FACULTY WORKING PAPER 93-0130

Quantile Smoothing Splines

Roger Koenker
University of Illinois

Pin Ng
University of Houston

Stephen Portnoy
University of Illinois

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

BEBR

FACULTY WORKING PAPER NO. 93-0130

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

April 1993

Quantile Smoothing Splines

Roger Koenker
Pin Ng
Stephen Portnoy

Quantile Smoothing Splines

Roger Koenker
University of Illinois

Pin Ng
University of Houston

Stephen Portnoy
University of Illinois

April, 1993

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 a penalized quantile regression problem. We characterize solutions, as splines, i.e. piecewise polynomials. and discuss computation by standard linear programming techniques. For sufficiently small values of the bandwidth parameter the solutions interpolate the specified quantiles of the response variable at the distinct design points, while for sufficiently large bandwidths solutions specialize to the linear regression quantile fit (Koenker and Bassett(1978)) to the observations. Because the methods estimate conditional quantile functions they possess an inherent robustness to extreme observations in the response variable. Remarkably, the entire path of solutions, in the quantile parameter or the bandwidth parameter, may be computed efficiently by parametric linear programming methods. Finally we note that the approach may be easily adapted to impose monotonicity, convexity, or other constraints on the fitted function. Two examples are provided to illustrate the use of the proposed methods.

KEYWORDS: Nonparametric regression, quantiles, splines, change-points, smoothing, bandwidth selection.

1. INTRODUCTION

Several authors have recently proposed methods for nonparametric estimation of conditional quantile functions: Truong (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. Cox and Jones in the discussion of Cole(1988), reviving a suggestion of Bloomfield and Steiger (1983), have recently proposed estimating quantile smoothing splines which minimize

$$\sum_{i=1}^{n} \rho_\tau(y_i - g(x_i)) + \lambda \int (g''(x))^2 dx$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check 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 cubic spline, thus generalizing the extensive literature on classical least-squares 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 $\rho_{1/2}(u) = \frac{1}{2}\, |u|$ case. However, the resulting quadratic program poses some serious computational obstacles. A recent paper by Bosch, Ye and Woodworth(1993) discusses an interior point algorithm for this problem. 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. Solutions with this L_1 form of the roughness penalty are linear splines and therefore provide a natural, automatic approach to estimating certain piece-wise linear change-point models. An application of this sort to the relationship between maximal running speed and body mass of terrestrial mammals is provided in Section 3.

2. QUANTILE SMOOTHING SPLINES

2.1. The L_1 Roughness Penalty

In prior work (Koenker, Ng, and Portnoy(1992)), we have considered the problem of minimizing

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

with $0=x_0<x_1< \cdots <x_n<x_{n+1}=1$, over the (Sobolev) space W_1^2 of continuous functions on [0, 1]

with absolutely continuous first derivative and absolutely integrable second derivative. However, the argument given there that the solution to (1.1) is a parabolic spline, i.e. piecewise quadratic, is incorrect. Indeed the problem expressed in (1.1) is ill-posed since the infimum of $R_{\tau,\lambda}$ is not attained by an element of W_1^2. The situation can be rectified by reformulating the problem somewhat adopting the approach of Fisher and Jerome (1975) and Pinkus (1988) who consider closely related problems of optimal interpolation.

We should begin by briefly reviewing some results on optimal interpolation. For an integer $k \geq 2$, and $p \in [1, \infty)$, let

$$\|f\|_p = (\int_0^1 |f(x)|^p)^{1/p}$$

and W_p^k denote the Sobolev space of real functions on [0, 1] with $k-1$ absolutely continuous derivatives and k^{th} derivative existing almost everywhere as a function in $L_p[0, 1]$. We wish to consider the problem of finding the smoothest interpolant of the points $\{(x_i, y_i), i = 1, \cdots, n\}$ in the sense of solving

$$\inf\{\|g^{(k)}\|_p : g \in W_p^k,\ g(x_i) = y_i,\ i=1, \cdots, n\} \tag{1.2}$$

The case $p=2$ is best known, yielding splines of degree $2k-1$ with knots at the points $\{x_i, i = 1, \cdots, n\}$. We are primarily interested in the case of $p = 1$ and $p = \infty$, which have been treated by deBoor (1976), Fisher and Jerome (1975) and Pinkus (1988).

For $p = 1$, apparently Fisher and Jerome (1975) were the first to observe that (1.2) has no solution for $g \in W_1^k$. They showed that if W_1^k is expanded to include functions whose k^{th} derivatives are measures, the expanded problem does have a solution s, as a spline of degree $k-1$, that the total variation of its $(k-1)^{\text{th}}$ derivative, $V(s^{(k-1)})$, coincides with the extremal value of (1.2), and that the measure $s^{(k)}$ is concentrated on n, or fewer, points. Pinkus (1988) has refined this characterization somewhat and has provided considerable further generalization.

To bridge the gap between the smoothing problem posed in (1.1) and the optimal interpolation problem (1.2), we may simply observe that any solution, $\hat{g}$, to the former must also solve the latter in the sense that $\hat{g}$ must interpolate itself at the observed $\{x_i\}$ and therefore must minimize the roughness penalty, subject to a given fidelity constraint. Thus to determine *the form* of the solution to the smoothing problem it suffices to consider the interpolation problem.

It remains to consider the question of knot selection for the $p = 1$ case. Pinkus (1988), under somewhat restrictive conditions on the y_i's, notes that for $k = 2$, the case of primary interest here, the knots of the optimal spline coincide with the observed x_i. That this is true for any configuration of y_i's can be argued as follows. Let f be any interpolator of the points $\{(x_i, y_i): i = 1, \cdots, n\}$ with an absolutely continuous first derivative. Recall, e.g. Natanson (1974, p. 259), that the total variation of an absolutely continuous function is the integral of the absolute value of its derivative, thus we may write,

$$V(f') = \int |f''(x)|dx$$

Now, by the mean value theorem, let $u_i \in (x_i, x_{i+1})$ be such that $f'(u_i) = (y_{i+1} - y_i)/(x_{i+1} - x_i)$ for $i = 1, \cdots, n-1$. Then,

$$V(f') \geq \sum_{i=1}^{n-1} |\int_{u_i}^{u_i+1} f''(x)dx| \geq \sum_{i=1}^{n-1} |f'(u_{i+1}) - f'(u_i)| = V(\hat{f}')$$

where $\hat{f}$ is the piecewise linear interpolator with knots at the x_i. Finally, note for any continuous piecewise linear g there exists a sequence of functions $\{g_n\}$ with absolutely continuous first derivative such that

$$\lim_{n \to \infty} V(g_n') = V(g')$$

and thus by the foregoing argument $\hat{f}$ minimizes $V(g)$ for all such g.

Thus, following Pinkus (1988) if we expand our original space slightly from the Sobolev space

$$W_1^2 = \{g: g(x) = a_0 + a_1 x + \int_0^1 (x-y)_+ h(y)dy, \ \ h \in L_1, \ a_i \in \mathbf{R}, \ i = 0,1\}$$

to

$$U^2 = \{g: g(x) = a_0 + a_1 x + \int_0^1 (x-y)_+ \, d\mu(y), \ \ V(\mu) < \infty, \ \ a_i \in \mathbf{R}, \ i = 0,1\}$$

and replace the L_1 penalty on g'' with a total variation of penalty on g' we obtain,

Theorem 1. The function $g \in U^2$ minimizing $\sum \rho_\tau(y_i - g(x_i)) + \lambda V(g')$ is a linear spline with knots at the points x_i, $i = 1, \cdots, n$.

Having established the form of the solution, it is straightforward to develop an algorithm to compute $\hat{g}$. We may write

$$\hat{g}(x) = \alpha_i + \beta_i(x - x_i) \quad x \in [x_i, x_{i+1}) \ \ i = 0, \cdots, n$$

and by the continuity of $\hat{g}$

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

Thus writing $h_i = x_{i+1} - x_i$, we have $\beta_i = (\alpha_{i+1} - \alpha_i)/h_i$. The penalty may thus be expressed as

$$V(\hat{g}') = \sum_{i=1}^{n-1} |\beta_{i+1} - \beta_i| = \sum_{i=1}^{n-1} |(\alpha_{i+2} - \alpha_{i+1})/h_{i+1} - (\alpha_{i+1} - \alpha_i)/h_i|$$

Thus parameterizing $\hat{g}$ by the n-vector $\alpha = (\hat{g}(x_i))$, we may write the original problem as a linear program,

$$\min_{\alpha \in \mathbf{R}^n} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha_i) + \lambda \sum_{j=1}^{n-1} |d_j'\alpha|$$

where $d_j' = (0, \cdots, 0, h_j^{-1}, -(h_{j+1}^{-1} + h_j^{-1}), h_{j+1}^{-1}, 0, \cdots, 0)$, $j = 1, , \cdots, n-1$. In the important special (median) case of $\tau = 1/2$, we can view this as simple data augmentation and therefore as an ordinary least absolute deviation (LAD) regression problem. For general τ, further modifications described in detail in Koenker and Ng (1993) and implemented for "S" (Becker, Chambers and Wilks (1988)) are conceptually straightforward. Since we have n free parameters, α, and $2n - 1$ pseudo-observations, solutions must have n zero pseudo residuals by complementary slackness. And in our case these zeros correspond to either (i) exact interpolation of observations, so $\hat{\alpha}_i = y_i$, or (ii) linearity of $\hat{g}$ at an internal knot, i.e., $\beta_{i+1} = \beta_i$ for some index i. Obviously, the parameter λ controls comparative "advantage" of these two alternative means of reducing the objective function. When λ is sufficiently large *all* the $\hat{\beta}_i$ will be equal and the solution will be the bivariate linear quantile regression fit as in Koenker and Bassett (1978). When λ is sufficiently small, all n observations will be interpolated when the design points are unique, otherwise the τth quantiles at each distinct design point are interpolated.

2.2. Bandwidth Choice

As in any smoothing problem, choice of "bandwidth", here represented by the parameter λ, is critical. For quantile smoothing splines, the problem of computing a family of solutions for various λ is greatly eased by the fact that the problem is a *parametric* linear program in the parameter λ. An important implication of this fact is that we may initially solve the much smaller 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 a potentially rather large problem. It might be noted that n would need to be quite large by the usual standards of applied statistics in order that the resulting problem would actually loom large by the standards of contemporary linear programming.

Each transition to a new solution of the parametric linear program in λ involves a single simplex pivot of an extremely sparse tableau, and hence solving for a broad range of λ is quite efficient. The situation is quite analogous to the problem of solving for the entire family of quantile regression solutions in the parameter τ, described originally in Bassett and Koenker (1982), and described in greater detail in Koenker and d'Orey(1987).

An interesting, and important aspect of the way that solutions depend upon the penalty parameter λ involves the number of interpolated points. In the classical 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 are usually based on the trace of various quasi-projection matrices in the least-squares theory. See, for example, Hastie and Tibshirani (1990) for a cogent 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 may be used in conjunction with the function $R[\hat{g}]$ itself to implement data-driven bandwidth choice. The criterion

$$SIC(p_\lambda) = log\,(n^{-1} \sum_{i=1}^{n} \rho_\tau(y_i - \hat{g}(x_i)) + \tfrac{1}{2} n^{-1} p_\lambda \log n,$$

which may be interpreted as the Schwarz(1978) criterion for quantile smoothing spline problem seems to perform well in some limited applications. Machado(1993) considers similar criteria for parametric quantile regression and more general M-estimators of regression.

2.3. The L_∞ Penalty

Replacing the L_1 roughness penalty with the L_∞ penalty, we have

$$R_{\tau,\lambda}[g] = \sum \rho_\tau(y_i - g(x_i) + \lambda \sup_x |g''(x)|.$$

Again we may focus on the corresponding interpolation problem which has an extensive literature. Favard (1940) was apparently the first to show that the problem of minimizing $||g^{(k)}||_\infty$ over

$$\{g : g^{(k)} < \infty,\ g(x_i) = g_0(x_i) \quad i = 1, \cdots, n\}$$

for fixed function g_0, had a solution which was a polynomial spline of degree k, with k^{th} derivative zero outside $[x_1, x_n]$ and less than $n - k$ knots all simple inside (x_1, x_n). Later Karlin (1975), deBoor (1976) and Fisher and Jerome (1975) clarified the critical role of perfect splines, i.e. splines of degree k, with $|g^{(k)}|$ constant as solutions to this "optimal interpolation" problem. As the discussion in Powell (1981, Chapters 23-24, see especially Figure 23.1) makes evident the perfect spline solutions may not be terribly appealing. They are obviously required to be "uniformly rough", an undesirable feature unless the target function has this property. Furthermore, the knot locations of the optimal perfect spline depend upon the configuration of interpolated observations, a fact which greatly complicates their computation. Nevertheless we believe it is still interesting and worthwhile to consider the L_∞ penalty and to this end we propose minimizing over quadratic splines with knots at the observed x_i. Now

$$\hat{g}(x) = \alpha_i + \beta_i(x - x_i) + \gamma_i(x - x_i)^2 \quad x \in [x_i, x_{i+1}) \quad i = 0, \cdots, n$$

This formulation allows us to rewrite the penalty as

$$||g''||_\infty = 2 \max_i |\gamma_i|.$$

Again we can formulate the problem as a linear program, but now the penalty appears in the role of linear inequality constraints. As above let $h_i = x_{i+1} - x_i$, $i = 1, \cdots, n-1$. From the continuity constraints of the quadratic spline we have

$$\gamma_i h_i^2 + \beta_i h_i + \alpha_i = \alpha_{i+1}$$

$$2\gamma_i h_i + \beta_i = \beta_{i+1}$$

for $i = 1, ..., n-1$, with $\beta_0 = \beta_1$ and $\alpha_0 = \alpha_1$. We require that the quadratic spline be linear in the exterior intervals $[0, x_1)$ and $(x_n, 1]$. Obviously neither the roughness penalty nor the fidelity contribution are affected by this requirement, which gives us $\gamma_0 = \gamma_n = 0$. Eliminating the β's yields

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

so we have $3(n+1)$ parameters. and $2(n+1)$ linear constraints. Writing $\theta = (\gamma_1, \alpha_1, \cdots, \alpha_n)$ as the $(n+1)$ vector of free parameters we have

$$\gamma = K^{-1} B \theta = Q\theta$$

where $\gamma = (\gamma_1, \cdots, \gamma_{n-1})'$, K is a $(n-1) \times (n-1)$ banded matrix with

$$k_{1,1} = 1, \; k_{j,j} = h_j, \; k_{j,j-1} = h_{j-1} \qquad j = 2, \cdots, n-1$$

and B is a banded $(n-1) \times (n+1)$ matrix with

$$b_{1,1} = 1, \; b_{j,j} = h_{j-1}^{-1}, \; b_{j,j+1} = -(h_{j-1}^{-1} + h_j^{-1})$$

$$b_{n-1,n+1} = h_{n-1}^{-1} \qquad j = 2, \cdots, n-1.$$

Now introducing the parameter σ to represent the bound on g'', we may write the problem as

$$\min_{(\theta,\sigma)} \{ \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\theta) + \lambda\sigma, \quad Q\theta \in [-\sigma, \sigma]^{n-1} \}.$$

where $X = [0 \vdots I_n]$. Thus we are once again faced with a linear program, a modified version of the Bartels and Conn (1980) algorithm for linear, inequality-constrained LAD problems has been implemented for *S*.

The L_∞ roughness penalty may be viewed as uniform prior on g''. Each λ implies a corresponding upper bound on the magnitude $\|g''\|_\infty$. This is quite different than the L_1 case. Active constraints now correspond not to consecutive segments having the same slope as in the L_1 case, but to segments on which $|\gamma_i| = \sigma$, i.e. where $\hat{g}''$ hits the permissible upper bound. Of course in the limiting case as $\lambda \to \infty$ so $\sigma \to 0$, the solution is, as with the L_1 penalty, the linear τ^{th} regression quantile estimate. In contrast to the piecewise linear form for $\hat{g}$ with a few "elbows" where g' jumps which characterizes the solution for the L_1 penalty, the L_∞ penalty enforces a uniform bound on g'' and this straightens the elbows and introduces modest curvature over longer segments to compensate. The L_∞ solution seems visually much smoother than the piecewise linear L_1 solution. Which seems preferable is obviously application dependent. The comments on bandwidth selection above apply also to the L_∞ case.

2.4. Extensions

Clearly there is considerable scope for other forms of the roughness penalty. We have focused on the L_1 and L_∞ penalties on g'', but other L_p norms are possible as are other differential operators. Indeed it might be interesting to explore the estimation of conditional mean models using the L_1 and L_∞ penalties, if efficient algorithms for the resulting quadratic programming problems could be developed. The simple forms considered above have the virtue that their linear programming formulation makes efficient computation immediately practical.

In many practical applications there will often be the question of extending these methods to multivariate settings. The additive spline models of Hastie and Tibshirani (1990) and others naturally suggest themselves. Some preliminary plots for bivariate x look quite promising. 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. 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.

Rates of convergence for these splines undoubtedly parallel classical results for least-squares smoothing splines. For the case of regression splines (i.e., for unconstrained p-dimensional classes of B-splines, with $p = p_n \to \infty$), He and Shi (1992) show that the L_2 convergence rate is $n^{-2/5}$. We hope to report further on the asymptotic behavior of quantile smoothing splines in future work.

3. SOME EXAMPLES AND ILLUSTRATIONS

Our first example, based on Chappell (1989), explores the relationship between maximal running speed and body mass of terrestrial mammals. The data, collected and described in detail by Garland (1983) is plotted in Figure 3.1; 107 species are represented. Two groups are identified for special treatment by Chappell: "hoppers" which, like the kangaroo, ambulate by hopping and are labeled by the plotting character h in the figure, and "specialized", labeled s, which like the hippopotamus, the porcupine, and man "were judged unsuitable for the inclusion in analyses on account of lifestyles in which speed does not figure as an important factor." For reference we have included Chappell's quadratic, and single-changepoint, log linear models. Both are estimated by least-squares, both omit the s observations and fit an additive shift effect for the "hoppers".

In Figure 3.2 we illustrate two cubic smoothing splines estimated by minimizing the penalized least squares criterion

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

The solid line is the fit when the entire sample is included, the dotted line excludes the special animals labeled s. In both cases λ is chosen by generalized cross-validation as described, for example, in Craven and Wahba (1979). One can immediately see the lack of robustness of the least squares splines to the slower special animals.

Next we fit the entire family of median smoothing splines using the L_1 penalty. There are 182 distinct curves corresponding to λ's ranging from 0 to ∞. In Figure 3.3 we plot 3 of these curves for $\lambda = \{1.01, 12.23, 41.16\}$. The dimension of the fitted functions represented by the number of interpolated points is given in the legend. Like the least squares spline in the previous figure, these estimates are based on all the observations. However, unlike the least squares splines which estimate the conditional mean function, these median splines have an inherent robustness to outliers in the vertical direction. As in parametric quantile regression, points may be moved up or down in the plot without effecting the fitted function so long as they do not cross it. The property follows immediately from the fact that the subgradient of the objective function depends upon the y_i only through the signs of the residuals not their magnitude. See Koenker and Bassett (1978, Thm 3.5).

Using the SIC criterion to choose λ selects the solid line with a single break. This fit is remarkably similar to Chappell's preferred single changepoint model, particularly in view of the fact that we have done none of the preliminary data editing which seems essential to the success of the least squares based methods. The simple piecewise linear form of the L_1 splines make them a natural technique for estimating linear changepoint models.

In Figure 3.4 we illustrate several distinct quantile smoothing splines for the same data, again based on the L_1 penalty. Here the upper quantiles are of particular interest since they represent the envelope of biological feasibility. In this figure we have again chosen λ for the median and 90^{th} percentile by the SIC criterion; however, SIC produces a rather rough fit with $p(\hat{\lambda}) = 8$ for the 25^{th} percentile and $p(\hat{\lambda}) = 7$ for the 75^{th}. So we have selected somewhat larger λ's for these curves to achieve a more consistent degree of smoothness. Even so, the viewer will note that the 75^{th} and 90^{th} percentile curves cross in Figure 3.4 indicating, perhaps, that the 75^{th} may still be somewhat oversmoothed, or simply that there is not enough data to distinguish these two quantiles for the larger animals.

In Figure 3.5 we illustrate the same four quantiles this time employing the L_∞ penalty. Now the solutions are quadratic splines and visually smoother than the linear splines produced by the L_1 penalty. Nevertheless the qualitative features of the estimated curves are quite similar. Again, SIC was quite successful except at the .25 quantile. In this case we again increased λ to achieve a more consistent degree of smoothness.

Our second example is the well-known motorcycle data which has been widely analysed in the nonparametric regression literature. The data, which appear in Härdle (1990), are accelerometer readings taken through time from an experiment on the efficacy of motorcycle crash helmets. The x-coordinate is the time in milliseconds after a simulated impact, and the response variable y is the acceleration (in g) of the head of the test dummy. In Figure 3.6 we illustrate the classical smoothing spline estimate, with GCV λ, as well as the L_1 median smoothing spline based on SIC selected λ. An interesting feature of the piecewise linear L_1 estimate is that unlike other estimates, it does not suggest that the dummy anticipates the crash, accelerating its head prior to the initial deceleration.

In Figure 3.7 and 3.8 we illustrate several quantiles for the motorcycle data using both the L_1 and L_∞ penalties. We might observe that the piecewise linear L_1 estimate are very similar to the L_1-penalty estimates using quadratic splines which have appeared in our previous work. For reasonable λ the quadratic splines were also essentially piecewise linear with brief quadratic segments to connect the linear stretches. In Figure 3.7, 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 delete it from their plots.

4. ACKNOWLEDGEMENT

The authors wish to thank Cornelius Gutenbrunner, and Mike Osborne for stimulating discussions, Esther Portnoy for pointing out the important Schuette reference, Rick Chappell for providing the mammal data, and Alan Pinkus for helpful email guidance on the optimal interpolation literature. Research support from NSF Grant SES 91-21776 is also gratefully acknowledged.

REFERENCES

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

Bassett, G.W. and Koenker, R.W. (1982) An empirical quantile function for the iid linear model, *J. Am. Stat. Assoc,* 77, 407-415.

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, *Trans. ACM Math. Software,* 6, 594-608.

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

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

Bosch, R. Ye, Y. and Woodworth, G. (1993) A convergent algorithm for quantile regression with smoothing splines, University of Iowa, Department of Statistics, Technical Report.

Chappell, R. (1989) Fitting bent lines to data, with applications to allometry. *J. Th. Biology*, 138, 235-256.

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

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.

Craven,P. and Wahba, G. (1979) Smoothing noisy data with spline functions *Numerische Mathematik*, 31, 377-403

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

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

Favard J. (1940). Sur l'interpolation, *J. Math. Pures Appl.* 19, 281-306.

Fisher S.D. and Jerome, J.W. (1985). Spline Solutions to L_1 Extremal Problems in One and Several Variables, *J. Approx. Theory, 13, 73-83.*

Garland, T. (1983) The relation between maximal running speed and body mass in terrestrial mammals, *J. Zool. Lond.* 199, 157-170.

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

Hastie, T.J. and Tibshirani, R.J. (1990) *Generalized Additive Models*, London: Chapman-Hall.

He, X. and Shi, P. (1992). Convergence rate of B-spline estimators of nonparametric conditional quantile functions, National University of Singapore, Technical report.

Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity, *J. of the Am. Statis. Assoc*, 87, 58-68.

Karlin, S. (1975) Interpolation properties of generalized perfect splines and the solution of certain extremal problems. *Trans. of the Am. Math. Soc.* 206, 25-66.

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

Koenker, R. and Ng, P.T. (1993). Computing quantile smoothing splines, *Proceedings of the 1992 Interface Conference.*

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

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.

Machado, J.A.F.(1993). Robust Model Selection and M-Estimation. *Econometric Theory,* forthcoming.

Natanson (1974). *Theory of Functions of a Real Variable,* New York: Ungar.

Pinkus A. (1988). On smoothest interpolants, *SIAM J. Math. Anal.* 19, 1431-1441.

Powell M.J.D. (1981). *Approximation Theory and Methods,* Cambridge: Cambridge U. Press.

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.

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

Truong, 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 J. of Sci. and Stat. Comp,* 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.*

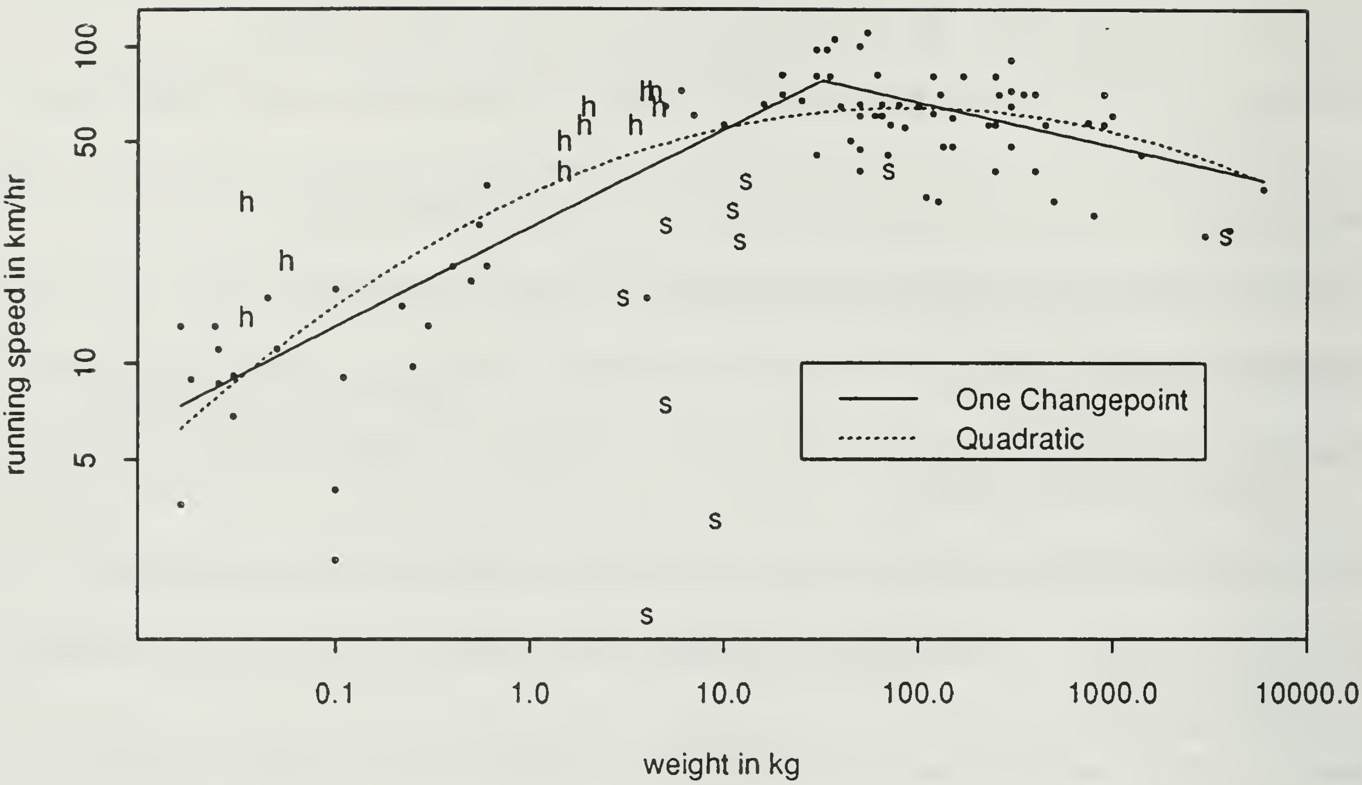

Fig 3.1: Body Mass vs. Maximal Recorded Running Speed of 107 Terrestrial Mammals with two estimated models from Chappell(1989)

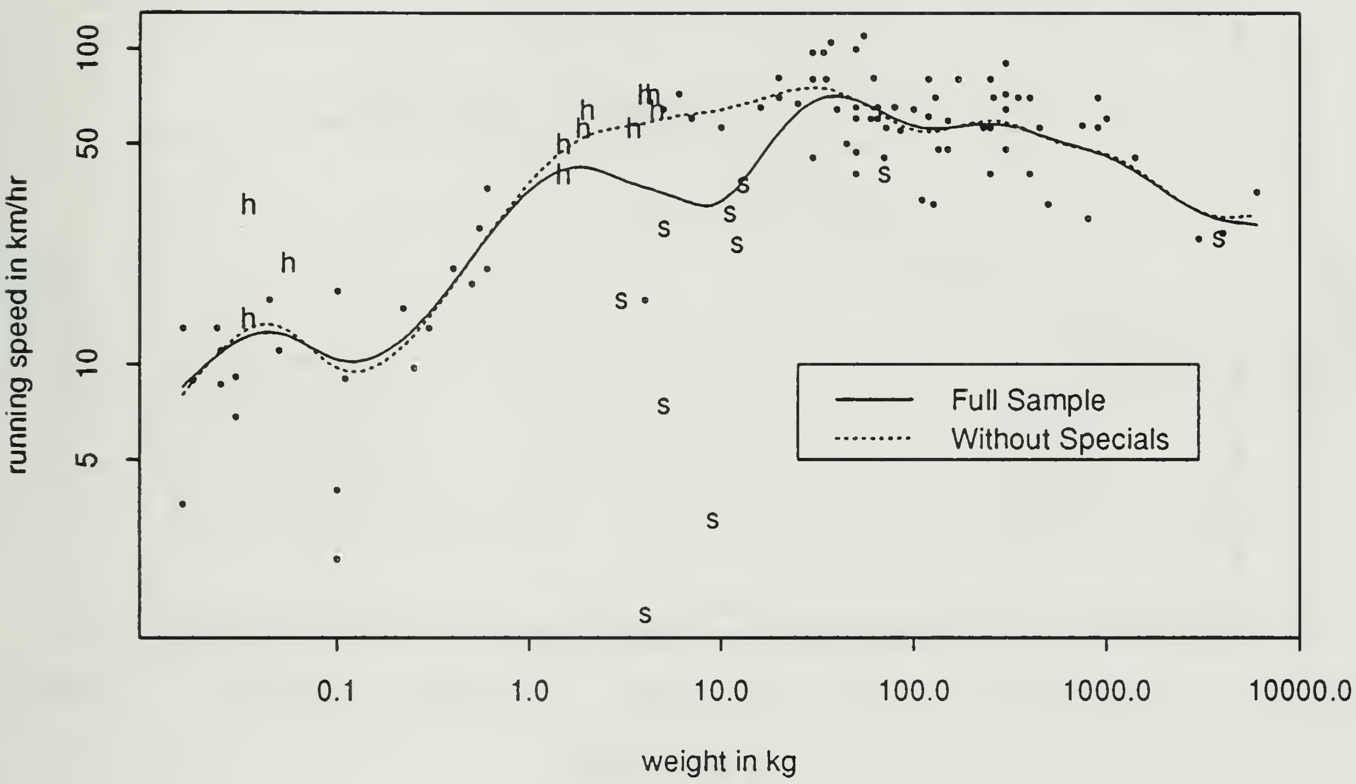

Fig 3.2: Mammal Data and two Least-Squares (Cubic) Smoothing Splines

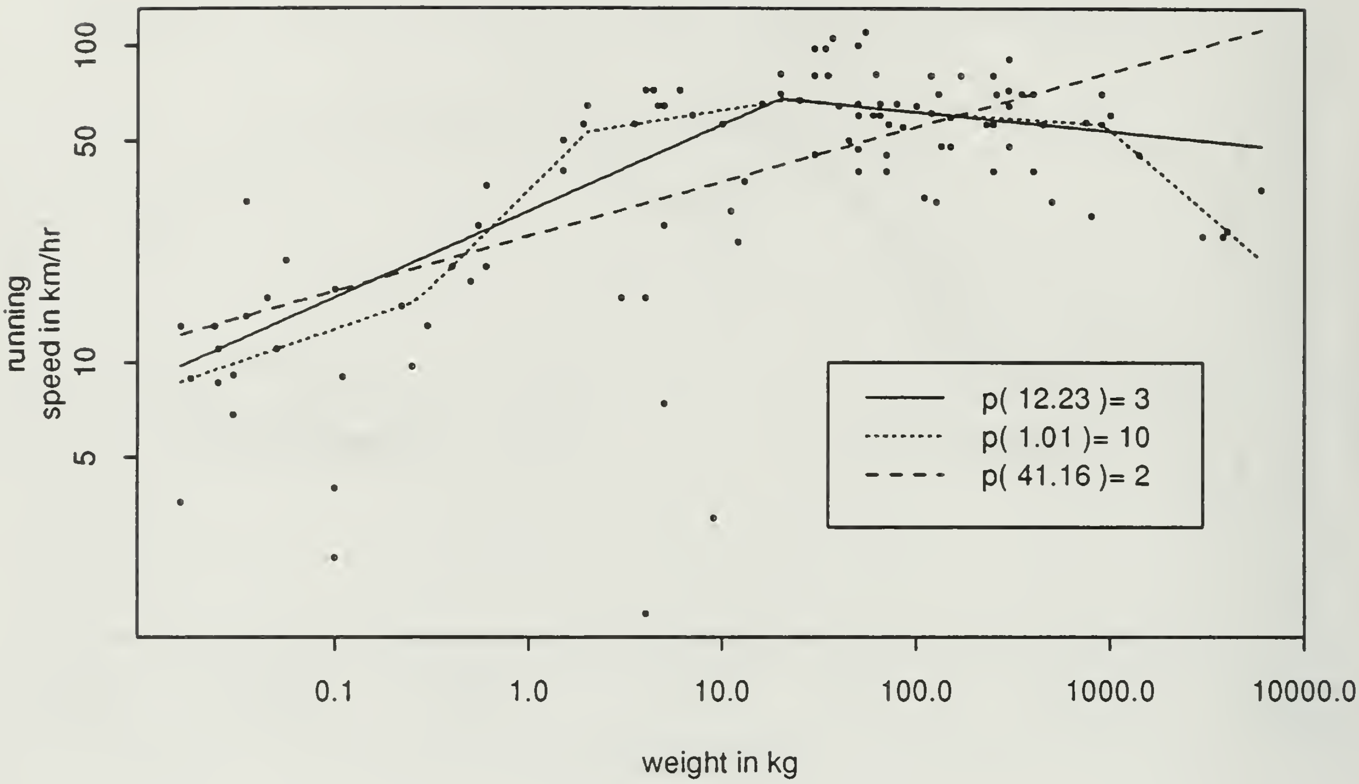

Fig 3.3: Mammal Data and three Median L_1 Smoothing Splines: Effective dimension of the spline is $p(\lambda)$ indicated in the legend

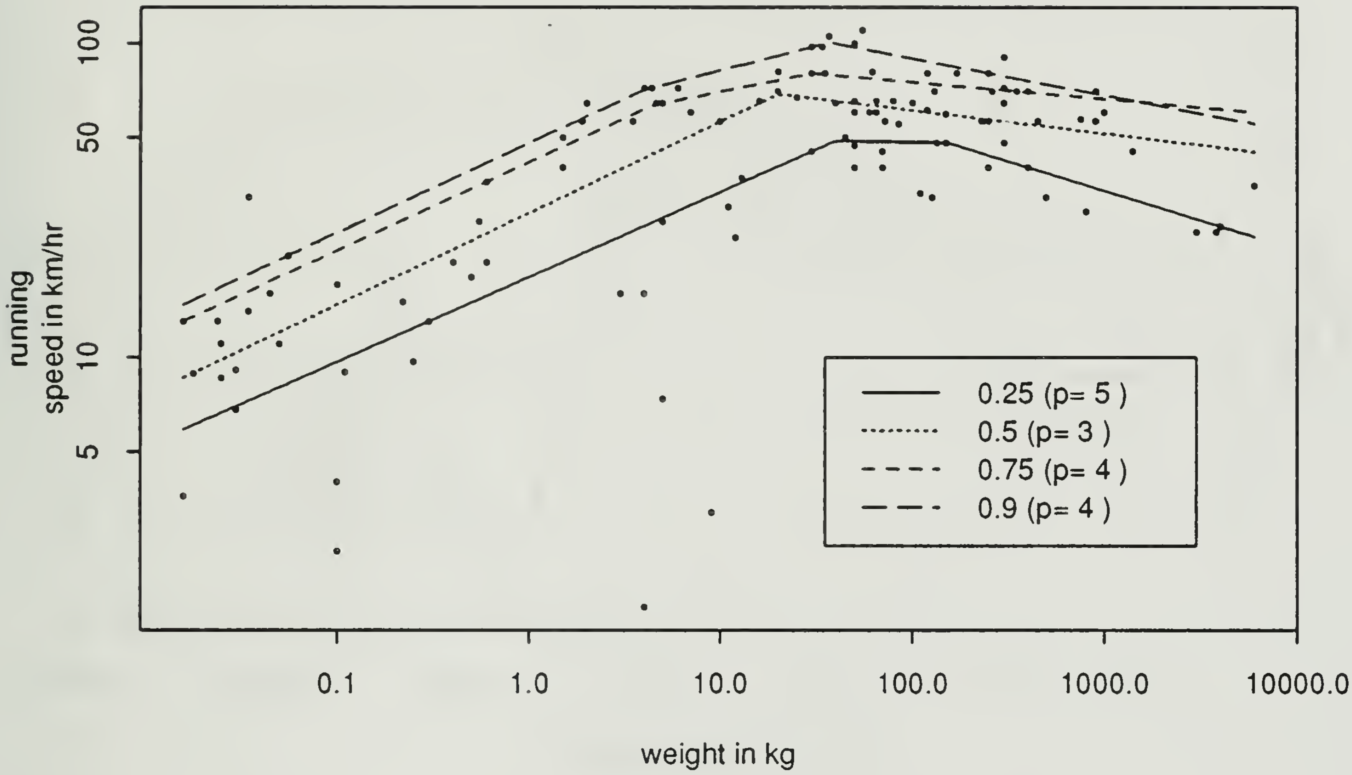

Fig 3.4: Mammal Data and four Quantile L_1 Smoothing Splines

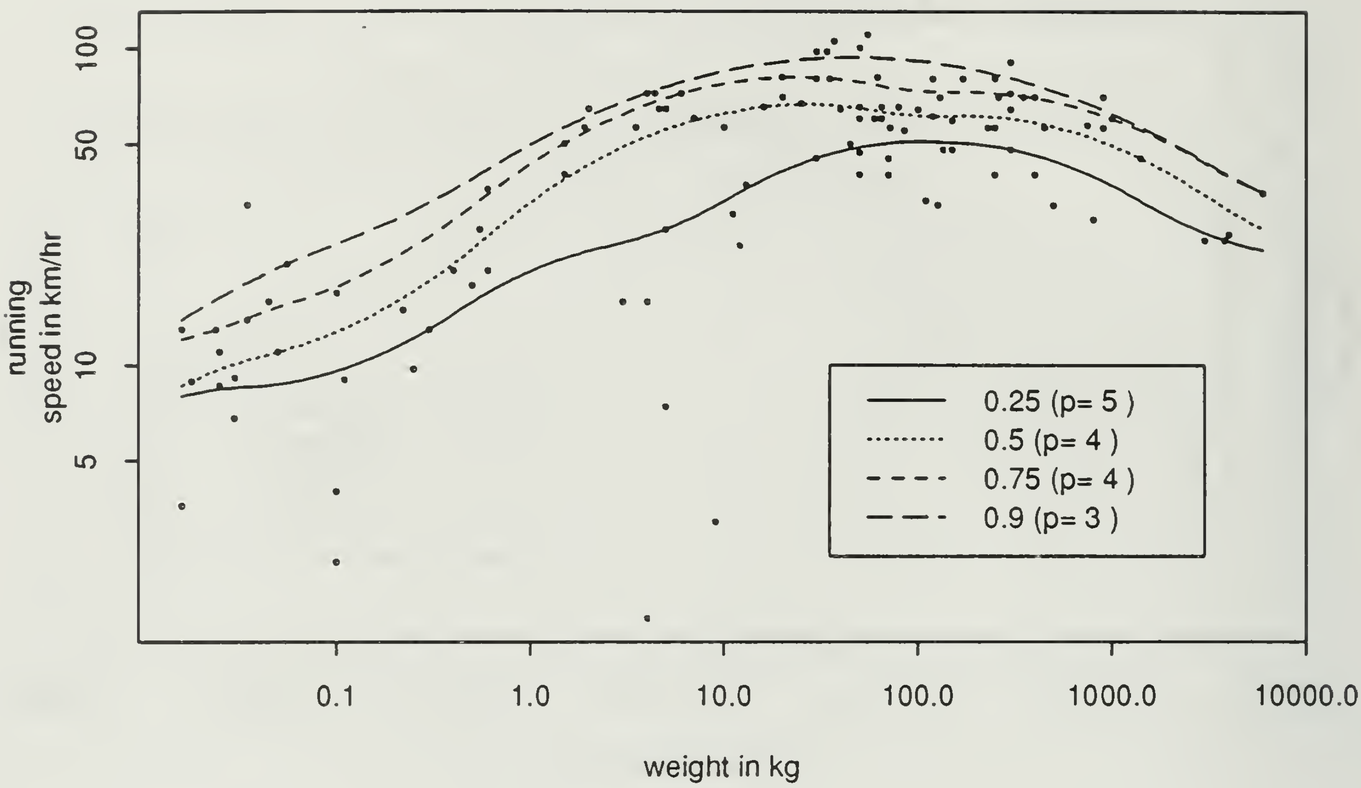

Fig 3.5: Mammal Data and four Quantile L_∞ Smoothing Splines

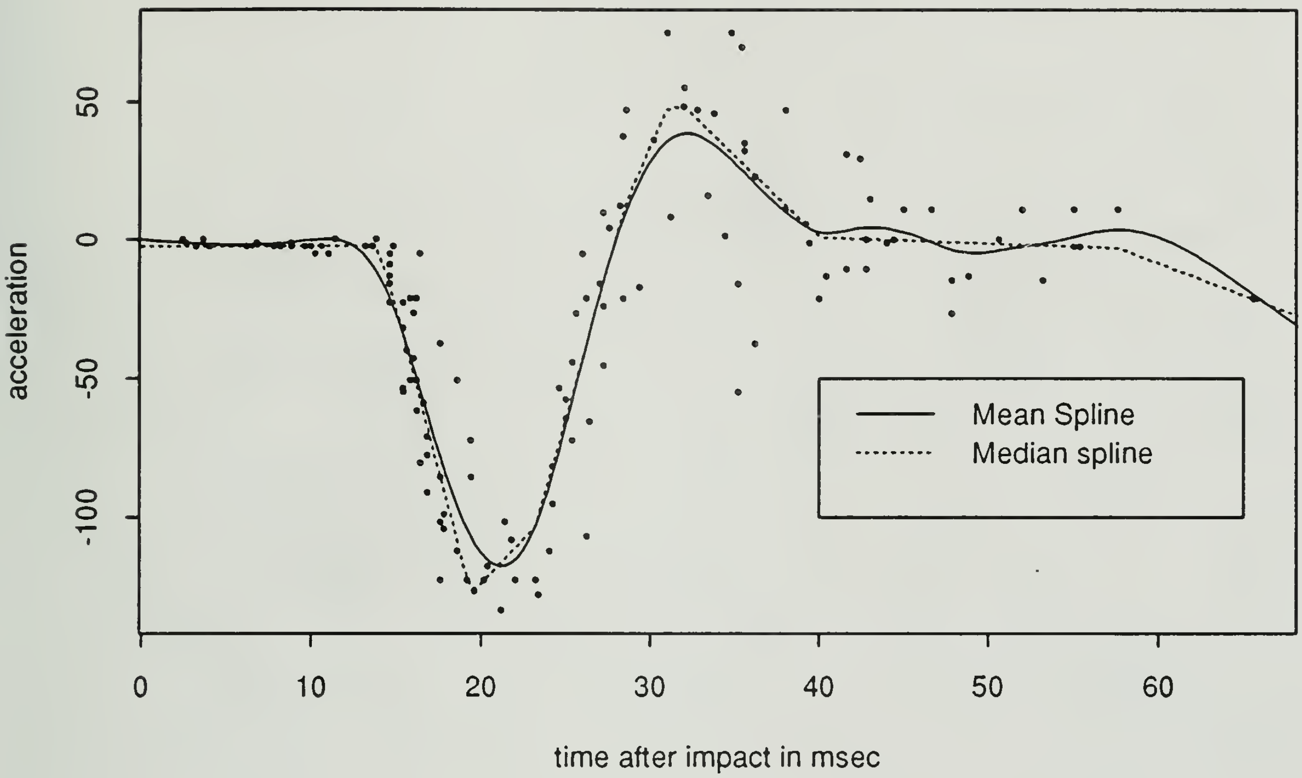

Fig 3.6: Motorcycle Data and Mean and Median Smoothing Splines

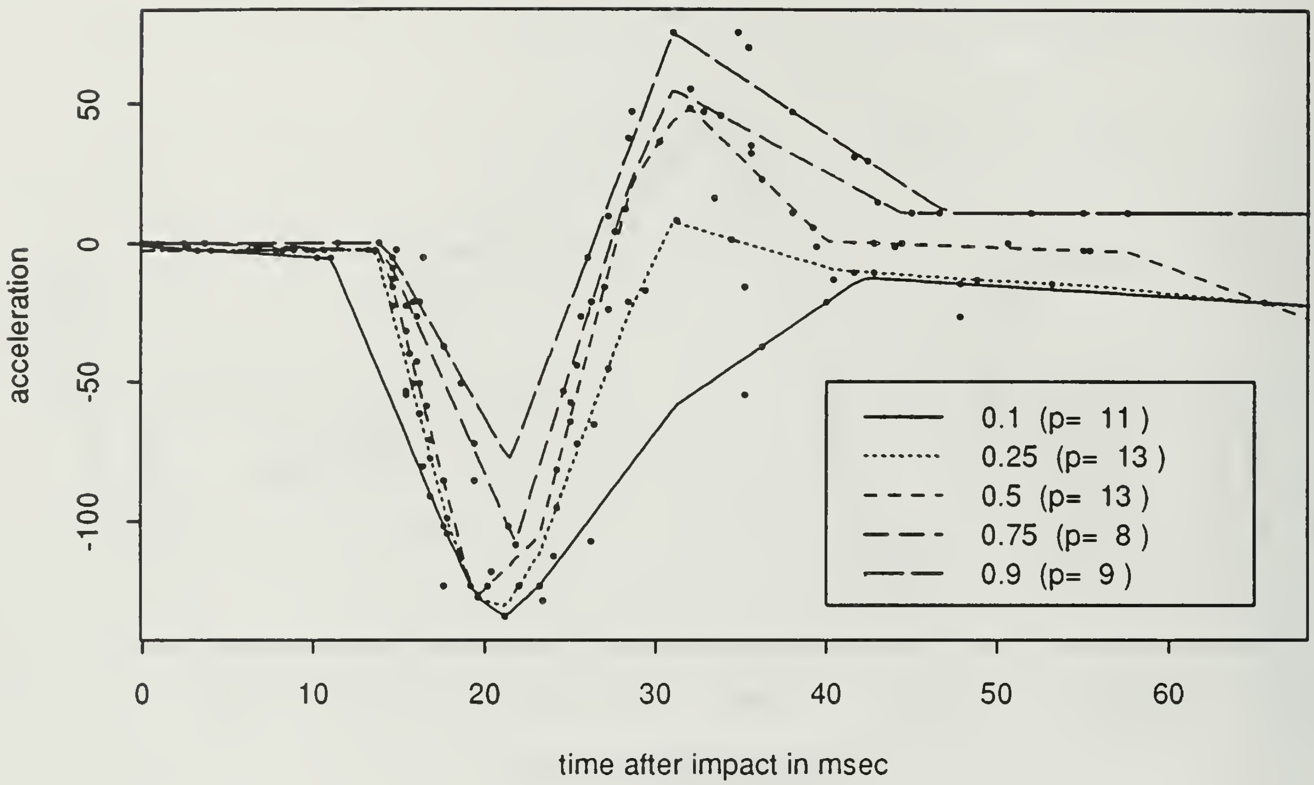

Fig 3.7: Motorcycle Data and four Quantile L_1 Smoothing Splines