yT} & \cdots & \mathbf{V}_{m1}^{yT} & \cdots & \mathbf{V}_{m(n-1)}^{yT} \end{array}\right]^T$$

is an $(mn-m)$ matrix with rows

$$\begin{aligned} \mathbf{V}_{i,j}^y &= \Big[B_0(x_i)\left\{B_0^{(1)}(y_{j+1}) - B_0^{(1)}(y_j)\right\}, \cdots, \\ &\quad B_0(x_i)\left\{B_{n+1}^{(1)}(y_{j+1}) - B_{n+1}^{(1)}(y_j)\right\}, \cdots, \\ &\quad B_{m+1}(x_i)\left\{B_0^{(1)}(y_{j+1}) - B_0^{(1)}(y_j)\right\}, \cdots, \\ &\quad B_{m+1}(x_i)\left\{B_{n+1}^{(1)}(y_{j+1}) B_{n+1}^{(1)}(y_j)\right\}\Big], \end{aligned}$$

and

$$\mathbf{V}^x = \left[\begin{array}{ccccccc} \mathbf{V}_{11}^{xT} & \cdots & \mathbf{V}_{(m-1)1}^{xT} & \cdots & \mathbf{V}_{1n}^{xT} & \cdots & \mathbf{V}_{(m-1)n}^{xT} \end{array}\right]^T$$

is an $(mn-n)$ matrix with rows

$$\begin{aligned} \mathbf{V}_{i,j}^x &= \Big(\left\{B_0^{(1)}(x_{i+1}) - B_0^{(1)}(x_i)\right\} B_0(y_j), \cdots, \\ &\quad \left\{B_0^{(1)}(x_{i+1}) - B_0^{(1)}(x_i)\right\} B_{n+1}(y_j), \cdots, \\ &\quad \left\{B_{m+1}^{(1)}(x_{i+1}) - B_{m+1}^{(1)}(x_i)\right\} B_0(y_j), \cdots, \\ &\quad \left\{B_{m+1}^{(1)}(x_{i+1}) B_{m+1}^{(1)}(x_i)\right\} B_{n+1}(y_j)\Big). \end{aligned}$$

The weight ω is a $(3mn-m-n)$-vector as

$$\omega = \left[\begin{array}{c} 1/2 + (\tau - 1/2)\, sgn\,(\tilde{y}_1 - \tilde{x}_1\gamma) \\ \vdots \\ 1/2 + (\tau - 1/2)\, sgn\,(\tilde{y}_{mn} - \tilde{x}_{mn}\gamma) \\ \lambda_1 \mathbf{1} \\ \lambda_2 \mathbf{1} \end{array}\right],$$

where $\mathbf{1}$ is a vector of 1's.

REFERENCES

Antoch, J. and Janssen, P. (1989) Nonparametric Regression M-quantiles. *Statistics & Probability Letters*, 8, 355-362.

Bartels, R. and Conn, A. (1980) Linearly Constrained Discrete L_1 Problems. *ACM Transactions on Mathematical Software*, 6, 594-608.

Bhattacharya, P.K. and Gangopadhyay, A.K. (1990) Kernel and Nearest-neighbor Estimation of a Conditional quantile. *Annals of Statistics*, 18, 1400-1415.

Chaudhuri, P. (1991) Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation. *Annals of Statistics*, 19, 760-777.

Cleveland, W.S. and Grosse, E. (1991) Computational Methods for Local Regression. *Statistics and Computing*, 1, 47-62.

Efron, B. (1991) Regression Percentiles Using Asymmetric Squared Error Loss. *Statistica Sinica*, 1, 93-125.

Friedman, J.H. (1991) Multivariate Adaptive Regression Splines. *Annals of Statistics*, 19, 1-67.

Gal, T. (1979) *Postoptimal analyses, parametric programming and related topics.* New York: McGraw-Hill.

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

Gu, C. and Wahba, G. (1993) Smoothing Spline ANOVA With Component-wise Bayesian 'Confidence Intervals'. *Journal of Computational and Graphical Statistics*, 2, 97-117.

Härdle, W. (1989) *Applied nonparametric regression.* Cambridge: Cambridge University Press.

He, X. (1997) Regression quantiles without crossing. *The American Statistician*, 51, 186-192.

He, X. and Shi, P. (1994) Convergence Rate of B-spline Estimators of Nonparametric Conditional Quantile Function. *Journal of Nonparametric Statistics*, 3, 299-308.

He, X. and Shi, P. (1996) Bivariate tensor-product B-splines in a partly linear model. To appear in *Journal of Multivariate Analysis*.

Hendricks, W. and Koenker, R. (1992) Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity. *Journal of the American Statistical Association*, 87, 58-68.

Hu, C.L. and Schumaker, L.L. (1985) Bivariate Natural Spline Smoothing. In *Delay Equations, Approximation and Application*, (eds. G. Meinardus and G. Nürnberger). Boston: Birkhäuser Verlag, Basel.

Kimeldorf, G.S. and Wahba, G. (1970) A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. *Annals of Mathematical Statistics*, 41, 495-502.

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

Ng, P. (1996) An algorithm for quantile smoothing splines. *Computational Statistics and Data Analysis*, 22, 99-118.

Portnoy, E. (1994) Bivariate Schuette Graduation. In *ARCH: Proceedings 28th Actuarial Research Conference*, 1994.1, pp. 127-134. Society of Actuaries.

Portnoy, E. (1995) Bivariate Schuette Graduation of Race-specific Mortality Rates. In *ARCH: Proceedings 29th Actuarial Research Conference*, 1995.1, pp. 131-140. Society of Actuaries.

Portnoy, S. (1997). Local Asymptotics for Quantile Smoothing Splines. *Annals of Statistics*, 25, 414-434.

Schuette, D.R. (1978) A Linear Programming Approach to Graduation. *Transaction of the Society of Actuaries*, 30, 407-445.

Schumaker, L.L. (1981) *Spline Functions: Basic Theory.* New York: John Wiley & Sons.

Schumaker, L.L. and Utreras, F. I. (1990) On Generalized Cross Validation for Tensor Smoothing Splines. *SIAM Journal of Scientific and Statistical Computing*, 11, 713-731.

Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464.

Shen, X. (1997) On the method of penalization. *Statistica Sinica*, in print.

Stone, C.J. (1991) Asymptotics for doubly flexible logspline response models. *Annals of Statistics*, 16, 1832-1854.

Stone, C.J. (1994) The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation. *Annals of Statistics*, 22, 118-171.

Wahba, G. (1990) *Spline Models for Observational Data.* Philadelphia: SIAM, Society for Industrial and Applied Mathematics.

Wang, F.T. and Scott, D.W. (1994) The L_1 Method for Robust Nonparametric Regression. *Journal of the American Statistical Association*, 89, 65-76.

Welsh, A.H. (1996) Robust Estimation of Smoothing Regression and Spread Functions and Their Derivatives. *Statistica Sinica*, 6, 347-366.

White, H. (1990) Nonparametric Estimation of Conditional Quantiles Using Neural Networks. *Computing Science and Statistics: Proceedings of the Symposium on the Interface*, 22, 190-199.

Zelterman, D. (1990) Smooth Nonparametric Estimation of the Quantile Function. *Journal of Statistical Planning and Inference*, 26, 339-352.

Figure 1: Scatter Plots of Baseball Data

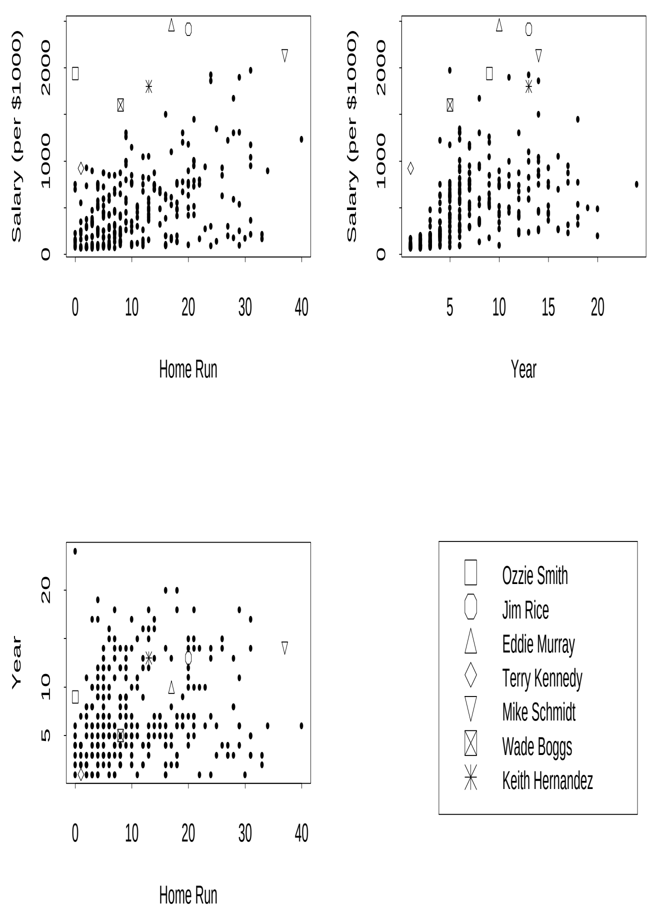

Figure 2: First Quartile Smoothing Spline for Baseball Data

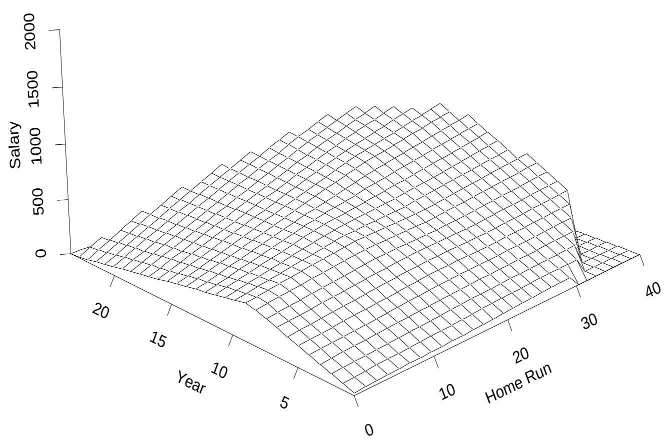

Figure 3: Median Smoothing Spline for Baseball Data

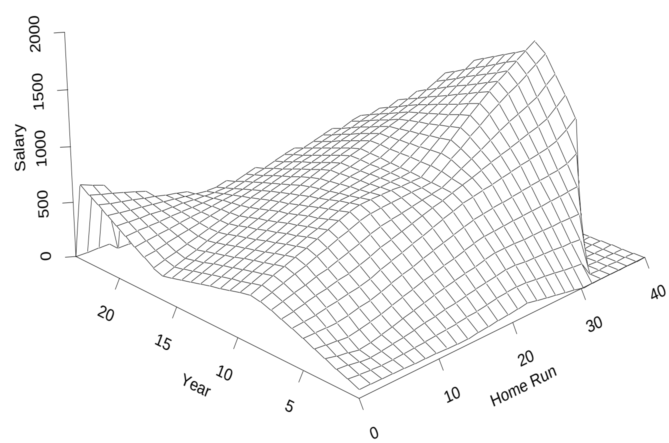

Figure 4: Third Quartile Smoothing Spline for Baseball Data

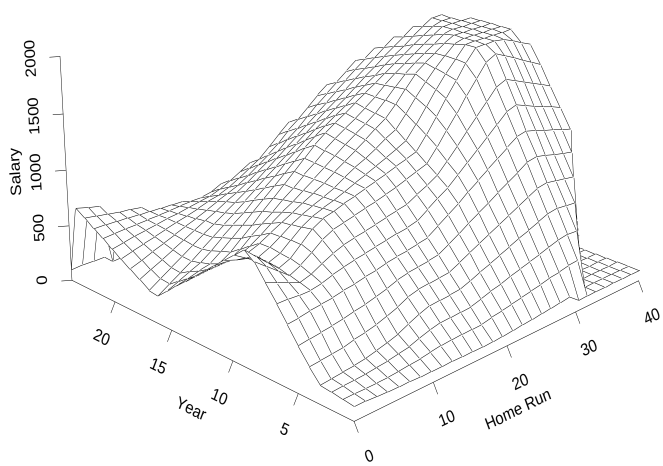

Figure 5: Thin-plate Spline Fit of Baseball Data