ACTUARIAL RESEARCH CLEARING HOUSE
1992 VOL. 1

NONPARAMETRIC REGRESSION METHODS BASED ON REGRESSION QUANTILES

Actuarial Research Conference
August, 1991

Stephen Portnoy
University of Illinois
Department of Statistics, 101 Illini Hall
725 S. Wright Street
Champaign, IL 61820

ABSTRACT

Traditionally, regression models have been analyzed in terms of models for the conditional mean. However, it is becoming increasingly clear that estimation of the conditional median and other quantiles provides a more complete and satisfactory approach to the analysis of regression models. In particular, estimation of conditional quantiles for nonparametric models should provide valuable information concerning departures from standard model assumptions (like non-linearity and heteroscedasticity). Such estimates are also required to analyze larger or smaller conditional quantiles of particular interest by themselves; for example, to model heavy users of some commodity, say, electricity, or behavior of students who score unusually well on a standardized exam.

For linear models, we review the regression quantile estimators introduced by Koenker and Bassett, which are especially natural and quickly computable by linear programming methods. Two approaches to generalizing regression quantile ideas to nonparametric situations are presented. (1) A "spline" approach may be based on minimizing a linear combination of the regression quantile loss function and an L_1-norm smoothing penalty. This approach generalizes the work of Schuette for smoothing actuarial data and provides an rather nice computational algorithm (based on linear programming) for one-dimensional data. (2) A "kernel" approach may be based on minimizing a weighted regression quantile loss function with weights given by a standard kernel. This approach also permits linear programming to provide efficient computation, and allows known results for regression quantiles in non-i.i.d. cases to give a complete asymptotic theory.

1. REGRESSION QUANTILES: THE LINEAR CASE

Koenker and Bassett ((1978) and (1982)) developed an elegant approach to generalizing the notion of "sample quantile" to the regression model. Following their development, consider the model, $Y_i = x_i'\beta + u_i$ for $i = 1, \ldots, n$. Let $0 \leq \theta \leq 1$, and choose $\hat{\beta}(\theta)$ to minimize

$$\sum_{i=1}^{n} \rho_\theta(y_i - x_i'\beta) \ , \qquad \rho_\theta(u) \equiv \theta u^+ + (1-\theta)u^-$$

As an example, consider the 1755 Boscovitch Ellipticity Data (see Stigler, 1986). Figure 1 gives the arc length of one degree of latitude at 5 locations on the surface of the earth. A perfectly spherical earth would lead to a horizontal linear fit. The apparent positive slope indicates ellipticity (in a north-south plane). The plot gives the interval of θ-values for which each indicated linear fit is the regression quantile given by $\hat{\beta}(\theta)$. The Koenker-Bassett approach offers several advantages.

- Natural: $\hat{\beta}(\theta)$ estimates $\beta + \xi_\theta$ where ξ_θ is the θ^{th} quantile of the error distribution

- Computable using linear programming:

$$\text{minimize} \quad \sum_{i=1}^{n} (\theta v_i + (1-\theta) w_i)$$

$$\text{subject to} \quad v_i - w_i = y_i - x_i'\beta \ ; \quad v_i \geq 0 \ ; \quad w_i \geq 0$$

As θ varies, use "parametric programming" to find breakpoints:

$$0 = \theta_0 < \theta_1 < \cdots < \theta_{J_n} = 1 \ , \quad \text{for some} \quad J_n \leq \binom{n}{p}$$

such that $\hat{\beta}(\theta)$ is constant for $\theta_j < \theta < \theta_{j+1}$, and each $\hat{\beta}(\theta)$ is determined by p observations $y_i = x_i'\hat{\beta}$, $i = i_1, i_2, \ldots, i_p$.

Note: $J_n = O_p(n \log n)$ (under conditions).

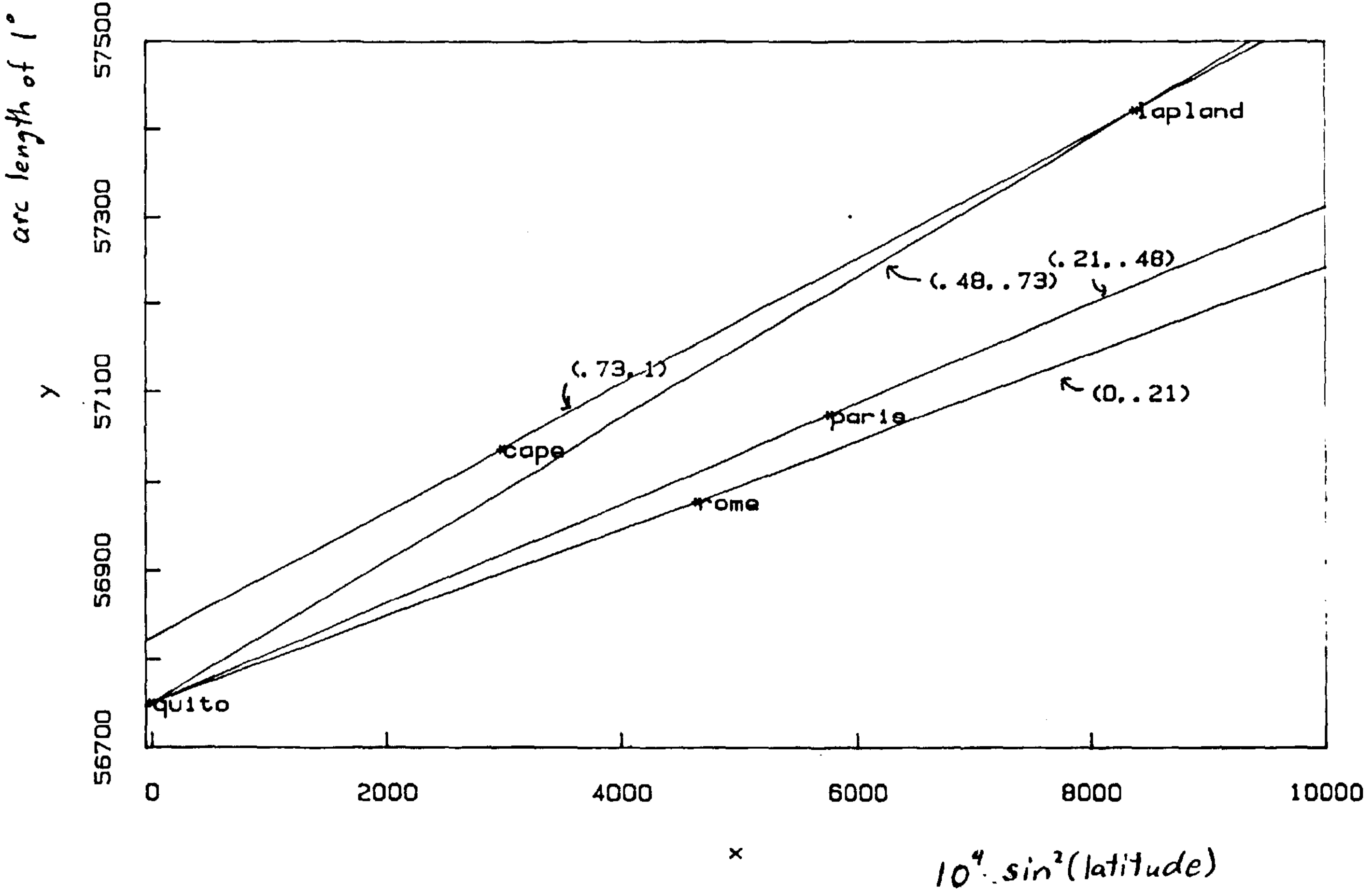
- Provides quantile function estimator:

$$\hat{Q}(\theta) = \bar{x}'\hat{\beta}(\theta) \ ,$$

where $\bar{x}$ is the average of the design vectors x_i. Furthermore, $\hat{Q}$ can be inverted to define an estimator of the error distribution:

$$\hat{F}_n(u) = \inf\{\, \theta : \ u \leq \bar{x}'\hat{\beta}(\theta) \,\}$$

1. Boscovitch Ellipticity Data



Some Applications:

- Trimmed LS (Ruppert-Carroll, 1980): Delete observations below the θ^{th} and above the $(1 - \theta)^{th}$ regression quantile planes and use least squares applied to the remaining observations. Here, the asymptotic results directly generalize those for the one-dimensional trimmed mean, whereas trimming based on residuals gives different asymptotics.

- General L-estimator (Koenker-Portnoy, 1987): Define $\hat{\beta} = \int_0^1 w(\theta)\hat{\beta}(\theta)\, d\theta$, where $w(\theta)$ is an appropriate score function.

- Outlier Identification (Portnoy, 1990): The observations fit by the extreme planes $\hat{\beta}(0)$ and $\hat{\beta}(1)$ can be used to give a high-breakdown method for searching for outliers.

- Tests for non-stationarity: Koenker-Bassett (1982), Portnoy-Welsh (1991), Efron (1991), Portnoy (1991b).

General Asymptotic Result (Portnoy, 1991): Let u_i be **non-stationary** (and "nearly" m-dependent). Let $\bar{F}(\theta) = \frac{1}{n} \sum_{i=1}^{n} F_i(\theta)$ (the average c.d.f. of the errors) and define

$$\beta(\theta) = \beta - (\bar{F}^{-1}(\theta), 0, \cdots, 0)' \qquad R_i(\theta) = x_i\{\theta - I[v_i \leq \bar{F}^{-1}(\theta)]\}$$

$$Q_\theta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i x_i' f_i(\bar{F}^{-1}(\theta)) \qquad b(\theta) = n^{-1/2} Q_\theta^{-1} \sum_{i=1}^{n} E\, R_i(\theta)$$

Then, under conditions, for $\varepsilon \leq \theta \leq 1 - \varepsilon$,

$$n^{1/2}(\hat{\beta}(\theta) - \beta(\theta)) - b(\theta) = n^{-1/2} Q_\theta^{-1} \sum_{i=1}^{n} (R_i(\theta) - ER_i(\theta))$$

$$+ \mathbf{O}_p(n^{-1/4} \log n)$$

$$\to_D \mathbf{N}_p(0, Q_\theta^{-1} \Sigma Q_\theta^{-1})$$

where $\Sigma = \lim_{n \to \infty} \mathrm{COV}\left(n^{-1/2} \sum_{i=1}^{n} R_i(\theta)\right)$

Note: $b(\theta)$ may tend to infinity; so this result permits rather general non-stationarity. It is also possible to apply this result to the nonparametric regression problem discussed in section 4.

2. CONDITIONAL QUANTILES

Given data (Y_i, x_i), define the θ^{th} conditional quantile to be a function $g_\theta(x_i)$ such that $P\{Y_i \le g_\theta(x_i)\} = \theta$; i.e., $Y_i = g_\theta(x_i) + u_i$ where $P\{u_i \le 0\} = \theta$. Such conditional quantiles have many applications.

- *They are useful for assessing departures from model assumptions, especially* "heteroscedasticity". If u_i are identically distributed, all conditional quantiles are parallel. Otherwise they will be non-parallel, and the extent of lack of parallelism provides a test for heteroscedasticity (Efron, 1991; Portnoy, 1991).

- In heteroscedastic cases, the model for the mean may be nonlinear, but there may be some quantile that abeys a linear model. For example, let $Y_i = \alpha + \beta x_i + e^{x_i} u_i$ where u_i are i.i.d. with Negative Exponential distribution. Figure 2 gives several conditional quantile curves. Note that the quantile corresponding to $\theta = 1 - 1/e$ is the only one giving a linear fit.

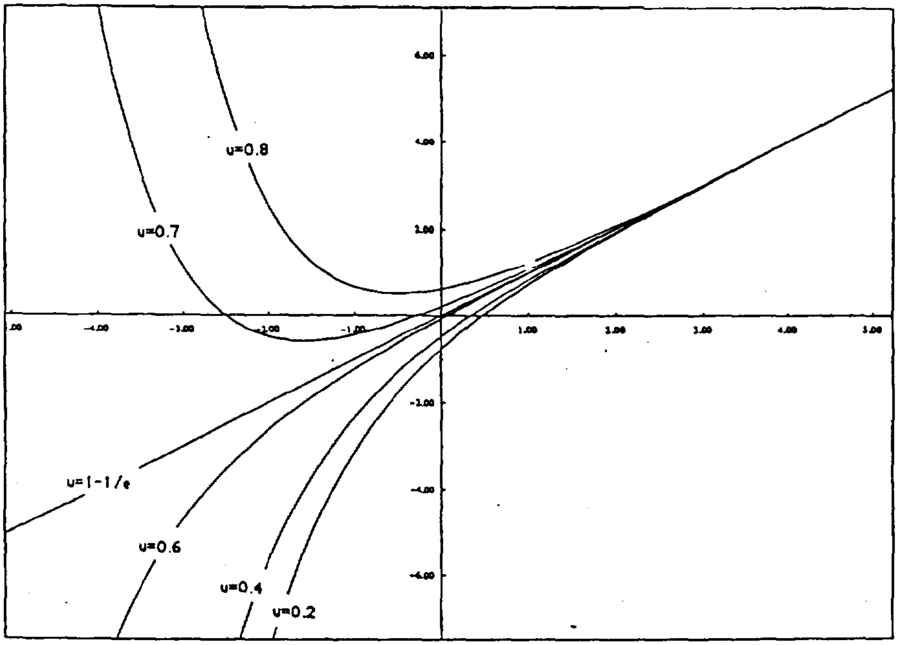


Figure 2. Selected quantile functions of the distribution of Y given x.

- Specific quantiles may be of independent interest. That is, more extreme individuals may depend rather differently on the independent variables than those near the median. For example, considering modelling:

Pollution levels in terms of source or geographic characteristics

Students test performance in terms of study (or social) characteristics

Mortality rates for individuals in a specific risk category -- those at higher (or lower) risk may have very different mortality curves

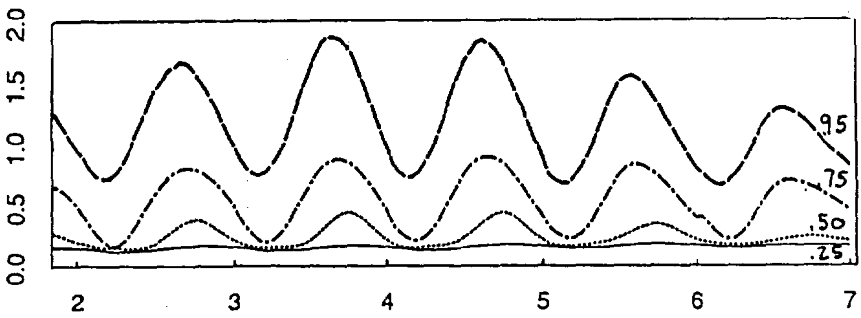
As a specific example, consider the data in Hendricks and Koenker, *JASA* (forthcoming) concerning electricity demand (by household) over time in terms of weather characteristics. Here

$$Y(t) = \sum \alpha_i \ \phi_i(t) + u(t)$$

where $\phi_i(t)$ represent various time periodicities (hourly, daily, weekly) and weather covariates (current temp., max temp. over past 24 hours, etc.)

In figure 3, note the very different response patterns for periods of high demand. Clearly the low demand quantile curve corresponds to background use, while the high demand curves reflect use during active periods of the day (particularly, air conditioning).

Figure 3.



3. SMOOTHING SPLINES

Consider the model,

$$\text{Model:} \qquad Y_i = g(x_i) + u_i$$

In the classical approach, we minimize over appropriate functions $g(x)$

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

This approach has extensive development and theory (Wahba, 1990; Silverman, 1985)

virtues:

- Elegant theory (Reproducing Kernel Hilbert Space)
- Sparse linear computations
- Optimality at Gaussian

problems:

- *Squared error and L_2 penalty chosen for math convenience* -- hard to interpret and *not* "natural"
- Not robust (poor if u_i non-normal)
- Computational problems as λ varies (each λ gives a separate problem)
- Hard to include monotonicity or convexity constraints (quadratic programming may be quite difficult)
- Need specific distributional assumption to estimate quantiles

We can try replacing squared error with $\sum \rho_\theta(y_i - g(x_i))$, but this still leads to a difficult quadratic programming problem. Thus, consider minimizing (over g):

$$\sum \rho_\theta(y_i - g(x_i)) + \lambda \int |g''(x)| \, dx$$

Result: The optimal $\hat{g}$ has the form

$$g(x) = \alpha_i (x - x_i)^2 + \beta_i (x - x_i) + \gamma_i$$

$$\text{for} \quad x_i \le x < x_{i+1} \qquad i = 1, \ldots, n$$

(see Koenker and Ng, 1991)

Thus, the problem yields the Schuette (1978) formulation

$$\min \quad \sum \rho_\theta(y_i - \gamma_i) + \lambda \sum \Delta_i |\alpha_i|$$

subject to continuity constraints; here $\Delta_i \equiv x_{i+1} - x_i$.

The solution is again given by **parametric programming** (in λ):

$$\min \quad \sum(\theta u_i + (1-\theta)v_i) + \lambda \sum \Delta_i(a_i + b_i) \qquad \text{subject to}$$

$$u_i - v_i = y_i - \gamma_i \qquad u_i \geq 0 \qquad v_i \geq 0$$

$$a_i - b_i = \alpha_i \qquad a_i \geq 0 \qquad b_i \geq 0 \qquad \alpha_0 = \alpha_n = 0$$

$$\alpha_i \Delta_i^2 + \beta_i \Delta_i + \gamma_i = \gamma_{i+1} \qquad 2\alpha_i \Delta_i + \beta_i = \beta_{i+1}$$

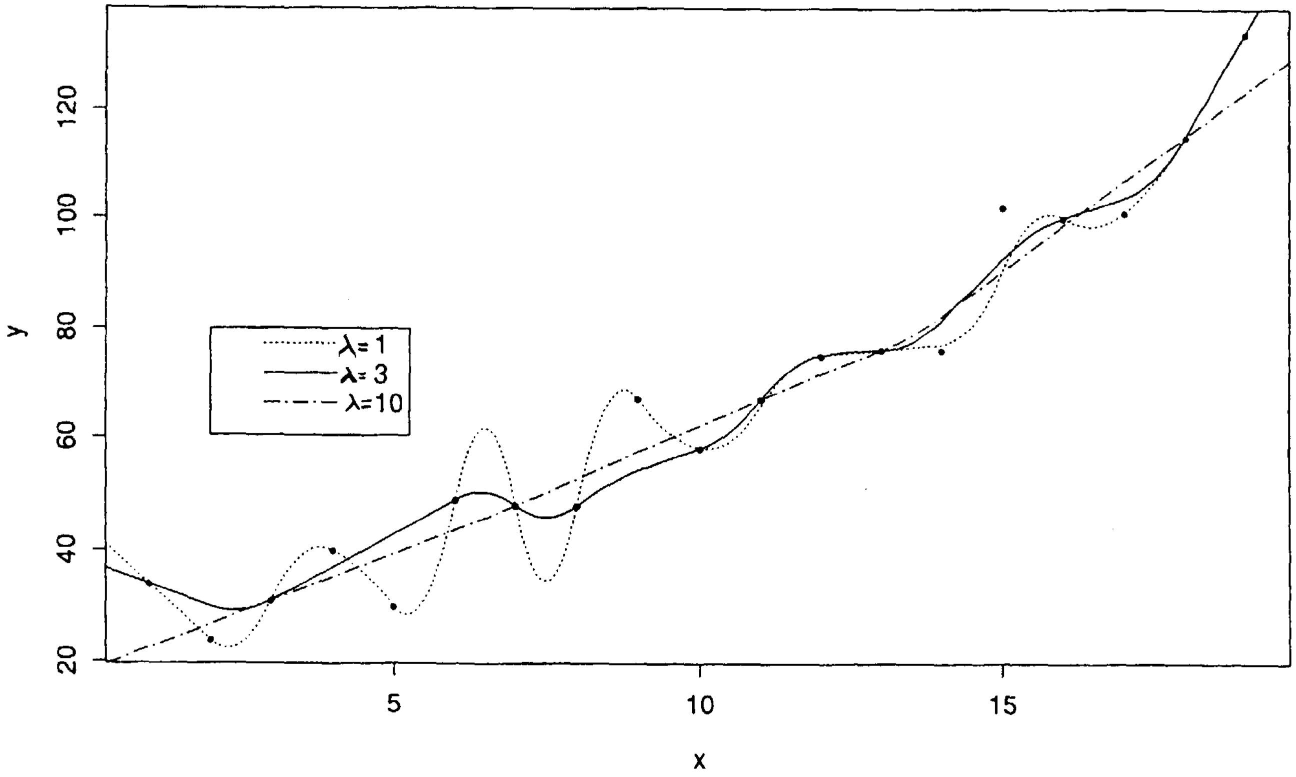
Advantages:

- natural: absolute discrepancies are more intuitive
- robust to outlying errors
- gives quantiles directly
- computationally easy: λ small -- $g(x)$ fits all obs., λ large -- $g(x)$ = global linear fit

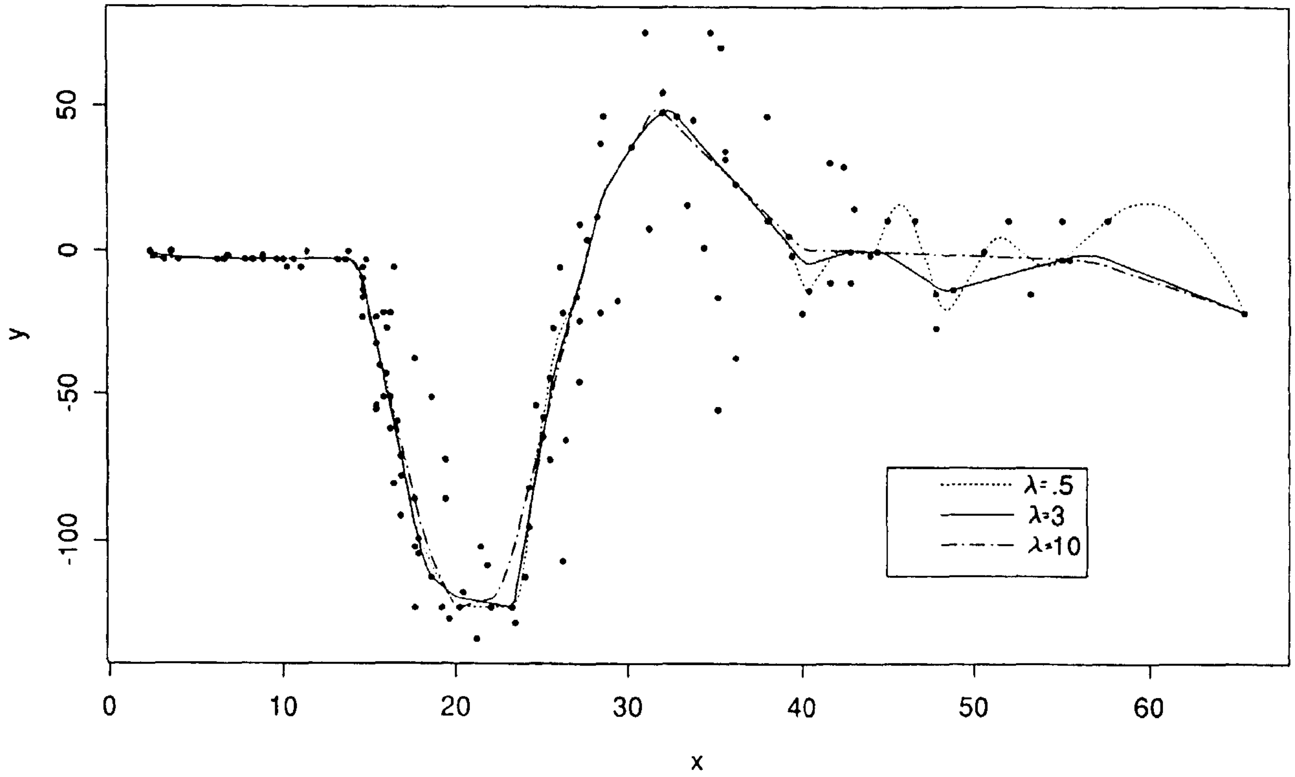
  As λ decreases, get breakpoints and new solution with one simplex pivot (so can get solutions for all λ)
- monotonicity and convexity constraints are easy: still get parametric *linear* programming problem!
- dimensionality: as λ increases, the solution at each new breakpoint either replaces a quadratic segment by a linear one or interpolates one new observation; thus permitting a direct assessment of dimension as the difference between the number of interpolated observations and the dimension of the interpolating segments.

Figures 4 and 5 give examples of median smoothing splines for two data sets. Figure 6 gives the estimated 10th, 50th, and 90th percentile curves corresponding to a specific value of the smoothing parameter, λ.

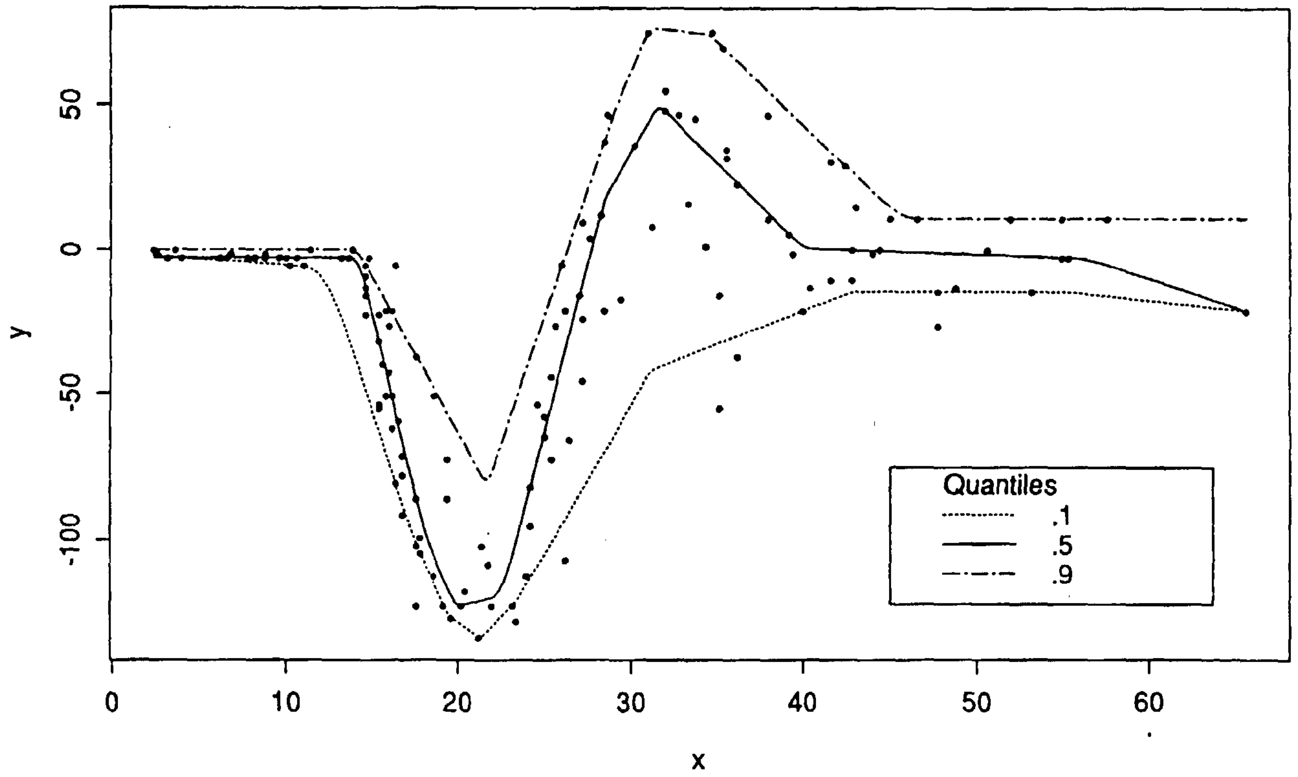
4. Median Smoothing Splines for Schuette's Example 1



5. Median Smoothing Splines for Motorcycle Data



6. Quantile Smoothing Splines for Motorcycle Data



4. A KERNEL METHOD: WEIGHTED REGRESSION QUANTILES

Choose $\hat{\beta}(x;\theta)$ to minimize (over β):

$$\sum w_i(x)\ \rho_\theta(y_i - x_i'\beta)$$

where for some kernel k and window width h_n,

$$w_i(x) \equiv \frac{1}{h_n}\, k\left[\frac{x - x_i}{h_n}\right] .$$

Assume $h_n \to \infty$, k has bounded support, and that $\sum x_i w_i(x) \to x$ (as $n \to \infty$). Then

$$\hat{g}_\theta(x) \equiv x'\hat{\beta}(x;\theta)$$

provides a nonparametric estimator of the conditional quantile.

Advantages:

- multivariate extension immediate
- Theory straightforward: the non-i.i.d. theorem applies directly with bias

$$b(x) = O_p(h_n^2) .$$

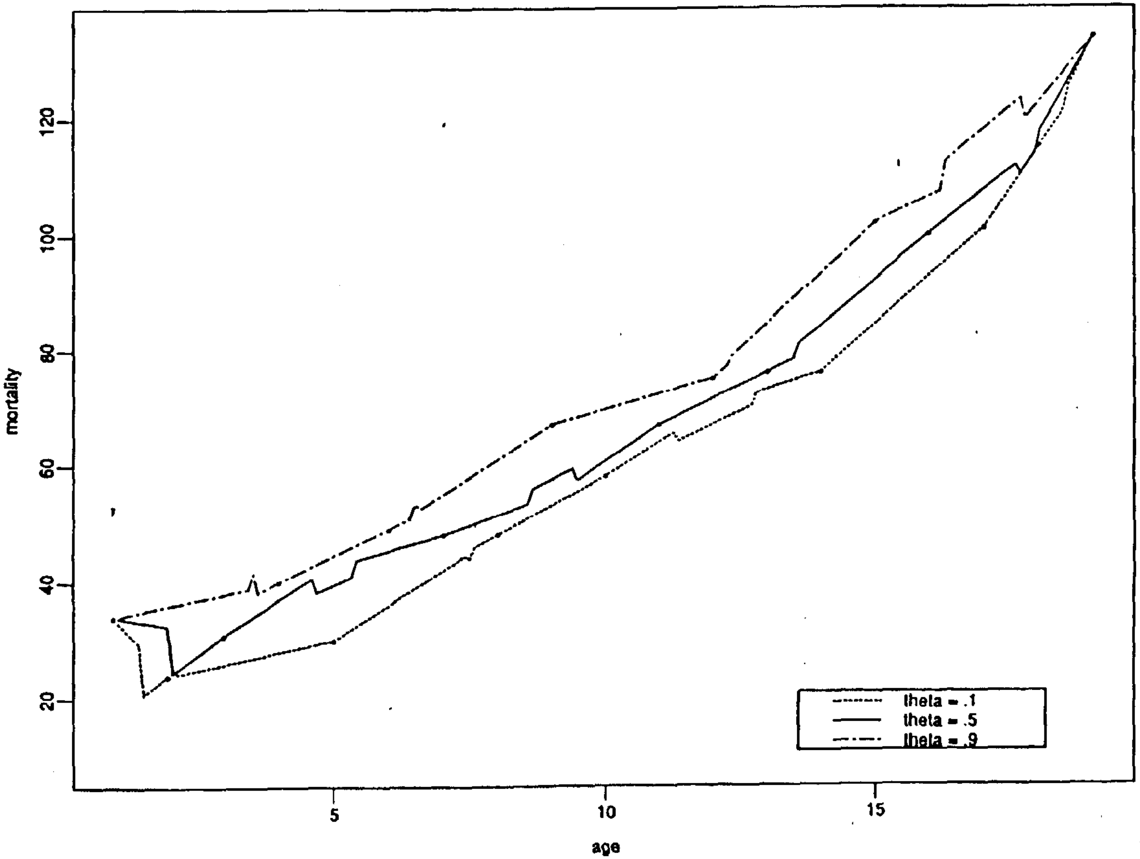
Theory may extend to splines using asymptotic equivalence of the methods (Messer, 1991).

Problems:

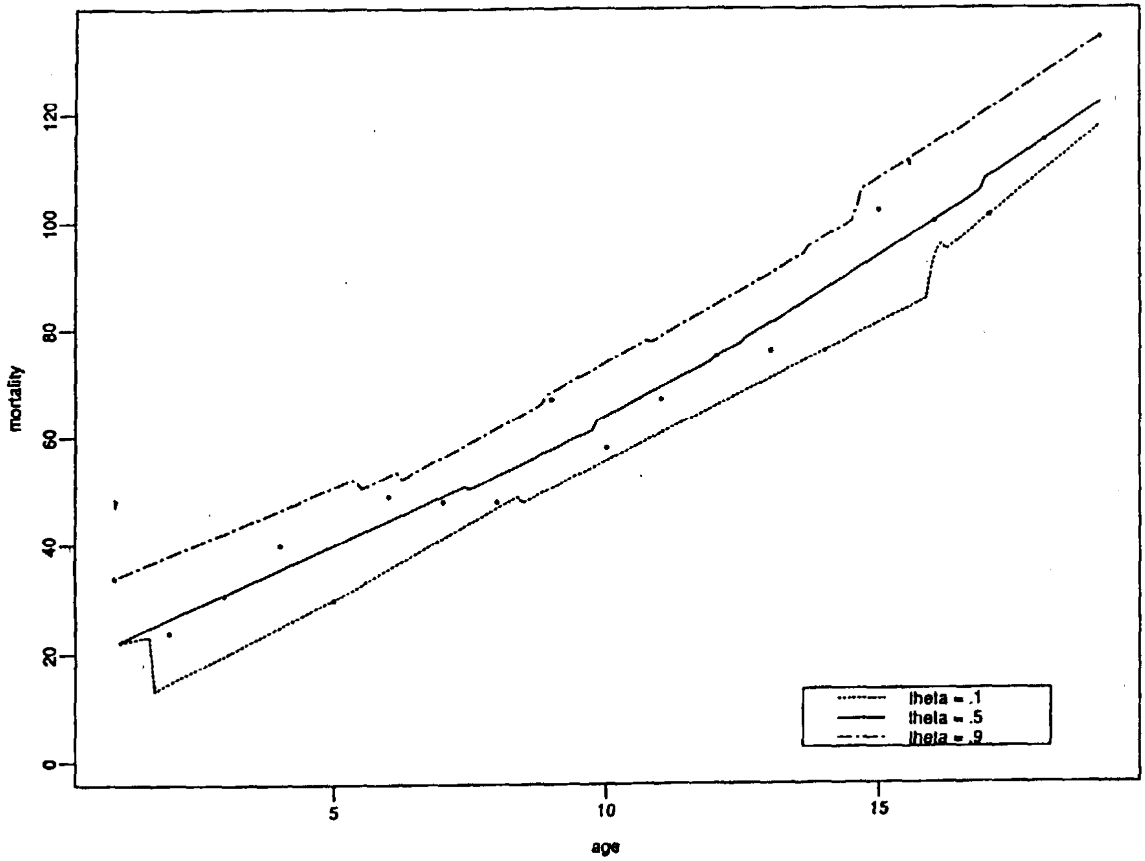
- oversmoothing at local max or min (see following example)
- choice of k and h_n (although known approaches should still work)
- how to predict outside convex hull of observations (e.g., future predictions for time series)
- computational: separate linear program for each x (in a given grid)

Some examples follow. Figures 7 and 8 give estimated quantile curves for the Schuette data using a triangular kernel and a logistic kernel respectively. The lack of smoothness requires some further work: perhaps using locally weighted quadratic regression quantiles (instead of linear ones) would provide improvement. Comments based on using a uniform kernel in Chaudhuri (1991) and by Wang and Scott (1991) suggest this is so. Figures 9, 10, and 11 give estimated quantile curves for the Motorcycle data for different window widths, h. Adaptive choice of h would probably improve the estimates substantially; but clearly choosing h large enough to smooth out the breaks results in significant oversmoothing. This method may not be appropriate for producing smooth curves, but should be useful for prediction at specific x-values since the asymptotic theory shows that the conditional quantile estimators can be used directly to provide predicition intervals.

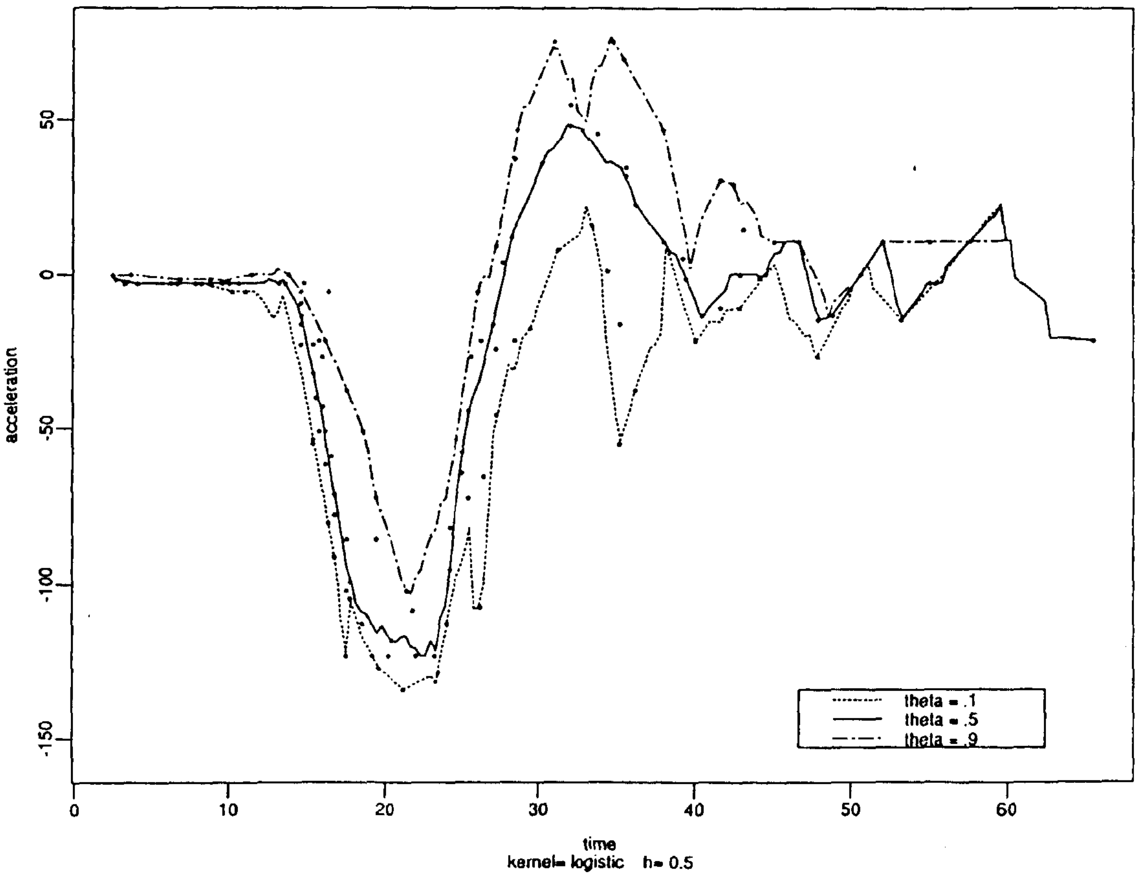
7. SCHUETTE DATA (triangular kernel)



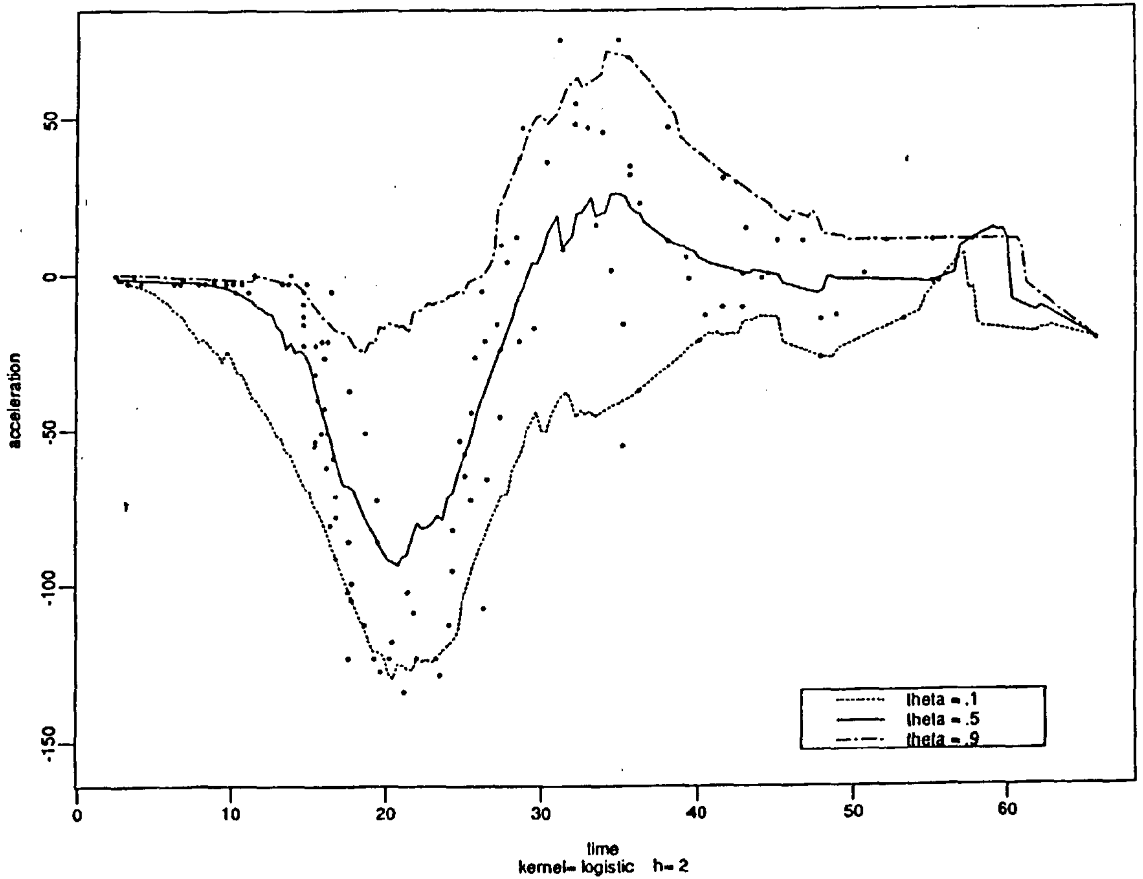
8. SCHUETTE DATA (logistic kernel)



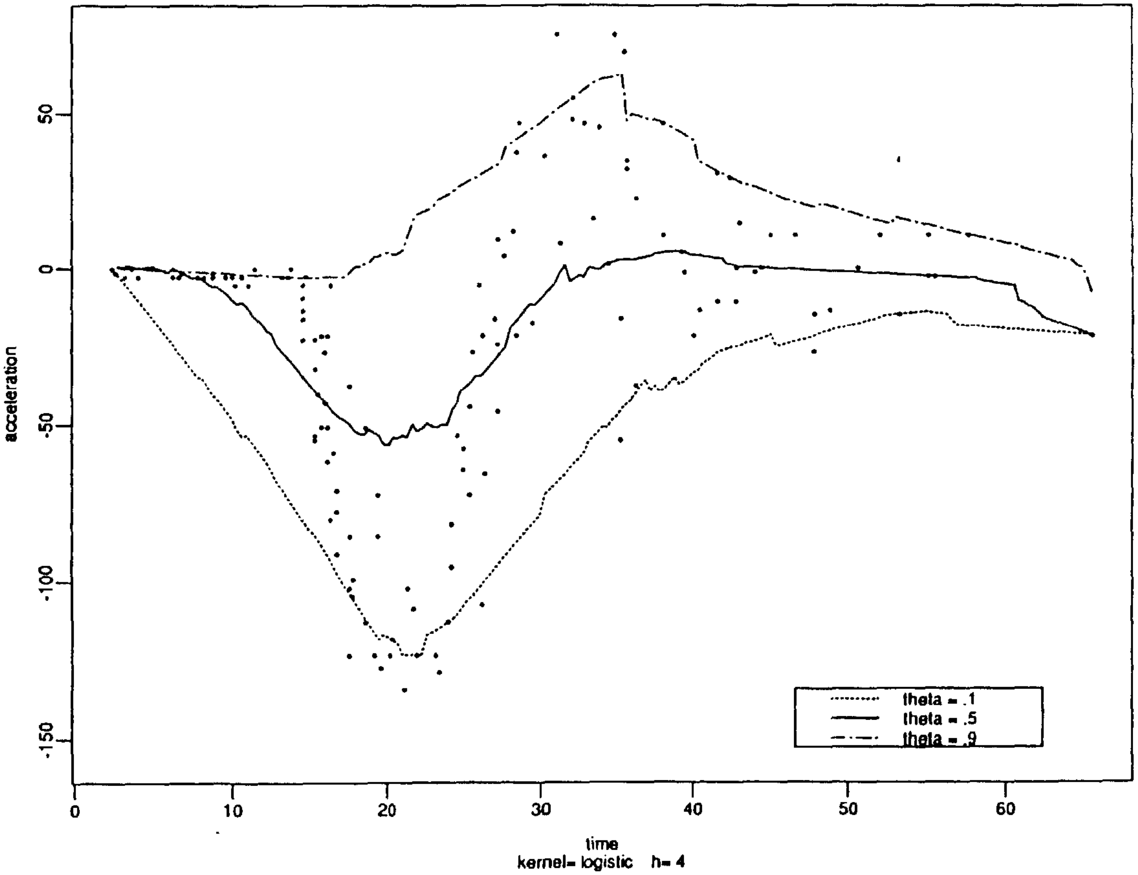
9. MOTORCYCLE DATA



10. MOTORCYCLE DATA



11. MOTORCYCLE DATA



REFERENCES

Bassett, G.W. and Koenker, R.W. (1982). An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association*. 77, 407-415.

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

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

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

Koenker, R. and Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles, *Econometrica, 50*, 43-61.

Koenker, R. and Ng, P. (1991). Quantile Smoothing Splines, to appear in: *Nonparametric Statistics and Related Topics* (ed: A.K.Md.E. Saleh), North Holland: New York.

Koenker, R.W. and Portnoy, S. (1987). L-Estimation for the Linear Model., *Journal of the American Statistical Association, 82*, 851-857.

Messer, K. (1991). A comparison of a spline estimate to its equivalent kernel estimate, *Ann. Statist., 19*, 817-829.

Portnoy, S. (1990). Regression quantile diagnostics for multiple outliers, *Directions in Robust Statistics and Diagnostics, II* (ed: Stahel and Weisberg), Springer-Verlag, New York.

Portnoy, S. (1991). Behavior of regression quantiles in non-stationary, dependent cases, *J. Multivar. Anal., 38*, 100-113.

Portnoy, S. (1991b). A regression quantile based statistic for testing non-stationarity of errors, to appear in: *Nonparametric Statistics and Related Topics* (ed: A.K.Md.E. Saleh), North Holland: New York.

Portnoy, S. and Welsh,. A. (1991). Exactly what is being modelled by the systematic component in a heteroscedastic linear regression; to appear: *Prob. Stat. Letters.*

Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association.* 75, 828-838.

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

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

Stigler, S.M. (1986) *The History of Statistics.* Cambridge: Harvard Press.

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

Wang, F. and Scott, D. (1991). The L_1 method for robust nonparametric regression, preprint.