Faculty Working Paper 91-0153

Quantile Smoothing Splines

Roger Koenker
Department of Economics
University of Illinois

Pin Ng
Department of Economics
University of Houston

Bureau of Economic and Business Research
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 91-0153

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

July 1991

Quantile Smoothing Splines

Roger Koenker
Department of Economics
University of Illinois at Urbana-Champaign

Pin Ng
Department of Economics
University of Houston

Quantile Smoothing Splines[1]

Roger Koenker [a] and Pin Ng [b]

[a] Department of Economics, University of Illinois, Champaign, Illinois 61820, U.S.A.

[b] Department of Economics, University of Houston, Houston, TX 77204-5882, U.S.A.

Abstract

Quantile smoothing splines defined as solutions to

$$\min_{g \in C^1} \sum \rho_\alpha(y_i - g(x_i)) + \lambda \int_0^1 |g''(x)|\,dx$$

with $\rho_\alpha(u) = (\alpha - I(u < 0))u$ are proposed as a simple, nonparametric approach to estimating conditional quantile functions. We show that solutions, $\hat{g}$, are parabolic splines, i.e. piecewise quadratics, on the mesh $\{x_1, ..., x_n\}$, and may be computed by standard l_1-type linear programming techniques. At $\lambda = 0$, $\hat{g}$ interpolates the α^{th} quantiles at the distinct design points, and for λ sufficiently large $\hat{g}$ is the linear regression quantile fit to the observations. The entire path of solutions, in the penalty parameter λ, may be efficiently computed by linear parametric programming methods.

1. INTRODUCTION

In practice nonparametric regression is virtually always the estimation of flexible models *for conditional mean functions.* Some recent theory has advanced robustified methods of estimating alternative measures of conditional central tendency, e.g., Cox (1983), Utreras (1981), and Härdle (1990). But aspects other than the central tendency of conditional distributions are frequently of substantial interest. What, for example, is the seasonal pattern of extreme temperatures, water levels, or pollution readings? Are estimated trends in mean incomes or SAT performance consistent with trends in extreme quantiles? Efron (1991) has persuasively argued that regression percentiles have a critical role to play in regression diagnostics for generalized linear models.

[1]The authors wish to thank Cornelius Gutenbrunner, Steve Portnoy and Mike Osborne for helpful discussions, and Esther Portnoy for pointing out the important Schuette reference. Research support from NSF SES 89-22472 is also gratefully acknowledged.

Several authors have proposed methods for nonparametric estimation of such conditional quantile functions. Troung (1989) following the nearest neighbor approach of Stone (1977), Samanta (1989) and Antoch and Janssen (1989) using kernel methods and White (1990) using neural networks have all suggested methods for nonparametric estimation of conditional quantile functions. Extending the parametric methods for estimating linear models for conditional quantiles of Koenker and Bassett (1978), Hendricks and Koenker (1990) discuss sieve-type regression spline models and apply them to electricity demand data. Cox (1988) and Jones (1988), reviving an idea of Bloomfield and Steiger (1983), have suggested estimating quantile smoothing splines which minimize

$$\sum \rho_\alpha(y_i - g(x_i)) + \lambda \int (g''(x))^2 dx \tag{1.1}$$

where $\rho_\alpha(u) = (\alpha - I(u < 0))u$ is the Czech function of Koenker and Bassett (1978). Here the parameter $\alpha \in [0, 1]$ controls the quantile of interest, while $\lambda \in \mathbf{R}_+$ controls the smoothness of the resulting estimate, thus generalizing the extensive literature on l_2 smoothing splines pioneered by Wahba (1990). This is an intriguing idea, and has also been mentioned, for example, in Eubank (1988) and Utreras (1981) in the median $\rho_{1/2}(u) = |u|$ case. However, the resulting quadratic program poses some serious computational obstacles. Obviously the computational virtues of the piecewise linear form of the first term of the objective function are sacrificed by the quadratic form of the smoothness penalty.

One is thus naturally led to ask: "why not replace $(g''(x))^2$ in the penalty by $|g''(x)|$?" The median special case of this problem has been studied in a remarkable paper by Schuette (1978) in the actuarial literature. We will show, expanding on Schuette's discrete version of the problem using finite differences, that minimizing

$$R_{\alpha,\lambda}[g] = \sum_{i=1}^{n} \rho_\alpha(y_i - g(x_i)) + \lambda \int_0^1 |g''(x)| dx \tag{1.2}$$

over $g \in C^1[0, 1]$ retains the linear programming form of the parametric version of the quantile regression problem and yields solutions which are parabolic, i.e., quadratic, splines. These quantile smoothing splines have several attractive features. Unlike the familiar l_2-smoothing splines, they are qualitatively robust to gross errors in the observations $\{y_i\}$. Like the l_2-case, as $\lambda \to 0$ the estimate becomes rougher: for λ sufficiently large the estimate coincides with the bivariate linear regression quantile estimate, while for λ in a neighborhood of zero the estimate interpolates the α^{th} sample quantiles at each distinct design point. As λ increases the number of interpolated observations decreases monotonically, indeed an important computational aspect of the quantile smoothing problem is that all of the distinct solutions corresponding to distinct λ may be found by performing a sequence of simplex pivots following methods of parametric linear programming, in very much the same manner that all of the distinct regression quantile solutions may be found by parametric variation of α, see Koenker and D'Orey (1987).

2. MEDIAN SMOOTHING SPLINES

Given observations $\{(y_i, x_i) : i = 1, ..., n\}$ with $0 < x_1 < ... < x_n < 1$ consider the problem of minimizing

$$R_\lambda[g] = \sum_{i=1}^{n} |y_i - g(x_i)| + \lambda\int_0^1 |g''(x)|dx \qquad (2.1)$$

over $g \in C^1[0, 1]$, the space of continuous functions on [0, 1] with continuous first derivative.

Definition. A function $g : [0, 1] \to \mathbf{R}$ is a *parabolic spline* with mesh $0 = x_0 < x_1 < \dots < x_n < x_{n+1} = 1$ if $g \in C^1[0, 1]$ and $g(x)$ is piecewise quadratic in the intervals $[x_i, x_{i+1})$, $i = 0, \dots, n$, that is, g has the form

$$g(x) = \alpha_i(x - x_i)^2 + \beta_i(x - x_i) + \gamma_i \quad x_i \le x < x_{i+1} \quad i = 0, \dots, n \qquad (2.2)$$

Theorem. There exists a parabolic spline $\hat{g}$ which solves (2.1).

Proof. Suppose g solves (2.1). We will show that there exists a parabolic spline $\hat{g}$ such that $R[\hat{g}] = R[g]$. Suppose, provisionally, that $sgn(g''(x))$ is constant on the intervals $[x_i, x_{i+1})$, $i = 0, \dots, n$, so we may write

$$\int_0^1 |g''(x)|dx = \sum_{i=0}^{n} |g'(x_{i+1}) - g'(x_i)| \qquad (2.3)$$

Note that we may set $\alpha_i = ½[g'(x_{i+1}) - g'(x_i)]/(x_{i+1} - x_i)$ for $i = 0, \dots, n$ and determine the remaining coefficients of $\hat{g}$ from the conditions

$$\hat{g}(x_i+) = \gamma_i = g(x_i+) \qquad i = 1, \dots, n$$

and

$$\hat{g}'(x_i+) = \beta_i = g'(x_i+) \qquad i = 1, \dots, n$$

The parabolic spline $\hat{g}$ thus constructed is in $C^1[0, 1]$, since g was, and clearly achieves the same value of R. Finally, note that if g'' changes sign on an interval between knots, say $[x_i, x_{i+1})$, the same construction and the fact that $\int |f| \ge \int f$ implies that the resulting $\hat{g}$ satisfies $R[\hat{g}] < R[g]$ which contradicts our hypothesis that g'' could change sign. □

Having established the form of the solution to (2.1), it is straightforward to develop an algorithm to compute $\hat{g}$. Using (2.2) the penalty becomes

$$\int_0^1 |g''(x)|dx = 2\sum_{i=1}^{n-1} h_i |\alpha_i|.$$

where $h_i = x_{i+1} - x_i$, $i = 1, \dots, n-1$. Note that like other smoothing spline problems $\hat{g}$ must be linear in the exterior intervals $[0, x_1)$ and $(x_n, 1]$, since otherwise, the penalty could be reduced without affecting the "fidelity" term. From the continuity conditions

$$\begin{aligned} \alpha_i h_i^2 + \beta_i h_i + \gamma_i &= \gamma_{i+1} \\ 2\alpha_i h_i + \beta_i &= \beta_{i+1} \end{aligned} \qquad i = 1, \dots, n-2$$

eliminating the β_i's we have

$$\alpha_i h_i + \alpha_{i+1} h_{i+1} = \left[\frac{\gamma_{i+2} - \gamma_{i+1}}{h_{i+1}}\right] - \left[\frac{\gamma_{i+1} - \gamma_i}{h_i}\right] \qquad i = 1, \dots, n-2.$$

Including the exterior segments $[0,x_1)$ and $[x_n,1]$ we have $3(n+1)$ free parameters and $2n$ linear continuity constraints plus the 2 conditions that $\alpha_0 = \alpha_n = 0$, which leaves us with $n+1$ free parameters. Thus, given the γ_i's we have one "free" α_i, say, α_1, and we may rewrite (2.1) as

$$\min_{\theta \in \mathbf{R}^{n+1}} \sum |\tilde{y}_i - \lambda_i \tilde{x}_i \theta| \tag{2.4}$$

where $\tilde{y}' = (y', 0'_{n-1})$, $\theta' = (\alpha_1, \gamma_1, ..., \gamma_n)$, $\lambda_i = I(i \le n) + \lambda I(i > n)$,

$$\tilde{X} = \begin{bmatrix} 0 & & \\ \vdots & I_n & \\ 0 & & \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ & A & \\ & & \end{bmatrix}_{(2n-1)\times(n+1)}$$

$A = HD^{-1}B$, $\quad H = diag(h)$,

$$D = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ h_1 & h_2 & 0 & & & & 0 \\ 0 & h_2 & h_3 & & & & 0 \\ \vdots & & & & & & \\ 0 & \cdot & \cdot & \cdot & 0 & h_{n-2} & h_{n-1} \end{bmatrix}_{(n-1)\times(n-1)}$$

and

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & h_1^{-1} & -(h_1^{-1}+h_2^{-1}) & h_2^{-1} & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ & & h_2^{-1} & -(h_2^{-1}+h_3^{-1}) & h_3^{-1} & & & & & \\ \vdots & \vdots & & & & \vdots & & & & \vdots \\ 0 & 0 & & 0 & & 0 & h_{n-2}^{-1} & -(h_{n-2}^{-1}+h_{n-1}^{-1}) & & h_{n-1}^{-1} \end{bmatrix}_{(n-1)\times(n+1)} .$$

Clearly, (2.4) is a conventional l_1-regression problem and can be solved with conventional software; an 'S' function (Becker and Chambers and Wilks (1988)) to accomplish this appears in the Appendix.

Once we notice that the median smoothing spline may be expressed as the solution to a particular l_1-regression problem, a number of other important features of the solution are immediately apparent. We have an l_1 regression with $(n+1)$ parameters and $(2n-1)$ observations; solutions must have $n+1$ residuals which are zero (by complementary slackness) and in our case these zero residuals correspond to either (i) exact interpolation of an observation, so $y_i = \hat{\gamma}_i$ or (ii) linearity of $\hat{g}$ in a particular subinterval of the design mesh, so $\hat{\alpha}_i = 0$ for some i. Obviously, the parameter λ controls the relative "advantage" of these two alternatives. When λ is sufficiently large *all* the $\hat{\alpha}_i$ will be zero and the solution will correspond to the bivariate linear l_1-fit. When λ is sufficiently small all n observations will be interpolated, and all but one of the $\hat{\alpha}_i$'s can be non-zero.

As in any smoothing problem, choice of "bandwidth" -- here represented by the parameter λ -- is critical. For median smoothing splines and quantile smoothing splines in general, this problem is ameliorated by the fact that the whole family of solutions to (2.1) for $\lambda \in [0, \infty)$ may be easily found by parametric linear programming. Suppose at some $\lambda = \lambda_0$ we consider the solution $\hat{g}$ satisfying $\hat{\gamma}_i = \hat{g}(x_i) = y_i$, $i = 1, \ldots, n$. As λ increases from 0, this solution remains optimal as long as the subgradient condition

$$-\lambda_j \le \Big(\sum_{i \in \bar{h}} sgn(\tilde{y}_i - \lambda_i \tilde{x}_i \hat{\theta}) \lambda_i \tilde{x}_i \tilde{X}_h^{-1} \Big)_j \le \lambda_j \tag{2.5}$$

holds. Here, following Bassett and Koenker (1978), we use h to index $n+1$ element subsets of $\mathbf{K} \equiv \{1, 2, \ldots, 2n-1\}$, $\bar{h} = \mathbf{K} - h$, and $\tilde{X}_h$ denotes the submatrix of $\tilde{X}$ with row indices in h. Noting that $\lambda_i = 1$ for $i \in \bar{h}_1$, and for $i \in \bar{h}_2$, it follows that $sgn(\tilde{y}_i - \lambda_i \tilde{x}_i \theta) = sgn(-A_i.\theta)$ since $\lambda > 0$, equation (2.5) may be rewritten as

$$\lambda_j \le r_j + s_j \lambda \le \lambda_j \quad j = 1, \ldots, n+1$$

with r_j and s_j the jth elements of the vectors:

$$r = \sum_{i \in \bar{h}_1} sgn(\tilde{y}_i - \tilde{x}_i \hat{\theta}) \tilde{x}_i \tilde{X}_h^{-1}$$

$$s = \sum_{i \in \bar{h}_2} sgn(\tilde{y}_i - \tilde{x}_i \hat{\theta}) \tilde{x}_i \tilde{X}_h^{-1}$$

where we have decomposed $\bar{h} = \bar{h}_1 \cup \bar{h}_2$ with $\bar{h}_1 \subset \{1, \ldots, n\}$ and $\bar{h}_2 \subset \{n+1, \ldots, 2n-1\}$, so the next λ is

$$\lambda_1 = \min\{ \min_{j \in h_1}\{\max\{\lambda_0, (1-r_j)/s_j, -(1+r_j)/s_j\}\}, \min_{j \in h_2}\{\max\{\lambda_0, -r_j/(1+s_j), r_j/(1-s_j)\}\}\}.$$

Continuing this iteration until $\lambda_{m+1} = \lambda_m$ yields all distinct solutions to the problem (2.1) with the solution for $\lambda \ge \lambda_m$ corresponding to the linear l_1 fit to the observations. An important implication of this fact is that we may initially solve the simpler linear l_1 problem corresponding to $\lambda = \infty$ and gradually relax the roughness penalty with a sequence of simplex pivots, thus avoiding a direct solution of the potentially rather large problem (2.4).

Each "click" to a new solution involves a single simplex pivot of an extremely sparse constraint matrix, and hence solving (2.1) for a broad range of λ appears quite feasible. One interesting aspect of the way that solutions $\hat{\gamma}(\lambda)$ depend upon the penalty parameter λ concerns the number of interpolated points. In the l_2 smoothing spline literature much has been made of the "effective dimensionality" or "degrees of freedom" of the estimated curves corresponding to various λ. Such measures of dimensionality are usually based on the trace of various quasi-projection matrices in the least-squares theory. See Buja, Hastie and Tibshirani (1989) for an extensive discussion. For the median smoothing spline the connection is more direct in the sense that there is an explicit trade-off between the number of interpolated points and the number of linear segments. Since "reasonable" smoothing suggests that the number of interpolated points is small relative to n, it is probably sensible to start the parametric programming at the linear l_1-solution rather than at $\lambda = 0$.

If the design is in "general position" so no two observations share the same design point, there must be at least 2 and at most n interpolated y_i's. Call this number p_λ. Clearly, p_λ is a plausible measure of the effective dimension of the fitted model with penalty parameter λ, and $n - p_\lambda + 1$, which corresponds to the number of linear segments in the fitted function, is a plausible measure of the degrees of freedom of the fit. Such decompositions might be used in conjunction with the function $R[\hat{g}]$ itself to implement data-driven bandwidth choice, for example, along the lines of Akaike (1974) or Schwarz (1978).i

3. QUANTILES AND OTHER EXTENSIONS

Smoothing splines for other quantiles may be estimated by replacing the $|\cdot|$'s in the fidelity term with the Czech function, as in (1.2). It seems natural to maintain the symmetric form of the penalty term. For a fixed quantile the entire path of solutions in the penalty parameter λ can, again, be found by methods of parametric linear programming. Likewise for λ fixed the entire set of distinct solutions for $\alpha \in [0, 1]$ may be found by similar methods.

A simple approach to inference using these methods may be developed along the lines suggested in Hendricks and Koenker (1990) for regression splines. For a given choice of quantile, α, and penalty, λ, we may easily compute the estimated quantile functions for $\alpha \pm h$, λ, providing a tolerance band for $\hat{g}_{\alpha,\lambda}$. A Siddiqui (1960) estimate may then be computed at each design point for the sparsity function and used to construct an estimate of the covariance matrix of θ_n.

There are a number of intriguing extensions incorporating further constraints. Monotonicity and convexity of the fitted function $\hat{g}$ may be readily imposed by imposing further inequality constraints on the parameters of the problem, using the variation reducing properties of splines developed in Schumaker (1981). See Ramsey (1988) for a discussion of some applications of these ideas to the l_2 smoothing spline case. Note however that while adding such inequality constraints to the l_2 problem results in a significant increase in complexity adding linear inequality constraints to the quantile smoothing spline problem does not alter the fundamental nature of the optimization problem to be solved.

Finally, there is the extension of these methods to multivariate settings where the additive spline models of Buja, Hastie and Tibshirani (1989) and others naturally suggest themselves. Clearly the nonlinear character of the present smoothers vitiate the attractive iterative "backfitting" algorithms available in the l_2-case. But feasible estimators may still be possible using a limited number of simplex pivots from an initial linear (in covariates) quantile function estimate.

Consideration of the asymptotic performance of quantile smoothing splines raises some challenging problems which we hope to address in future work. Unfortunately, it does not appear easy to adapt the asymptotic theory of l_2-smoothing splines to the present case. However, the asymptotic linearity results of Cox (1983) for M-type smoothing splines, if extendable to the penalty, may suggest a way forward. The recent work of White(1990) on neural network models for conditional quantile functions appears to offer a relatively straightforward approach to consistency for quantile smoothing splines.

4. SOME PICTURES IN LIEU OF A CONCLUSION

In Figure 4.1 we illustrate several estimates corresponding to various λ's of the median smoothing spline for Example 1 of Schuette (1978) which is a problem in actuarial graduation, i.e. smoothing of life tables. These figures match closely, but not exactly, Schuette's finite difference calculations reported in his Table 2. The interpolating nature of the solutions is immediately apparent in the Figure.

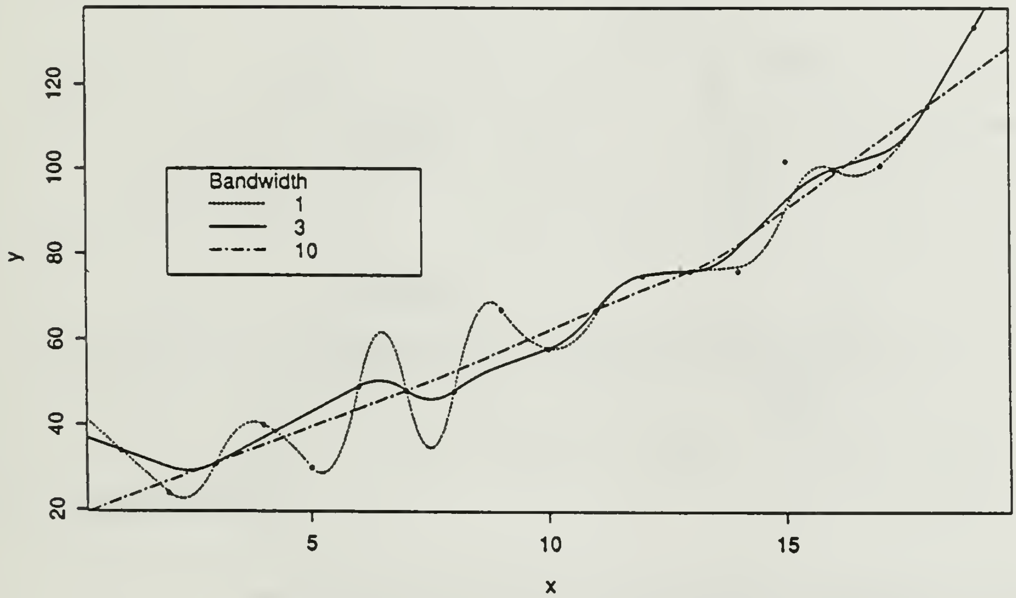

Figure 4.1: Median Smoothing Splines for Schuette's Example 1

In Figure 4.2 we illustrate several estimates of the median smoothing spline for the "motorcycle data" of Schmidt, Mattern and Schuler. See Hardle (1990) and Silverman (1985) for discussions of l_2 smoothing spline analyses of this data set. There is clearly some increase in dispersion in y for $x \in [30, 40]$ and one noticeable difference between these median estimates and the mean estimates is the reduced sensitivity of the former to the outlying points below the curve. Another aspect of this data set which perhaps deserves some comment is the fact that there are multiple observations at some design points, which we have assumed away in the treatment of Section 2. This is rectified in the *S* functions given in the Appendix in the following simple fashion. Knots are placed at each *distinct x* observation, and we parameterize the spline by the values it takes at those knots, and the value of second derivative in the interval between the first two knots. Obviously, the identity matrix in X must be replaced by a block diagonal matrix with columns of 1's in the diagonal blocks. It is this structure which implies as we noted in the introduction that the $\lambda = 0$ interpolative solution passes through the sample quantiles of the distinct design points. Of course, if the x_i are all distinct then we have a fully interpolative solution. An alternative to this approach which has been extensively explored in the spline literature is to introduce multiple knots at the same design point which then has the effect of relaxing the continuity requirement on the function and its derivative. We note in conclusion that the number of interpolated points, p_λ, for the 3 illustrated curves is 41, 26, 23 for $\lambda = .5$, 3, 10, respectively. Results of Portnoy (1984) and Welsh (1989) suggest that p^2/n must tend to zero for consistency, and this suggests since $n = 133$ here that even the $\lambda = 10$ estimate may be undersmoothed. Clearly, a salient advantage of this approach over the more straightforward fixed knot regression spline approach is that the data is allowed to choose regions in which the shape of the function changes rapidly.

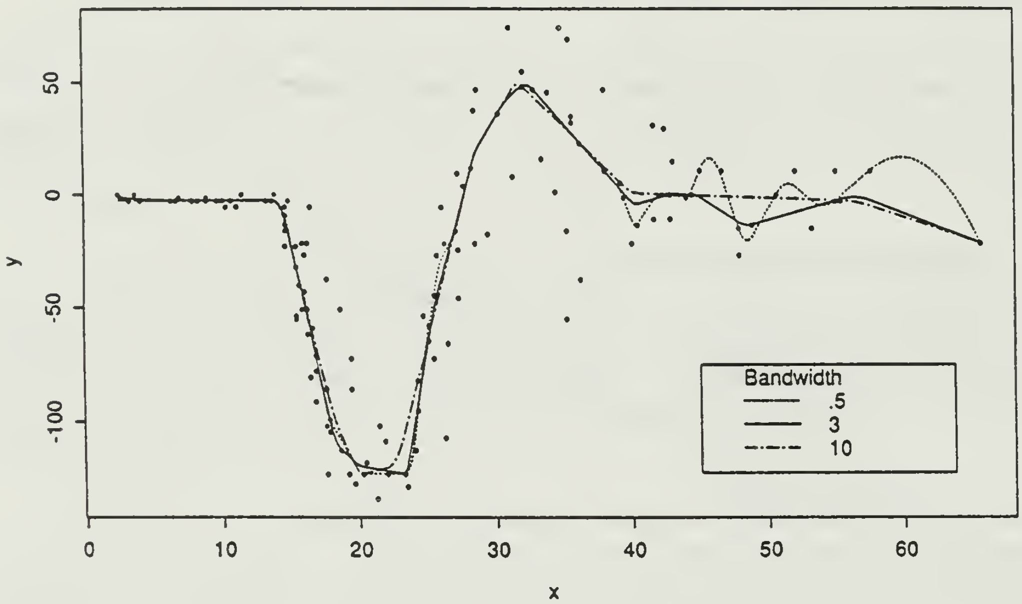

Figure 4.2: Median Smoothing Splines for Motorcycle Data

Finally, in Figure 4.3 we illustrate quantile smoothing splines for three distinct quantiles: $\alpha = .1, .5, .9$. The smoothing parameter λ is chosen to be 10 for all three estimates, although one might still argue that the resulting curves are undersmoothed. Clearly, these estimates reflect substantial heterogeneity in the conditional distribution of y over the observed range of x, thus vindicating the methods to some extent. Clearly, it is exactly this sort of nonhomogeneity that we would hope to identify with quantile methods. The computations for Figure 4.3 were done using a modified version of the regression quantile routine described in Koenker and d'Orey(1987) which implements the parametric programming aspect of the bandwidth selection. We hope to report further on this algorithm at a later date.

One feature of Figure 4.3 which is quite striking is the "near-piecewise linearity" of the estimated curves. The L_1 roughness penalty is perhaps too complaisant about the sharp elbows of $\hat{g}$, allowing $\hat{g}''$ to be large as long as this doesn't occur over long intervals. One possible response to this phenomenon is to restrict the number of knots to a more regular mesh. Another possible response is to replace the L_1 penalty with an L_∞ penalty which effectively controls the supremum of $\hat{g}''(\cdot)$. This problem too falls within the linear programming framework and we are currently exploring algorithms for its solution.

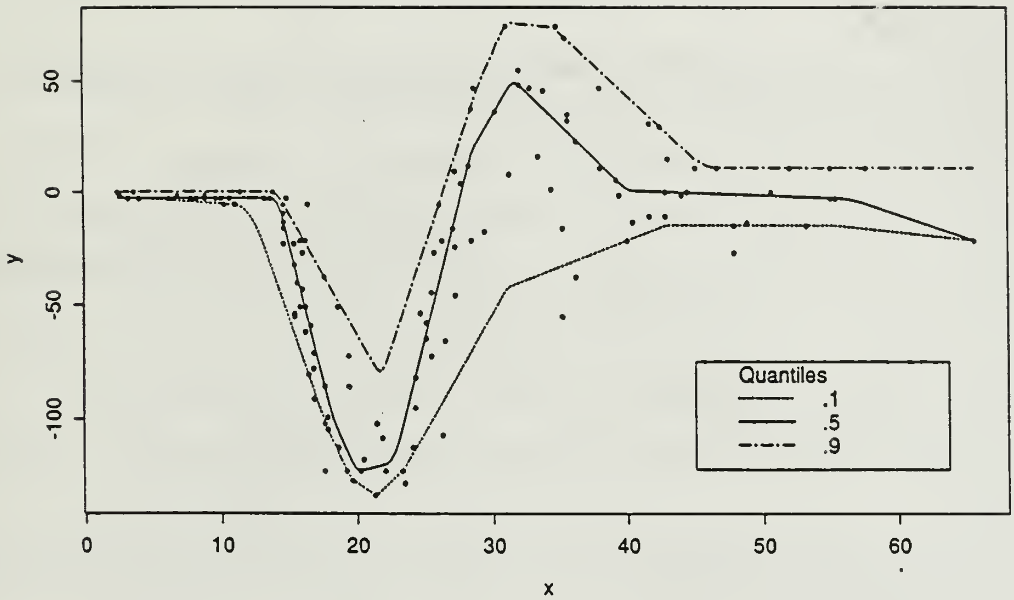

Figure 4.3: Quantile Smoothing Splines for Motorcycle Data

4. REFERENCES

1 Akaike, H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control,* 19, 716-723.

2 Antoch, J. and Janssen, P. (1989) Nonparametric regression m-quantiles, *Statistics and Probability Letters,* 8, 355-362.

3 Bassett, G. and Koenker, R. (1978) Asymptotic theory of least absolute error regression, *JASA,* 73, 618-622.

4 Becker, R.A., Chambers, J.M. and Wilks, A.R. (1988) *The New S Language,* Pacific Grove: Wadsworth.

5 Bloomfield, P. and Steiger, W.L. (1983) *Least Absolute Deviations: Theory, Applications and Algorithms,* Boston: Birkhäusen.

6 Buja, A., Hastie, T. and Tibshirani, R. (1989) Linear smoothers and additive models, with discussion, *Annals of Statistics,* 17, 453-555.

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

8 Cox, D.R. (1988) Discussion of Cole, *JRSS(A),* 151, 411.

9 DeBoor, C. (1978) *A Practical Guide to Splines,* New York: Springer-Verlag.

10 Efron, B. (1991) Regression percentiles using asymmetric squared error loss," *Statistica Sinica,* 1, 93-125.

11 Eubank, R. (1988) *Spline Smoothing and Nonparametric Regression,* New York: Dekker.

12 Härdle, W. (1990) *Applied Nonparametric Regression,* Cambridge: Cambridge University Press

13 Hendricks, W. and Koenker, R., (1990) Hierarchical spline models for conditional quantiles and the demand for electricity, *JASA,* forthcoming.

14 Jones, M.C. (1988) Discussion of Cole, *JRSS(A),* 151, 412-413.

15 Koenker, R. and Bassett G. (1978) Regression Quantiles, *Econometrica*, 46, 33-50.

16 Koenker, R. and d'Orey, V. (1987) Computing Regression Quantiles, *Applied Statistics*, 36, 383-393.

17 Portnoy, S. (1984) "Asymptotic behavior of M estimators of p regression parameters when p^2/n is large *I.* consistency," *Annals of Statistics,* 12, 1298-1309.

18 Ramsey, J.O. (1988) Monotone regression splines, with discussion, *Statistical Science,* 3, 425-461.

19 Samanta, M. (1989) Non-parametric estimation of conditional quantiles, *Statistics and Probability Letters,* 7, 407-412.

20 Schuette, D.R. (1978) A linear programming Approach to Graduation, with discussion, *Transactions of the Actuarial Society of America,* 407-445.

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

22 Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics,* 6, 461-464.

23 Siddiqui, M. (1960) Distribution of Quantiles from a Bivariate Population, *Journal of Research of the National Bureau of Standards,* 64B, 145-150.

24 Silverman, B.W. (1985) "Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion," *JRSS(B),* 47, 1-52.

25 Stone, C. (1977) Consistent Nonparametric Regression, with discussion, *Annals of Statistics,* 5, 595-645.

26 Troung, Y.K. (1989) Asymptotic properties of kernel estimators based on local medians, *Annals of Statistics,* 17, 606-617.

27 Utreras, F.I. (1981) On computing robust splines and applications, *SIAM Journal of Scientific and Statistical Computing,* 2, 153-163.

28 Wahba, G. (1990) *Spline Models for Observational Data,* CBMS-NSF Regional Conference Series in Applied Mathematics, No. 59.

29 Welsh, A.H. (1989) "On M-processes and M-estimation," *Annals of Statistics,* 17, 337-361.

30 White H. (1990) Nonparametric estimation of conditional quantiles using neural networks, unpublished.

5. APPENDIX

```
"l1ssh"<-
function(x, y, w, lambda)
{
#Compute l1 smoothing spline with penalty parameter lambda
#Solution is a quadratic spline:
#       g(x) = a_i * (x - x_i)^2 + b_i * (x - x_i) + c_i
#Inputs:
#       x--explanatory variable
#       y--explained variable
#       w--weights
#       lambda--penalty parameter
#Outputs:
#       xun--ordered vector of x with no duplicate values
#       g--smoothed value at each design point with
#          g[1] = a_1, g[i+1] = c_i
#
        ny <- len(y)
        ox <- order(x)
        xun <- unique(x[ox])
        h <- diff(xun)
        nh <- len(h)
        D <- diag(h, nrow = nh)
        D[row(D) == col(D) + 1] <- h[1:(nh - 1)]
        D[1, 1] <- 1
        B <- diag(1/h, nrow = nh)
        B[row(B) == col(B) + 1] <-  - (1/h[1:(nh - 1)] + 1/h[2:(nh)])
        B[row(B) == col(B) + 2] <- 1/h[2:(nh - 1)]
        B <- cbind(c(0, 1/h[1], rep(0, nh - 2)), B)
        B[1,  ] <- 0
        B <- cbind(0, B)
        B[1, 1] <- 1
        A <- diag(h) %*% solve(D) %*% B
        X <- matrix(0, ny, nh + 1)
        X[cbind(1:ny, category(x[ox]))] <- 1    #fidelity part of design matrix
        X <- rbind(cbind(0, diag(w[ox]) %*% X), lambda * A) #the whole X matrix
        return(x = xun, g = l1fit(X, c(w[ox] * y[ox], rep(0, nh)), int = F)$
                coef)
}


"qspline"<-
function(x, g, z)
{
#compute quadratic spline at points z given knots at x
#Inputs:
#       x--unique vector of ordered explanatory variable returned from
#          l1ssh(x,y,w,lambda)
#       g--smoothed value at each unique design point, component returned
#          from l1ssh(x,y,w,lambda)
#       z--points where smoothed values are desired
#Output:
#       return function values at points z
#
        BIG <- 1e+301
        n <- len(x)
        x <- sort(x)
        h <- c(0, diff(x))
        a <- rep(0, n + 1)
        a[2] <- g[1]
        g[1] <- g[2]
        b <- rep( - a[2] * h[2] + (g[3] - g[2])/h[2], n + 1)
        for(i in 3:(n + 1)) {
                b[i] <- (2 * (g[i] - g[i - 1]))/h[i - 1] - b[i - 1]
        }
        a[3:n] <- diff(b[3:(n + 1)])/(2 * h[3:n])
        k <- cut(z, c(x, BIG)) + 1      #obtain bin numbers for z
        k[z <= x[1]] <- 1
        x <- c(x[1], x)
        dz <- z - x[k]
        return(g[k] + b[k] * dz + a[k] * dz^2)
}
```