FACULTY WORKING PAPER 92-0170

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. 92-0170

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

October 1992

Quantile Smoothing Splines

Roger Koenker
Department of Economics

Pin Ng
Department of Economics

Quantile Smoothing Splines

Roger Koenker
Department of Economics,
University of Illinois,
Champaign, IL 61820, U.S.A.
roger@ysidro.econ.uiuc.edu

Pin Ng
Department of Economics,
University of Houston,
Houston, TX 77204-5882, U.S.A.

September, 1992

Abstract

Although nonparametric regression has traditionally focused on the estimation of conditional *mean* functions, nonparametric estimation of conditional *quantile* functions is often of substantial practical interest. We explore a class of quantile smoothing splines, which are defined as solutions to

$$\min_{g \in \mathcal{G}} \sum \rho_\tau(y_i - g(x_i)) + \lambda \left(\int_0^1 |g''(x)|^p dx \right)^{1/p}$$

with $\rho_\tau(u) = (\tau - I(u < 0))u$ and $p \geq 1$. For the particular choices $p=1$ and $p=\infty$ and $\mathcal{G} = \{ g \in C^1[0,1]: (\int_0^1 |g''(x)|^p dx)^{1/p} < \infty \}$, we show that solutions, $\hat{g}$, are parabolic splines, i.e. piecewise quadratics, on the mesh $\{x_1, \ldots, x_n\}$, and may be computed by standard l_1-type linear programming techniques. At $\lambda = 0$, $\hat{g}$ interpolates the τ^{th} quantiles at the distinct design points, and for λ sufficiently large $\hat{g}$ is the linear regression quantile fit (Koenker and Bassett(1978)) to the observations. In fact, the entire path of solutions, in the penalty parameter λ, may be efficiently computed by parametric linear programming methods. For a somewhat more general class of quantile smoothing splines we establish that $(E_X(\hat{g}(X) - g_0(X))^2)^{1/2} = O_p(n^{-2/5}(\log n)^2)$, is achievable, under mild conditions on the true conditional quantile function of Y given X. Finally we note that the approach may be easily adapted to impose monotonicity and/or convexity constraints on the fitted function.

Some Key Words: Robust Nonparametric Regression, Quantiles, Smoothing Splines, Penalized Likelihood, Parametric Linear Programming, Sieve Estimators.

1. INTRODUCTION

Several authors have recently proposed methods for nonparametric estimation of conditional quantile functions: Troung (1989) following the pioneering work of Stone (1977) on nearest neighbor methods, Chaudhuri(1991), Samanta (1989) and Antoch and Janssen (1989) using kernel methods, and White (1991) employing neural networks. Hendricks and Koenker (1992) discuss regression spline models and apply them to electricity demand data. D.R. Cox and M.C. Jones in the discussion of Cole(1988), reviving a suggestion of Bloomfield and Steiger (1983), have recently proposed estimating quantile smoothing splines which minimize

$$\sum \rho_\tau(y_i - g(x_i)) + \lambda \int (g''(x))^2 dx$$

where $\rho_\tau(u) = (\tau - I(u < 0))u$ is the Czech function of Koenker and Bassett (1978). Here the parameter $\tau \in [0, 1]$ 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 Cox(1983), Eubank (1988) and Utreras (1981) in the median $\rho_{1/2}(u) = ½ \,|u|$ 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 (1.1) retains the linear programming form of the parametric version of the quantile regression problem and yields solutions which are easy to compute.

1.1. The L_1 Roughness Penalty

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

$$R_{\tau,\lambda}[g] = \sum_{i=1}^{n} \rho_\tau(y_i - g(x_i)) + \lambda \int_0^1 |g''(x)|\, dx \tag{1.1}$$

over $g \in \mathcal{G}$, the space of continuous functions on [0, 1] with continuous first derivative and absolutely integrable second derivitive.

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

$$g(x) = \alpha_i (x - x_i)^2 + \beta_i (x - x_i) + \gamma_i \quad x_i \le x < x_{i+1} \quad i = 0, ..., n \tag{1.2}$$

Theorem 1.1. There exists a parabolic spline $\hat{g}$ which solves (1.1).

Proof. Suppose g solves (1.1). We will show that there exists a parabolic spline $\hat{g}$ such that $R[\hat{g}] = R[g]$. Suppose, provisionally, that $\mathrm{sgn}(g''(x))$ is constant on the intervals

$[x_i, x_{i+1})$, $i = 0, ..., n$, so we may write

$$\int_0^1 |g''(x)| dx = \frac{1}{2} \sum_{i=0}^{n} |g'(x_{i+1}) - g'(x_i)| \tag{1.3}$$

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

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

and

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

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 $\int |f| \geq \int f$ 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 (1.1), it is straightforward to develop an algorithm to compute $\hat{g}$. Using (1.2) the penalty becomes

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

where $h_i = x_{i+1} - x_i$, $i = 1, ..., n-1$. This enables us to express the problem (1.1) as a l_1-type linear program. A number of important features of the solution are immediately apparent from the fact that the problem is a linear program. See Koenker and Ng(1992) for algorithmic details. Since we have $(n+1)$ free parameters and $(2n - 1)$ pseudo-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 $y_i = \gamma_i$ or (ii) linearity of $\hat{g}$ in a particular subinterval of the design mesh, so $\alpha_i = 0$ for some i. Obviously, the parameter λ controls the comparative "advantage" of these two alternatives. When λ is sufficiently large *all* the α_i will be zero and the solution will correspond to the bivariate linear regression quantile fit. When λ is sufficiently small all n observations will be interpolated, and all but one of the α_i's can be non-zero.

As in any smoothing problem, choice of "bandwidth", here represented by the parameter λ, is critical. For quantile smoothing splines, this problem is ameliorated by the fact that the whole family of solutions to (1.1) for $\lambda \in [0, \infty)$ may be easily found by parametric linear programming. An important implication of this fact is that we may initially solve the simpler linear quantile regression 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. Each transition to a new solution involves a single simplex pivot of an extremely sparse constraint matrix, and hence solving (1.1) for a broad range of λ quite efficient. An interesting aspect of the way that solutions $\hat{g}(\lambda)$ depend upon the penalty parameter λ involves the number of interpolated points. In the classical 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 *e.g.* Buja, Hastie and Tibshirani

(1989) for an extensive discussion. For the quantile 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 quantile regression 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 Schwarz (1978).

1.2. The L_∞ Penalty

If we replace the L_1 roughness penalty with the L_∞ penalty, we have the objective function

$$R_{\tau,\lambda}[g] = \sum_{i=1}^{n} \rho_\tau(y_i - g(x_i)) + \lambda \sup_x |g''(x)| \tag{1.4}$$

which we would like to minimize over $g \in \mathcal{G}$. Again we can characterize the solution as a quadratic spline.

Theorem 1.2. There exists a parabolic spline which solves (1.4).

Proof. Suppose g solves (1.4). Let $\hat{g}$ be of the form (1.2) with $(\alpha_i, \beta_i, \gamma_i)$'s defined exactly as in the proof of Theorem 2.1. Clearly, $\hat{g} \in C^1$ and has the same "fidelity to the data" as g. That $\hat{g}$ isn't rougher than g in $\|\cdot\|_\infty$ follows immediately from

$$\sup_{x \in A_i} |g''(x)| h_i \geq \int_{A_i} |g''(x)| dx \geq \left| \int_{A_i} g''(x) dx \right| = \sup_{x \in A_i} |\hat{g}''(x)| h_i,$$

for $A_i = [x_i, x_{i+1})$, $i = 1, \cdots, n-1$. $\square$

This characterization of the solution allows us to rewrite the penalty as

$$\|\hat{g}''(x)\|_\infty = 2 \max_i |\alpha_i| .$$

We can now parametrize the $\hat{g}$ as in the previous case and formulate (1.4) as a linear program, but the penalty is now relegated to role of linear inequality constraints. With the L_1 penalty, as we have seen, a direct tradeoff between the number of interpolated responses and the number of linear segments between design observations existed. With the L_∞ penalty this tradeoff is altered. Active constraints now correspond not to $\alpha_i = 0$ but to $|\alpha_i| = \Delta$, the upper bound for g''; thus the tradeoff is between segments that attain the prescribed bound, and observations which are interpolated. Of course, in the limiting case $\Delta = 0$ the solution is, as with the L_1 penalty with $\lambda = \infty$, the linear τth regression quantile estimate. Clearly, the L_1 penalty favors a piecewise linear form for $\hat{g}$ with a few sharp elbows where g'' can be very large. The L_∞ penalty enforces a uniform bound on

g'' and thus straightens the elbows and introduces a modest curvature in the longer segments to compensate.

1.3. Extensions

In many practical applications there is an immediate question of extending these methods to multivariate settings. 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.

There are a number of intriguing extensions incorporating further constraints. Monotonicity and convexity of the fitted function $\hat{g}$ may be readily imposed by simply imposing further linear inequality constraints on the parameters of the problem. The final section includes examples of montonically constrained estimates. While adding such inequality constraints to the corresponding l_2 problem results in a significant increase in complexity adding linear inequality constraints to the quantile smoothing spline problems does not alter the fundamental nature of the optimization problem to be solved.

In situations in which the derivitives of g_0 need to be estimated it may be worthwhile to consider higher order derivitive penalties as is sometimes done in the classical smoothing spline literature. This might be the case in the growth curve analysis of Cole(1988) for example.

2. ASYMPTOTICS

To explore the asymptotic behavior of quantile smoothing splines we will assume that the observations on (Y_i,X_i) pairs are independent and identically distributed. We would like to estimate the τth conditional quantile function of $Y|X$ which we will denote

$$g_0(x) = F_{Y|X}^{-1}(\tau|X=x)$$

We will assume further that $g_0 \in \mathcal{G} \equiv \{g \in C^1[0,1] \,|\, \sup|g''|<\infty\}$ and that the conditional density exists and is strictly bounded away from 0 and ∞ when evaluated at $g_0(x)$, that is, there exists $m > 0$ and $M < \infty$ such that for all x in $[0,1]$ we have

$$m \le f_{Y|X}(g_0(x)) \le M.$$

The lower bound is an identifiability condition insuring the uniqueness of g_0, while the upper bound is required to exclude pathologically rapid convergence. We will consider a somewhat more general class of estimators which solve

$$\min_{g \in \mathcal{G}_n} \sum_{i=1}^{n} \rho_\tau(y_i - g(x_i))$$

where

$$\mathcal{G}_n = \{g \in \mathcal{G} \,|\, g(x) = \sum_{j=1}^{p_n} c_j B_j(x), \ \int |g''| \le \Delta_n\},$$

and the functions $\{B_j(x): i=1,...,p_n\}$ constitute a B-spline basis for the parabolic splines on the uniform mesh $\{0,0,0,h_n,2h_n, \cdots ,1,1,1\}$, where $h_n = p_n^{-1}$, see e.g. deBoor(1978).

The choice of the restricted B-spline definition of $\mathcal{G}_n$ is partly for theoretical convenience, but it also has an important practical rationale. The B-spline formulation of (1.1) is considerably easier to compute than the full smoothing spline formulation when n is very large. It represents a compromise between the classical smoothing spline and the more restrictive regression splines, see Wahba(1990). Note that there is a direct connection between the choice of Δ_n in the definition of $\mathcal{G}_n$ and the λ parameter in the previous section. Our objective will be to explore conditions on p_n, and Δ_n which insure consistency of $\hat{g}_n$ in the L_2-norm,

$$\|\hat{g}_n - g_0\|_2 \equiv (E_X(\hat{g}_n(X) - g_0(X))^2)^{1/2}.$$

Theorem 2. If $p_n = O(n^{1/5}/\log n)$ and $\Delta = O(\log n)$, then under the conditions of the preceeding paragraph,

$$\|\hat{g}_n - g_0\|_2 = O_p(n^{-2/5}(\log n)^2).$$

Proof. We will sketch the proof which relies heavily on recent work of Shen and Wong(1992) on convergence rates for sieve estimators. In Shen and Wong's notation the contribution of the ith observation to the pseudo-likelihood (negative fidelity) in our case is

$$l(g(x),y) = -\rho_\tau(y - g(x)).$$

As in Bassett and Koenker(1986), for $g(x) \ge g_0(x)$, we have

$$E_{Y|X}[\rho_\tau(Y-g) - \rho_\tau(Y-g_0)] = \int_{g_0(x)}^{g(x)} (F_{Y|X}(y) - \tau)dy$$

while for $g(x) \le g_0(x)$, the sign and limits of integration are reversed. Thus, using the density bound, there exists a $\delta>0$ such that for any y satisfying $|y-g_0(x)| \le \delta$,

$$|F_{Y|X}(y) - \tau| \ge \tfrac{1}{2}m\,|y-g_0(x)|,$$

and therefore

$$E_{Y|X}[\rho_\tau(Y-g) - \rho_\tau(Y-g_0)] \ge m\,|g - g_0|^2,$$

and

$$\inf_{\{\|g-g_0\|_2 \ge \varepsilon\}} E_{Y|X}[\rho_\tau(Y-g) - \rho_\tau(Y-g_0)] \ge m\varepsilon^2,$$

This verifies Condition C1 of Shen and Wong with $\alpha=1$. Their Condition C2 is trivially satisfied with $\beta=1$ since

$$|\rho_\tau(Y-g) - \rho_\tau(Y-g_0)| \le |g-g_0|.$$

Let $H(\varepsilon,A) = \log K(\varepsilon,A)$, denote the ε-entropy of the set A where $K(\varepsilon,A)$, denotes the number of $\|\cdot\|_\infty$ balls of radius ε required to cover A. To verify Condition C3 of Shen and Wong we must compute $H(\varepsilon,\mathcal{F}_n)$ where

$$\mathcal{F}_n = \{\rho_\tau(y-g) - \rho_\tau(y - g_0) : g \in \mathcal{G}_n\}.$$

Since the set $\mathcal{G}_n$ is isomorphic to the finite dimension parameter space Θ_n with elements $\theta = (\gamma_1, \alpha_1, \cdots, \alpha_{p_n})$ where $\gamma_1 = g(0)$ and $\alpha_i = g''(x)$ for $x \in ((i-1)h, ih]$, we have

$H(\varepsilon, \mathcal{F}_n) \le H(\varepsilon, \mathcal{G}_n) \le H(\varepsilon, \Theta_n)$, and since

$$\int_0^1 |g''(x)|\, dx = p_n^{-1} \sum_{i=1}^{p_n} |\alpha_i| \le \Delta_n$$

Theorems IX and X of Kolmogorov and Tihomirov (1959) imply,

$$H(\varepsilon, \mathcal{F}_n) \le 2p_n \log(p_n \Delta_n) \log(1/\varepsilon).$$

Thus, choosing $p_n = n^{1/5}/logn$, and $\Delta_n = \log n$, Condition C3 is satisfied with $A_3 = 2/5$, $r_0 = 1/10$ and $r = 0^+$, observing Shen and Wong's convention that $\varepsilon^{-0^+} = log\,(1/\varepsilon)$.

Theorem 1 of Shen and Wong(1992) now implies that

$$\|\hat{g}_n - g_0\|_2 = O_p(\max\{n^{-\upsilon}\,,\ \|\pi_n g_0 - g_0\|_2\,,\ K^{1/2}(\pi_n g_0, g_0)\}),$$

where $\pi_n g_0$ denotes an element of $\mathcal{G}_n$, $K(g_1, g_0) = E[\rho_\tau(Y - g_1) - \rho_\tau(Y - g_0)$ and

$$\upsilon = 2/5 - \log\log n/(2\log n).$$

Expanding $K(g_1, g_0)$ around g_0 we obtain

$$K(g_1, g_0) \le E f_{Y|X}(g_0(X))\, |g_1(X) - g_0(X)|^2 \le M \|g_1(X) - g_0(X)\|_2^2.$$

For n sufficiently large, Powell(1981, Theorem 20.3) establishes that,

$$\|\pi_n g_0 - g_0\|_\infty \le 3/2 p_n^{-2} \|g_0''\|_\infty.$$

Thus, given our choices of p_n and Δ_n, the second and third terms in the max expression are $O_p(n^{-2/5}\,(logn)^2)$. Noting that $n^{-\upsilon} = n^{-2/5}\sqrt{\log n}$ completes the proof. □

An essentially identical argument would yield the same rate of convergence for the L_∞ roughness penalty estimator.

3. PICTURES

In Figure 3.1 we illustrate three different quantile smoothing splines estimated using the L_1 roughness penalty. The data is the well known motorcycle data described for example in Härdle(1990). In Figure 3.2 we illustrate three comparable estimates using the L_∞ penalty for the same problem. The piecewise linearity of the L_1 estimates is already apparent in these figures. Since there are knots at each design point, apparent kinks, "elbows", in the fitted curve correspond to large values of the (piecewise constant) second derivitive on short intervals between adjacent design points. In Koenker, Ng, and Portnoy(1992) we contrast these estimates with those from a kernel estimate of the conditional quantile functions and conclude that the splines perform considerably better. Note that in the L_∞ picture the three estimated quantile functions cross at the penultimate point; this is apparently due to the wide separation of the last design point from the others, a fact that has prompted other investigators to omit it from their plots.

In Figures 3.3 and 3.4 we illustrate two quantile smoothing spline estimates for the data appearing in Scheutte(1978) which is a typical example of actuarial "graduation" of mortality tables. In this example it may be reasonable to impose monotonicity on the fitted curves, so we contrast estimates based on both L_1 and L_∞ penalties with their monotonically constrained counterparts. Note that since the first derivitive of $\hat{g}$ is piecewise linear, it suffices to constrain the first derivitive at each of the knots to be positive, which requires n additional linear inequality constraints. Here the estimates are

Figure 3.1 Quantile Smoothing Splines for Motorcycle Example: L_1 penalty $\lambda = 3$, and $\tau \in \{.1,.5,.9\}$

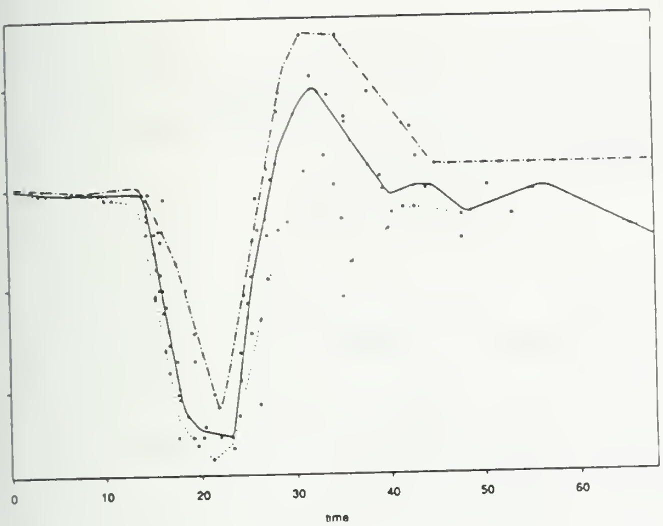


Figure 3.2 Quantile Smoothing Splines for Motorcycle Example: L_∞ penalty $\lambda = 50$, and $\tau \in \{.1,.5,.9\}$

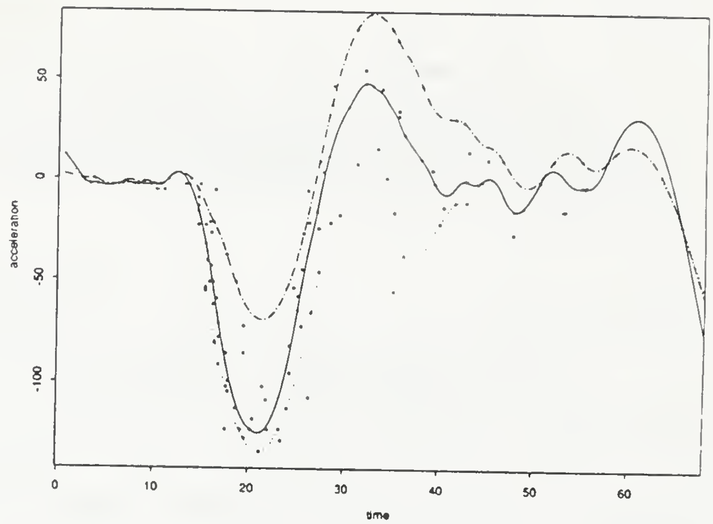


Figure 3.3 Median Smoothing Splines for Schuette Example: L_1 penalty Monotone vs. Unconstrained Fit $\lambda = .25$

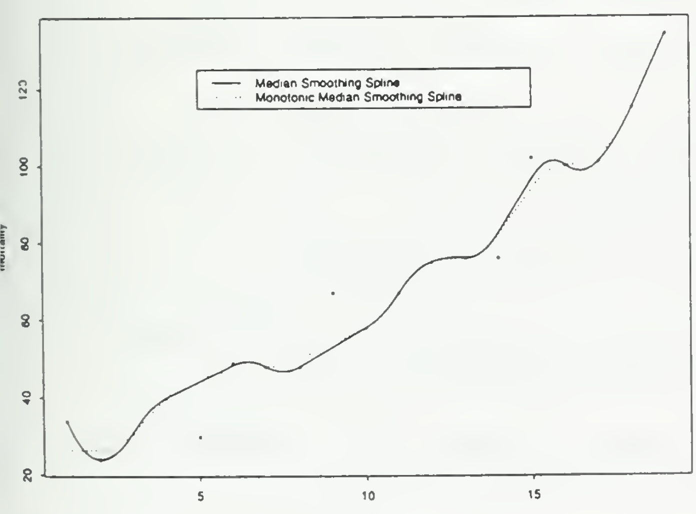
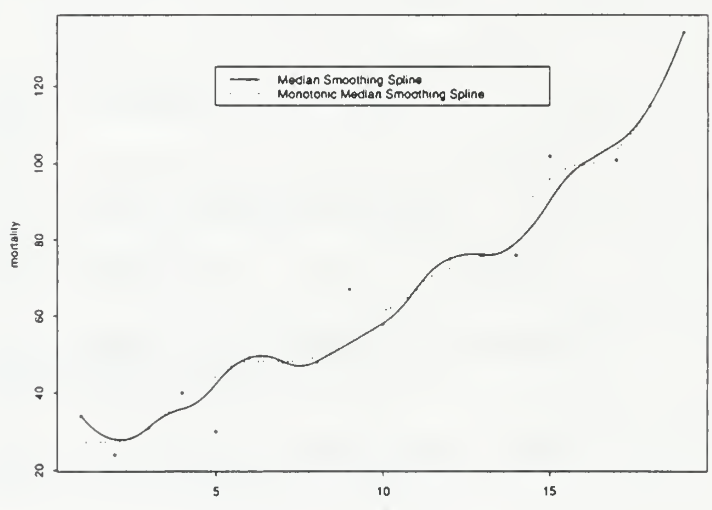


Figure 3.4 Median Smoothing Splines for Schuette Example: L_∞ penalty Monotone vs. Unconstrained Fit $\lambda = 5$



probably somewhat "undersmoothed" and the piecewise linearity of the L_1 penalty estimate is somewhat less apparent.

All of the computing was carried in the S language of Becker, Chambers, and Wilks(1988). The underlying algorithms are based on modifications of Bartels and Conn(1980) algorithm for constrained l_1 regression.

4. Acknowledgement

The authors wish to thank Cornelius Gutenbrunner, Steve Portnoy, and Mike Osborne for stimulating discussions, Wing Wong for sending us his Technical Report on Sieves, and Esther Portnoy for pointing out the important Schuette reference. Research support from NSF Grant SES 91-21776 is also gratefully acknowledged.

REFERENCES

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

Bassett, G.W. and Koenker, R.W. (1986). Strong consistency of regression quantiles and related empirical processes, *Econometric Theory,* 2, 191-201.

Bartels, R. and Conn, A. (1980) Linearly constrained discrete l_1 problems, *Transactions of the ACM on Mathematical Software,* 6, 594-608.

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

Buja, A., Hastie, T. and Tibshirani, R. (1989) Linear Smoothers and Additive Models, *Annals of Statistics,* 17, 453-555.

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

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

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

Cole, T.J. (1988) Fitting smoothing centile curves to reference data, with discussion, *Journal of Royal Statistical Society, Ser. A, 151, 385-418.*

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

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

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

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.

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

Koenker, R. and Ng, P.T. (1992) Computing quantile smoothing splines, *Proceedings of the 1992 Interface Conference.*

Koenker, R., Ng, P.T., and Portnoy, S.L.(1992) , Nonparametric estimation of conditional quantile functions *Proceedings of the 2nd Neuchatel Conference on Statistical Data Analysis based on the l_1 Norm and related Methods*, New York: North-Holland.

Kolmogorov, A.N. and Tihomirov, V.M. (1959) ε-capacity of sets in function spaces, *AMS Translations 2*, 17, 277-364.

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

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

Schuette, D. R. (1978). A linear programming approach to graduation, *Trans. Soc. of Actuaries, 30*, 407-445.

Shen, X. and Wong, W.H. (1992). Convergence Rates for Sieve Estimators, Technical Report 293, Department of Statistics, University of Chicago.

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

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

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

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

White H. (1991) Nonparametric estimation of conditional quantiles using neural networks, *Proceedings of the 1991 Interface Conference.*