The Statistician (2003)
52, *Part* 3, *pp.* 331–350

Quantile regression: applications and current research areas

Keming Yu,

University of Plymouth, UK

Zudi Lu

Chinese Academy of Mathematics and System Sciences, Beijing, People's Republic of China

and Julian Stander

University of Plymouth, UK

[Received September 2001. Revised January 2003]

Summary. Quantile regression offers a more complete statistical model than mean regression and now has widespread applications. Consequently, we provide a review of this technique. We begin with an introduction to and motivation for quantile regression. We then discuss some typical application areas. Next we outline various approaches to estimation. We finish by briefly summarizing some recent research areas.

Keywords: Check function; Conditional distribution; Quantile; Regression fitting; Skew distribution

1. Statistical motivations and basic concepts of quantile regression

1.1. Population and sample quantiles

The term quantile is synonymous with percentile; the median is the best-known example of a quantile. We know that the sample median can be defined as the middle value (or the value half-way between the two middle values) of a set of ranked data, i.e. the sample median splits the data into two parts with an equal number of data points in each. Usually, the sample median is taken as an estimator of the population median m, a quantity which splits the distribution into two halves in the sense that, if a random variable Y can be measured on the population, then $P(Y \leqslant m) = P(Y \geqslant m) = \frac{1}{2}$. In particular, for a continuous random variable, m is a solution to the equation $F(m) = \frac{1}{2}$, where $F(y) = P(Y \leqslant y)$ is the cumulative distribution function. As an example of the use of the median, consider the distribution of salaries. This is typically skewed to the right since relatively few people earn large salaries. As a consequence, the sample median provides a better summary of typical salaries than the mean.

More generally, the 25% and 75% sample quantiles can be defined as values that split the data into proportions of one- and three-quarters, and vice versa. Correspondingly, in the continuous case the population lower quartile and upper quartile are the solutions to the equations $F(y) = \frac{1}{4}$ and $F(y) = \frac{3}{4}$ respectively. Generally, for a proportion p $(0 < p < 1)$, and in the contin-

Address for correspondence: Keming Yu, Department of Mathematics and Statistics, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK.
E-mail: kyu@plymouth.ac.uk

 0039–0526/03/52331

uous case, the $100p\%$ quantile (equivalently, the $100p$th percentile) of F is the value y which solves $F(y) = p$; we assume that this value is unique.

A further generalization is exemplified by studies of the growth of children, where it is often of interest to know the position of a particular child relative to the overall distribution of heights of children of that age. The upper and lower quartiles for the conditional distribution of heights Y given age X therefore provide useful information. In this case the quantiles depend on the value of the covariate X and can be found by solving $F(y|x) = p$, where $F(y|x) = P(Y \leqslant y|X = x)$. Clearly, extensions to several covariates are then possible; see below.

1.2. From standard regression to quantile regression

Regression is used to quantify the relationship between a response variable and some covariates. Standard regression has been one of the most important statistical methods for applied research for many decades. For example, some years ago a university union wished to examine the relationship between the earnings of professors and the number of years that they had been professors. The union collected data on the salaries of 459 US statistics professors and the number of years for which they had been professors during the period from 1980 to 1990; see Bailar (1991). A standard linear regression model for this is

$$y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon, \tag{1}$$

where $\mathbf{x} = (1, x)^{\mathrm{T}}$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}}$, y is salary, x is the number of years as professor and ε is a Gaussian error. More complicated models, such as polynomial regression models, may also be used to model this relationship. In Fig. 1 we present the data, together with the best fitting quadratic regression curve.

Unfortunately, the curves in Fig. 1 provide an inadequate picture of the salary distribution in the sense that the change in shape of the salary distribution with years as professor is not displayed. This is simply because the standard regression fit models only the average relationship between salary and years as professor.

To give a more complete picture of the relationship between salary and years as professor, we present in Fig. 2 the 25%, 50% and 75% sample quantiles. The resulting curves are called quantile regression curves and clearly can be smoothed in some way. However, the change in shape of the salary distribution is much more clearly displayed. An approximation of the full salary probability distribution can be produced from the quantile regression curve corresponding to a range of values of p. In the next section we provide a formal definition of the $100p\%$ quantile regression curve.

1.3. From least squares estimation to the 'check function'

To discuss the basic motivation for the estimating method for quantile regression, we begin with the well-known least squares estimation.

Consider the simple regression model (1). The parameter vector $\boldsymbol{\beta}$ is usually estimated through the quadratic loss function $r(u) = u^2$, i.e., given a data set of observations $\{\mathbf{x}_i, y_i\}_{i=1}^n$, estimation is performed by minimizing

$$\sum_{i=1}^{n} r(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})^2$$

over $\boldsymbol{\beta}$.

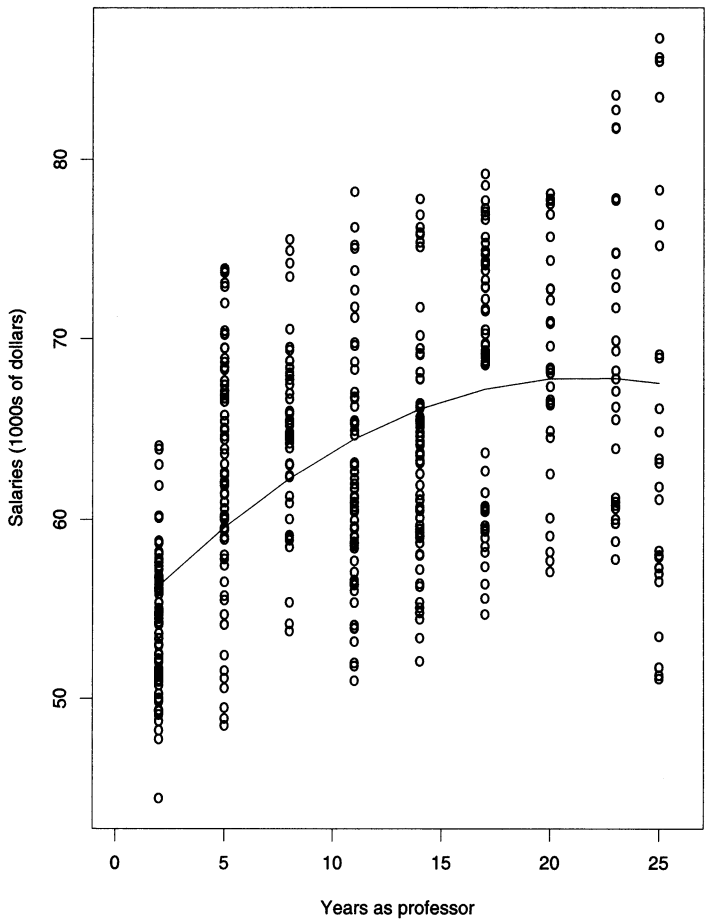

Fig. 1. Salaries of 459 US statistics professors as a function of years as professor with the fitted quadratic regression curve

Least squares regression is concerned with the estimation of the conditional expectation $E[Y|\mathbf{X}=\mathbf{x}]$, since this conditional expectation is the value of θ that minimizes the expected square loss function $E[(Y-\theta)^2|\mathbf{X}=\mathbf{x}]$ and $\Sigma_{i=1}^{n} r(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ is a sample estimate of this.

Similarly, median regression estimates the conditional median of Y given $\mathbf{X}=\mathbf{x}$ and corresponds to the minimization of $E[|Y-\theta||\mathbf{X}=\mathbf{x}]$ over θ. An associated loss function is $|u|$. However, it is more convenient to take the loss function to be $\rho_{0.5}(u)=0.5|u|$. Estimation proceeds by minimizing $\Sigma_{i=1}^{n} \rho_{0.5}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ over $\boldsymbol{\beta}$. We may rewrite $\rho_{0.5}(u)$ as $\rho_{0.5}(u)= 0.5u\, I_{[0,\infty)}(u) - (1-0.5)u\, I_{(-\infty,0)}(u)$, where

$$I_A(u) = \begin{cases} 1 & u \in A, \\ 0 & \text{otherwise} \end{cases}$$

is the usual indicator function of the set A. This definition may be generalized by replacing 0.5 by p to obtain a characterization of $100p\%$ quantile regression $q_p(\mathbf{x})$ at $\mathbf{x}$ as the value of θ that minimizes

$$E[\rho_p(Y-\theta)|\mathbf{X}=\mathbf{x}],$$

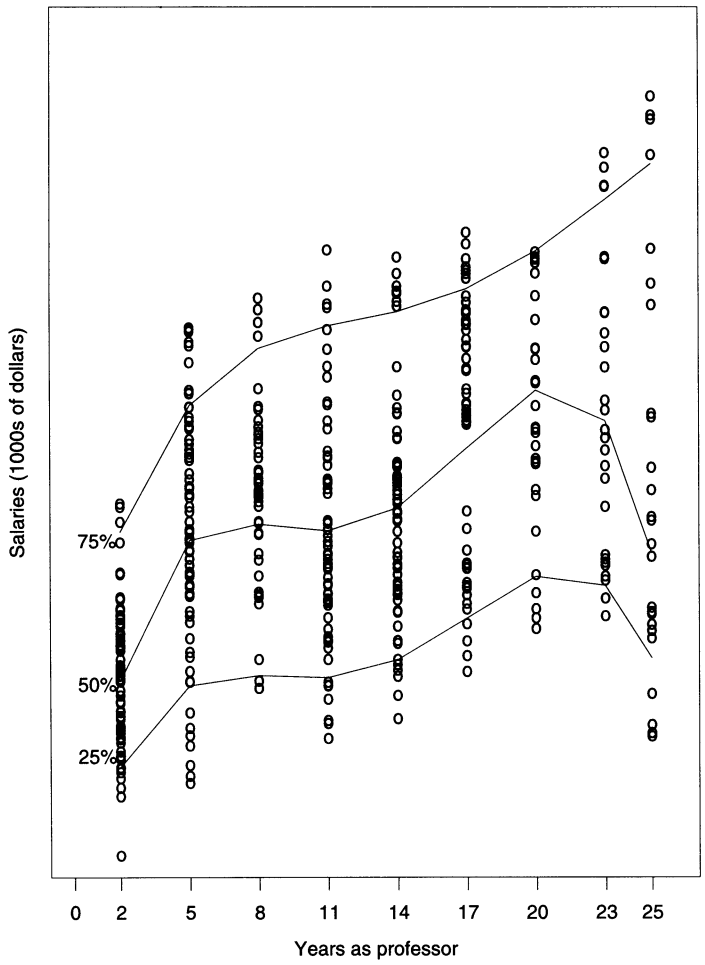

Fig. 2. Salaries of 459 US statistics professors as a function of years as professor: three quantile regression curves with $p = 0.25, 0.5, 0.75$ are shown

where

$$\rho_p(u) = pu\, I_{[0,\infty)}(u) - (1-p)u\, I_{(-\infty,0)}(u) \tag{2}$$

is called the 'check function'.

1.4. From conditional skew distributions to quantile regression

Fig. 3(a) displays weight against age for a sample of 4011 US girls (Cole, 1988). The intuitively reasonable notion of a relationship between weight and age is further supported by Fig. 3(b) which presents several smoothed quantile regression curves based on $p = 0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97$. These suggest that the associated conditional distributions are skew to the right. Two questions of interest are

(a) what is a typical weight profile as a function of age and
(b) what is a typical weight profile as a function of age for overweight and underweight people?

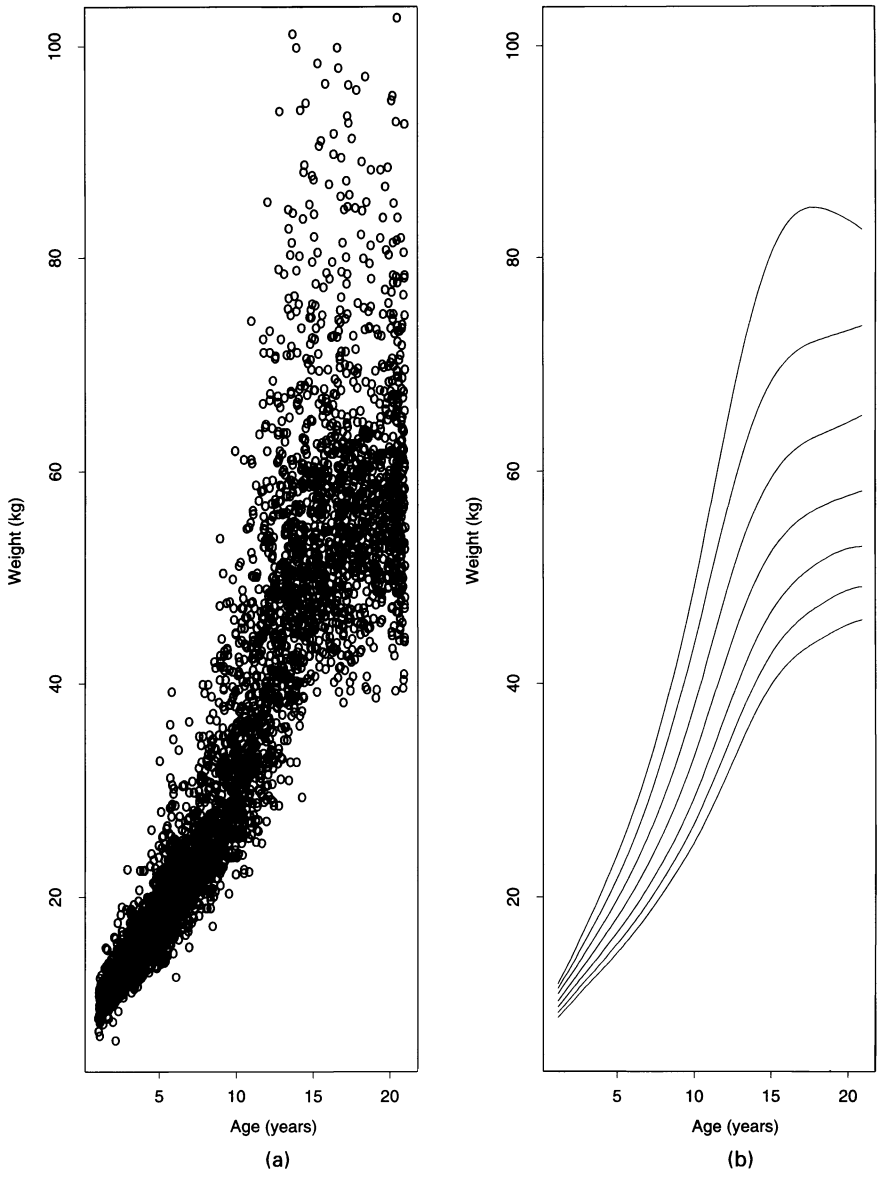

Fig. 3. Weight and age for a sample of 4011 US girls together with seven (p = 0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97) quantile regression curves: the curves were obtained by using the kernel smoothing method discussed in Section 3.3

A sensible answer to the first question is not provided by standard mean regression, as the mean at any specific year is pulled downwards. Hence, the median curve is a more appropriate curve to display. This median curve corresponds to the middle quantile regression curve displayed in Fig. 3(b). If it is thought that girls whose weights lie on or above the 97% curve for the population are overweight, then the appropriate curve to display is that based on quantile regression with $p = 0.97$. Similarly, the $p = 0.03$ quantile regression curve displays the relationship of the weight of underweight girls with age. In Section 3, we shall discuss the estimation of these smooth curves in detail.

1.5. From Gaussian likelihood to asymmetric Laplace density

Maximum likelihood estimation is one of the most popular methods for statistical inference. Consider again the regression model (1), where the model error $\varepsilon \sim N(0, \sigma^2)$ follows a Gaussian distribution with known standard deviation σ. The likelihood function for β based on a sample $\{\mathbf{x}_i, y_i\}_{i=1}^n$ from this model is given by

$$L(\beta) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\beta)^2\right\}.$$

Maximizing $L(\beta)$ over β yields the least squares estimates that were mentioned in Section 1.3. If we now assume that the model error ε has probability density function

$$f(\varepsilon) \propto \exp\left\{-\sum_{i=1}^{n}\rho_p(y_i - \mathbf{x}_i^{\mathrm{T}}\beta)\right\}$$

with ρ_p given in equation (2), then maximizing the associated likelihood function is equivalent to minimizing the check function. In fact, there is a standard probability density called the asymmetric Laplace density which takes the form

$$f(\varepsilon) = p(1-p)\exp\{-\rho_p(\varepsilon)\},$$

and realizations from this density can be simulated via the simple linear combination

$$\frac{1}{p}U - \frac{1}{1-p}V$$

of two independent exponential random variables U and V each with mean 1 (Yu and Moyeed, 2001).

1.6. From contaminated data to robust estimation

It is well known that a sample median is more robust to outlying observations than a sample mean for estimating the average location of a population. Suppose that a large, random batch of mixed 'good' and 'bad' pairs of independent observations (x_i, y_i), $i = 1, 2, \ldots, n$, is available for estimating the conditional mean $E[Y|X = x]$. We assume that a pair (x_i, y_i) is bad with probability π and good with probability $1 - \pi$, and that pairs (x_i, y_i) are distributed as

$$(X, Y) \sim \begin{cases} N(0, 0, r, 1, 1) & \text{if } (x_i, y_i) \text{ is good,} \\ N(0, 0, r, k, k) & \text{if } (x_i, y_i) \text{ is bad,} \end{cases}$$

where $N(\mu_1, \mu_2, r, \sigma_1^2, \sigma_2^2)$ denotes a bivariate normal distribution with correlation coefficient r, means μ_1 and μ_2 and variances σ_1^2 and σ_2^2. Thus, the (x_i, y_i) are independent realizations from the common underlying 'contaminated density'

$$f(x, y) = (1 - \pi)\, f_1(x, y) + \pi\, f_2(x, y),$$

where f_1 and f_2 are the density functions for $N(0, 0, r, 1, 1)$ and $N(0, 0, r, k, k)$ respectively. For a genuinely contaminated distribution, we assume that $k \neq 1$. A simple theoretic investigation presented by Yu and Jones (1998) established that the variance of a typical kernel smoother such as `ksmooth` of S-PLUS is much greater than the variance of a smooth quantile regression curve. This is further evidence that quantile regression is much more stable than mean regression for analysing this kind of 'contaminated data'.

1.7. Plan of the paper

So far we have defined and motivated quantile regression. In Section 2 we present some typical applications of quantile regression based on simple examples, illustrating how this methodology can be used effectively in many fields. Some methods and algorithms for estimation are sketched in Section 3. In Section 4 we briefly mention some current research areas, which include time series, statistical testing, additive models and Bayesian inference. We present our conclusions in Section 5.

2. Applications of quantile regression

In this section we present some typical applications of quantile regression to medical reference charts, survival analysis, financial economics, environmental modelling and the detection of heteroscedasticity.

2.1. Applications to reference charts in medicine

In medicine, reference (or centile) charts provide a collection of useful quantiles. These are widely used in preliminary medical diagnosis to identify unusual subjects in the sense that the value of some particular measurement lies in one or other tail of the appropriate reference distribution. The need for quantile curves rather than a simple reference range arises when the measurement (and hence the reference range) is strongly dependent on a covariate such as age, as Cole and Green (1992) and Royston and Altman (1994) have discussed. The chosen quantiles are usually a symmetric subset of $\{0.03, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.97\}$. An example of a reference chart is shown in Fig. 3, with the Y-variable being weight and the X-variable being age. How can these quantile regression curves be obtained?

An obvious approach is to use a known conditional distribution $F(y|x)$ to fit the underlying conditional distribution. The $100p\%$ quantile curve corresponds to $q_p(x) = F^{-1}(p|x)$. Now, if the distribution is normal, then estimating the $100p\%$ quantile curve is straightforward. If, however, the distribution is skew, as is more usual, then often a transformation to normality is applied. A typical transformation is the Box–Cox transformation to which we shall return in Section 3.2; see Cole (1988), Altman (1990) and Royston and Wright (2000).

For cases when it is not possible to make a transformation to normality, a nonparametric or semiparametric approach has been developed; see Cole and Green (1992) and Heagerty and Pepe (1999).

2.2. Applications to survival analysis

Applications to survival analysis include studying the effect of a specific covariate on the survival time of an individual. A given covariate may have a different effect on low, medium and high risk individuals. These effects can be understood by considering several quantile functions of survival time; see Koenker and Geling (2001) for details. Fig. 4 presents three quantile regression curves with $p = 0.1, 0.5, 0.9$ based on the 184 survival times of patients with covariate age between 12 and 64 years from the Stanford heart transplant survey (Crowley and Hu, 1977); see Yang (1999) for further details about censored median regression.

Cox's proportional hazard model is often used for survival analysis. Alternatively, the accelerated failure time approach that models the logarithm of the survival time as a function of covariates can be employed (Yang, 1999; Kottas and Gelfand, 2001). This is intuitively appealing, owing to its ease of interpretation.

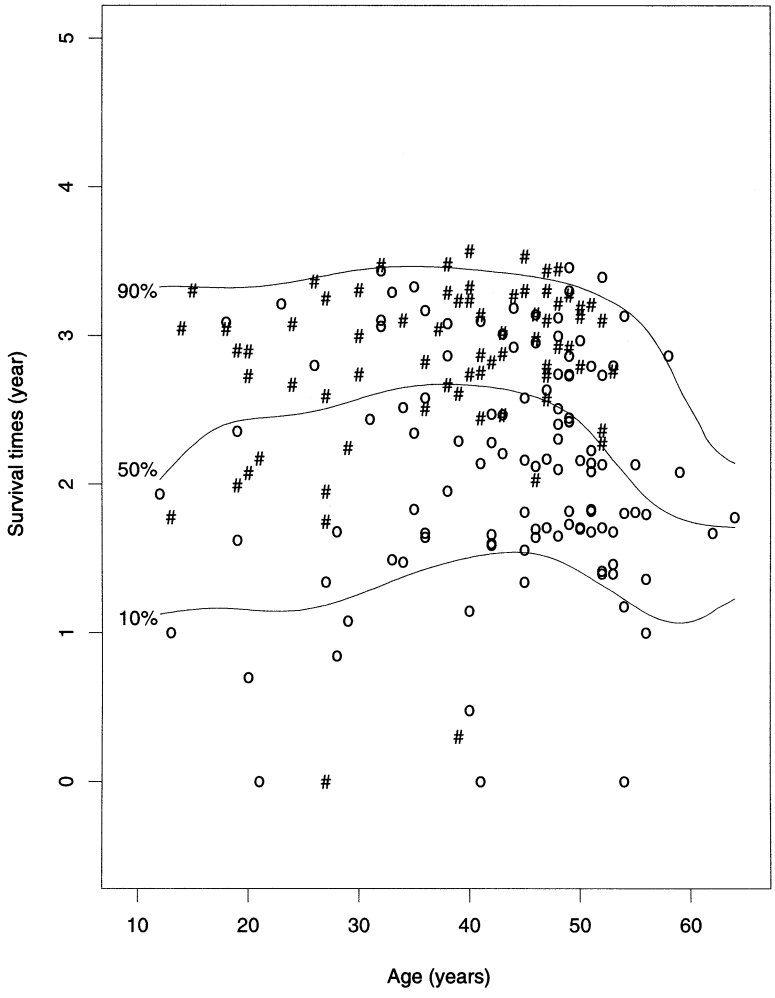

Fig. 4. Stanford heart transplant data with three quantile curves corresponding to $p = 0.1, 0.5, 0.9$: O, complete data; #, censored data

The basic model posits survival times T_i, $i = 1, \ldots, n$, that may be censored and that depend on covariates $\mathbf{x}_i$. In the absence of censoring, it is natural to consider the pairs $\{T_i, \mathbf{x}_i\}_{i=1}^n$ as a multivariate independently and identically distributed sample. If the ith observation has been censored, then we observe Y_i for T_i. The 'log'-transformation of T_i provides the usual accelerated failure time model, which regresses the logarithm of T_i linearly on $\mathbf{x}_i$, i.e.

$$\log(T_i) = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i,$$

where ε_i, $i = 1, \ldots, n$, are independently and identically distributed with an unknown distribution function. The mean of ε_i is not assumed to be 0 because we observe Y_i instead of T_i in the case of censoring and so the intercept term is not included in the vector $\boldsymbol{\beta}$. Because of this, mean regression analysis is not a good estimation technique for the accelerated failure time approach. However, the quantile regression technique that models the quantiles of the survival time, or a

monotone transform thereof, as a function of the covariates and the intercept is appropriate, as Yang (1999) illustrated.

2.3. Applications in financial and economics research

2.3.1. Value at risk, tails of distributions and quantiles

Financial regulations usually require banks to report their daily risk measures called value at risk (VAR). VAR models are the most commonly used measure of market risk in the financial industry (Lauridsen, 2000). Let Y be the financial return, so that the y satisfying $P(Y \leqslant y) = p$ for a given low value of p is the VAR. The variable Y may depend on covariates $\mathbf{x}$ such as exchange rates.

Clearly, VAR estimation relates to extreme quantile estimation through estimating the tail of financial return. The distribution of financial return could also be illustrated by several quantiles. For example, the common approach to estimating the distribution of one-period return in financial models is to forecast the volatility and then to make a Gaussian assumption (see Hull and White (1998)). Market returns, however, are frequently found to have more kurtosis than a normal distribution. Using data on 950 daily observations of the exchange rates for the British pound, German Deutschmark and Japanese yen quoted against the US dollar between June 25th, 1990, and March 23rd, 1994, Taylor (1999) employed a quantile regression approach for the distribution of multiperiod returns. A general discussion of using quantile regression for return-based analysis was given by Bassett and Chen (2001).

2.3.2. Economics

Quantile regression is useful in the study of consumptive markets as the influence of a covariate may be very different for individuals who belong to high, medium and low consumption groups. Similarly, changes in interest rates may have a different inference on the share prices of companies which belong to high, medium and low profits groups.

In particular, quantile regression is now regarded as a standard analysis tool for wage and income studies in labour economics; see, for example, Buchinsky (1995). It is also important to study how incomes are distributed among the members of a population, e.g. to determine tax strategies or for implementing social policies.

Other applications include modelling household electricity demand over time in terms of weather characteristics. The low quantile curves correspond to background use, whereas possibly the high quantile curves reflect high use during active periods of the day particularly due to air conditioning; see Hendricks and Koenker (1992).

2.4. Applications to environment modelling

Hydrology is concerned with modelling rainfall and river flow. The provision of fresh water requires an assessment of the chances of drought to help with the design of reservoirs, and the chances of high rainfalls to help with the design of flood drains and run-offs. The modelling of the tails of distributions and the knowledge of extreme quantiles are thus central to the statistics of hydrology. Suppose that $q_p(\mathbf{x})$ is the quantile function for a hydrological variable, such as annual maximum flood height; then, for some given high p_0, say 0.9, there is a $100(1 - p_0)\%$ chance of an exceedance beyond $q_{p_0}(\mathbf{x})$. The probability that the first exceedance occurs in year k is $P(K = k) = p_0^{k-1}(1 - p_0)$. The mean value of variable K is $1/(1 - p_0)$, which is called the return period T for a $100(1 - p_0)\%$ exceedance event. For example, if T is 100 years, then $q_{0.99}(\mathbf{x})$ is the value of a 99% chance of the 100-year flood exceedance.

In studying pollution data, models for mean concentration levels may be less relevant from a public health standpoint than comparable models for upper quantiles representing more extreme concentration levels; see, for example, Pandey and Nguyen (1999) and Hendricks and Koenker (1992) for discussion.

2.5. Applications to detecting heteroscedasticity

Recognizing heteroscedasticity is an important task for the data analyst. Quantile plots can provide a useful descriptive tool. These plots not only help to detect heteroscedasticity but also provide an impression of the location, spread and shape of the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$.

Quantile regression can be used to assess departures from the assumptions of the model $Y = \mathbf{x}^{\mathrm{T}}\beta + \varepsilon$. If the distribution of ε does not depend on the value of the covariate $\mathbf{X}$, all regression quantiles will be parallel. For example, the seven quantile curves for the US girls data in Fig. 3 are clearly not parallel, indicating heteroscedasticity.

3. Estimation methods and algorithms

We now discuss estimation methods and algorithms for quantile regression.

3.1. The parametric quantile regression model

To quantify the relationship between a response variable Y and covariates $\mathbf{x}$, we often assume that $E[Y|\mathbf{X} = \mathbf{x}]$ can be modelled by a simple linear combination $\mathbf{x}^{\mathrm{T}}\beta$. Similarly, the basic quantile regression model specifies the linear dependence of the conditional quantiles of Y on $\mathbf{x}$. In other words, we assume that the relationship between the $100p\%$ quantile of Y and covariates $\mathbf{x}$ is given by $q_p(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\beta$.

Given a data set $\{\mathbf{x}_i, y_i\}_{i=1}^n$, we can estimate the parameters β by minimizing

$$\sum_{i=1}^{n} \rho_p(y_i - \mathbf{x}_i^{\mathrm{T}}\beta).$$

There is no explicit solution for the regression coefficients under this parametric quantile regression model since the check function is not differentiable at the origin. However, using recent advances in interior point methods for solving linear programming problems discussed by Portnoy and Koenker (1997), this minimization can be performed by using the algorithm that was provided by Koenker and D'Orey (1987). The special case of median regression computation is available in the S-PLUS package, and a suite of functions for general parametric quantile regression in S-PLUS is available from `http://lib.stat.cmu.edu/`.

An 'interior' algorithm for general quantile regression fitting has been developed by Koenker and Park (1996) and S-PLUS functions for this algorithm are available from `http://www.econ.uiuc.edu/roger/research/rq/rq.html`.

3.2. The Box–Cox transformation quantile model

We mentioned parametric quantile regression estimation through the Box–Cox transformation in Section 2.1. The technical details related to this method are as follows: define a function $g(\lambda; y)$ as

$$g(\lambda; y) = \begin{cases} (y^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0, \\ \log(y) & \text{if } \lambda = 0. \end{cases}$$

Let $g^{-1}(\lambda; z)$ be the inverse function of $g(\lambda; y)$ with respect to y, so that $g^{-1}(\lambda; z) = (\lambda z + 1)^{1/\lambda}$ if $\lambda \neq 0$, and $\exp(z)$ if $\lambda = 0$. Then, if the $100p\%$ quantile of $g(\lambda; Y)$ is $q_p(g)$, the $100p\%$ quantile of Y is $g^{-1}\{\lambda; q_p(g)\}$. In particular, if there is a value of λ such that $g(\lambda; Y)$ and X are normally distributed, then properties of the bivariate normal distribution tell us that the $100p\%$ quantile of $g(\lambda; Y)$ takes the linear form $\beta_0 + \beta_1 x$, where β_0 and β_1 depend on p. Hence the $100p\%$ quantile of Y is given by $g^{-1}(\lambda; \beta_0 + \beta_1 x)$. The packages Statgraphics and S-PLUS for example allow Box–Cox transformations to be performed quickly.

Of course, there may be no value of λ that transforms the variable Y to normality. In that case, we can use a transformation based on the check function to find $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}}$ and λ by minimizing

$$\sum_{i=1}^{n} \rho_p\{y_i - g^{-1}(\lambda; \beta_0 + \beta_1 x_i)\};$$

see Buchinsky (1995) for further details.

Under the above transformation framework, we can also assume that the value of λ varies with covariates $\mathbf{x}$, and the $100p\%$ quantile of Y under the Box–Cox normal transformation may then be given by considering $g^{-1}\{\lambda(x); \beta_0 + \beta_1 x\}$. So instead of estimating a single value of λ we may estimate the curve $\lambda(x)$. In that case, we need smoothing estimation based on a kernel method or smoothing spline. Cole and Green (1992) provided an algorithm for penalized likelihood estimation of $\lambda(x)$. They derived the estimation of $\lambda(x)$ based on

$$l(\lambda) - \alpha \int \lambda''(x)^2 \, \mathrm{d}x,$$

where the log-likelihood function

$$l(\lambda) = \sum_{i=1}^{n} \{\lambda(x_i) \log(y_i) - \tfrac{1}{2} z_i^2\},$$

$z_i = g(\lambda; y_i)$, $\int \lambda''(x)^2 \, \mathrm{d}x$ is a roughness penalty and α is a smoothing parameter. Fortran code implementing this algorithm is available from Professor Cole.

3.3. *The nonparametric quantile regression model*

We have seen that the conditional distribution is a vital ingredient for quantile regression. Yu and Jones (1998) and Hall *et al.* (1999) have recently considered methods for estimating conditional distributions. For example, $F(y|x)$ may be estimated nonparametrically by

$$\hat{F}(y|x) = \sum_{i=1}^{n} w_i(x) \, I(y_i \leqslant y),$$

where w_i is a non-zero weight function depending on x_i and x and satisfying $\sum_{i=1}^{n} w_i(x) = 1$.

In Fig. 5 we show the associated nonparametric estimate of $q_{0.5}$ for the motor-cycle data set that was discussed by Silverman (1985).

However, some problems are associated with this type of kernel weighting estimation of the conditional distribution. First, this estimator is not a distribution function, since it is not monotone nor does it take values only between 0 and 1. For this reason, Hall *et al.* (1999) discussed a so-called adjusted Nadaraya–Watson estimator of the conditional distribution function which

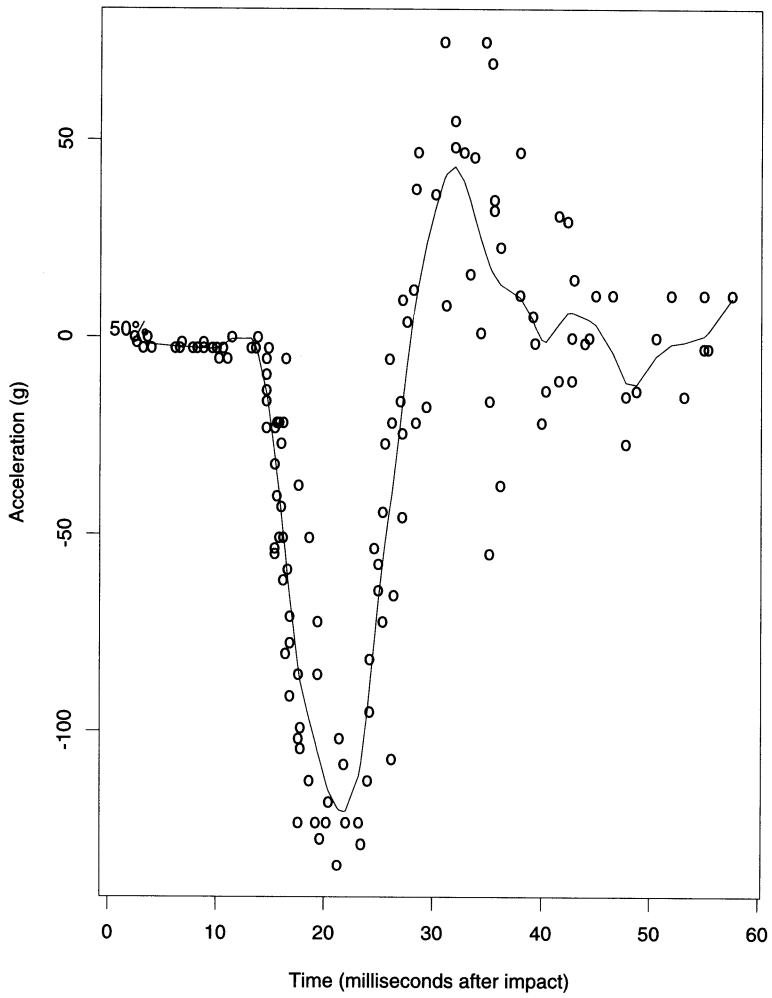

Fig. 5. Motor-cycle data set discussed by Silverman together with a nonparametric estimate of $q_{0.5}$ as a function of time

takes the form

$$\hat{F}(y|x) = \frac{\sum_{i=1}^{n} p_i(x)\, K_h(x_i - x)\, I(y_i \leqslant y)}{\sum_{i=1}^{n} p_i(x)\, K_h(x_i - x)},$$

where $K_h(x_i - x) = K\{(x_i - x)/h\}$ with bandwidth h and kernel density function K, and $p_i(x) \geqslant 0$, $\Sigma_{i=1}^{n} p_i(x) = 1$. Usually p_is satisfying these conditions are not unique.

A second problem is that quantile curves based on these estimators of $F(y|x)$ may cross one another, which is of course absurd. This is illustrated by Fig. 6, which shows kernel smoothing quantile regression estimation based on the $\hat{F}(y|x)$ of Hall *et al.* (1999) for the Stanford heart transplant data set that was discussed in Section 1.

To solve this problem, Yu and Jones (1998) used a 'double-kernel' approach. In this case, a bandwidth is needed for y as well as for x. The basic idea is to replace the indicator function

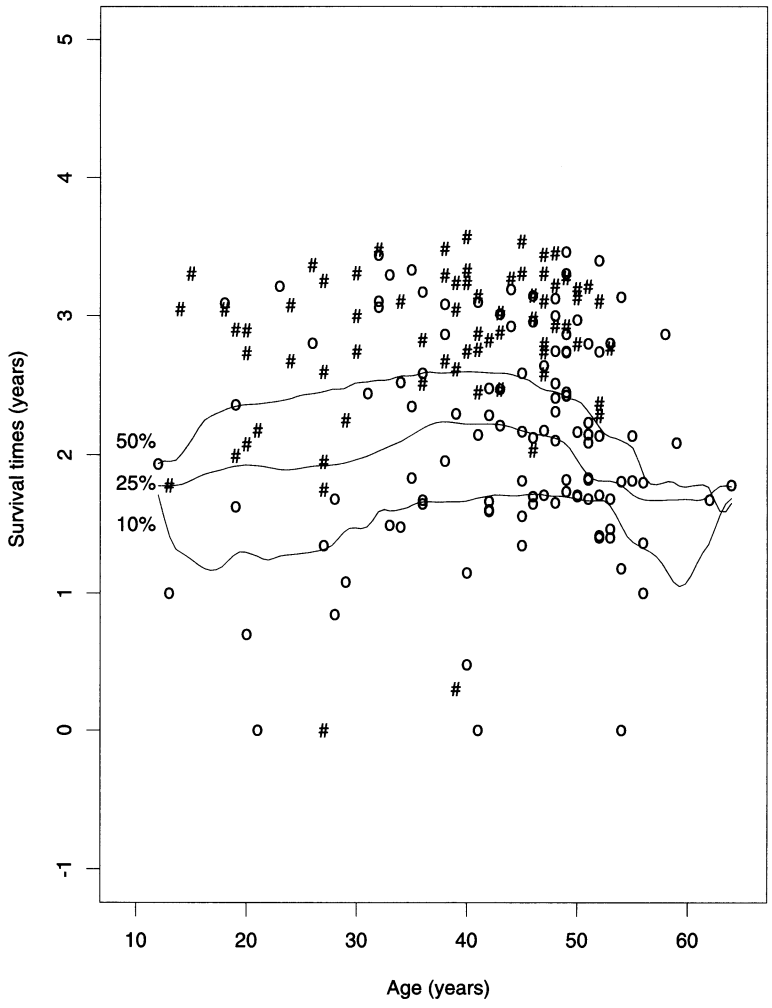

Fig. 6. Stanford heart transplant data with three quantile regression curves with $p = 0.1, 0.25, 0.5$, estimated by using the kernel smoothing technique: note that these curves cross

$I(A)$ in $\hat{F}(y|x)$ above by a continuous distribution function $\Omega\{(y_i - y)/b\}$: thus,

$$\hat{F}(y|x) = \sum_{i=1}^{n} w_i(x)\, \Omega\left(\frac{y_i - y}{b}\right),$$

where $\Omega(q) = \int_{-\infty}^{q} W(v)\, \mathrm{d}v$ is a distribution function with associated density function $W(u)$, $b > 0$ is a bandwidth in the y-direction and $w_i(x)$ are kernel-based weight functions.

If W is taken to be the uniform kernel density, $W(u) = 0.5\, I(|u| \leqslant 1)$, and if the weights

$$w_i(x) = K_h(x_i - x) \Big/ \sum_{i=1}^{n} K_h(x_i - x)$$

with bandwidth h in the x-direction, it can be shown that $\mathrm{d}\hat{q}_p(x)/\mathrm{d}p > 0$. It follows from this that $\hat{q}_p(x)$ is a monotone function of p, with the consequence that this method does not suffer from the crossing problem that is mentioned above. Yu and Jones (1998) discussed an iterative algorithm for estimating quantile functions under the double-kernel approach.

Under the check function set-up, kernel-based quantile regression estimation can be carried out by considering kernel-weighted check function minimization

$$\min_a \left\{ \sum_{i=1}^{n} \rho_p(y_i - a)\, K_h(x_i - x) \right\}, \tag{3}$$

where the minimizer $\hat{a} = \hat{a}(x)$ is the $100p\%$ quantile regression estimator. An iteratively reweighted least squares procedure for finding $\hat{a}$ is given in Yu and Jones (1998), with associated S-PLUS functions available from the first author. This algorithm is iterative, and convergence is guaranteed.

3.4. Bandwidth selection

Associated with the kernel fitting approach that was mentioned in Section 3.3 is the important issue of bandwidth selection. There are several different ways to select the bandwidth h in the x-direction. One method proposed by Ruppert *et al.* (1995) is based on the asymptotic mean-square error together with the 'plug-in' rule to replace any unknown quantity in the asymptotic mean-square error by its estimator. The resulting bandwidth is asymptotically optimal. However, a simpler rule of thumb is appealing especially if it is based on a straightforward approximation of the asymptotic mean-square error, as we always want to estimate several quantile regression functions simultaneously. One rule for selecting the bandwidth in the x-direction simply modifies the bandwidth h_{mean} that would be used for mean regression estimation and can be easily implemented as follows:

(a) obtain h_{mean} by employing the technique proposed by Ruppert *et al.* (1995), for example;
(b) calculate $h_p = C_p h_{\text{mean}}$, where

$$C_p = \left[\frac{p(1-p)}{\phi\{\Phi^{-1}(p)\}^2} \right]^{1/5}.$$

From the plot of C_p against p given in Fig. 7, we immediately deduce that h_p is symmetric about $p = 0.5$ and that, the more extreme the value of p, the more smoothing is required, with the increase becoming large for very large or small values of p.

Similarly, by minimizing the asymptotic mean-square error of the estimator over the bandwidth b_p, we can obtain the following relationship between b_p and h_p:

$$\frac{b_p h_p^3}{b_{1/2} h_{1/2}^3} = \frac{\sqrt{(2\pi)}\, \phi\{\Phi^{-1}(p)\}}{2\{(1-p)\, I(p \geqslant \frac{1}{2}) + p\, I(p < \frac{1}{2})\}};$$

see Yu and Jones (1998) for further details.

3.5. The semiparametric quantile regression model

The penalized likelihood estimation approach that was described in Section 3.2 is an example of a semiparametric fitting technique, as discussed by Cole and Green (1992). More generally, the response Y is assumed to depend in a parametric (linear) fashion on some, but not all, of the covariates. Such a model could be specified by replacing the linear model $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ by $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + g(t)$:

$$q_p = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + g(t)$$

where $\mathbf{x}$ and t are covariates for the $100p\%$ quantile of Y, and $\boldsymbol{\beta}$ and the function g are to be estimated. Such a situation may arise when a model is held to be appropriate on grounds of

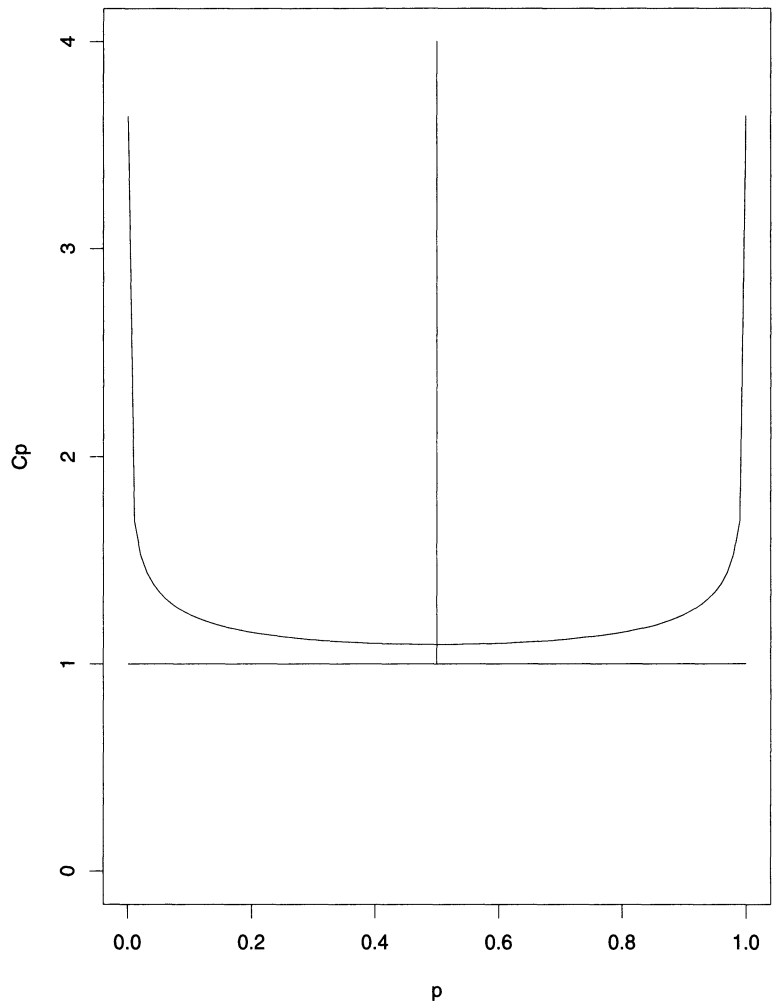

Fig. 7. Relationship between C_p and p: the function is symmetric about the vertical line $p = \frac{1}{2}$

theory or experience, but there are doubts about the homogeneity of the distribution of response over time or space. We usually estimate β and g by minimizing

$$\sum_{i=1}^{n} \rho_p\{Y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - g(t)\} + \alpha \int g''(t)^2 \, \mathrm{d}t.$$

Koenker *et al.* (1992) developed an algorithm to perform this minimization, and some associated S-PLUS functions are available from `http://lib.stat.cmu.edu/`.

3.6. The two-step method

The procedure that minimizes the check function is an iterative algorithm that sometimes takes a long time to converge. Yu (1999) presented a two-step method which avoids the minimization of the check function, but which has the same asymptotic property as the kernel estimators of Section 3.3. This two-step method proceeds by first producing a set of quantiles by using the k nearest neighbour approach of Bhattacharya and Gangopadhyay (1990) at each covariate

point. Then simple least square kernel smoothing is applied to these quantiles to obtain the final quantile regression estimators.

4. Recent research areas

We now briefly discuss some recent research areas.

4.1. Quantile regression for time series

Most research in quantile regression has assumed that the observations of the response variable Y are conditionally independent. Recently, several researchers have discussed different methods for time series quantile regression modelling. For example, a method based on estimating the conditional distribution is given by Cai (2002), whereas a method based on the check function is given by Gannoun *et al.* (2003). In the method of Cai (2002), the time series Y_i is assumed to be related to the time series X_i through the model $Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i$, where $\mu(X_i)$ is the regression function and ε_i is the model error. The dependence of $\sigma(X_i)$ on X_i means that the model is heteroscedastic. The method first estimates the conditional distribution of Y_i given X_i by the approach of Section 3.3 and then estimates the condition quantile by the inverse of the conditional distribution function. In the method of Gannoun *et al.* (2003) for the estimation of the conditional quantile of a strictly stationary real-valued process Z given the present and past records, the $100p\%$ quantile of Z is characterized as

$$q_p(\mathbf{x}) = \arg\min_{\theta \in R}\{E[\rho_p(Z-\theta)|\mathbf{X}=\mathbf{x}]\}.$$

Then, they used the kernel-based check function procedure of Section 3.3 for estimation. Quantile regression has also been applied to longitudinal data analysis in a similar way; see Lipsitz *et al.* (1997), who described an efficient quantile regression approach for CD4 cell count data by taking account of drop-outs.

4.2. Goodness of fit

Assessing the fit of quantile curves to data appears to have received little attention in the literature, with Koenker and Machado (1999) and Royston and Wright (2000) being two important exceptions. Typical goodness-of-fit issues here are assessing the performance of different polynomial models and comparing a parametric model against a nonparametric model. In the former case, for a parametric model $q_p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, we want to test whether covariate x_1 has any effect on the $100p\%$ quantile values of a response. Essentially, we require a statistical hypothesis test of this parametric model against the simpler parametric model $q_p(\mathbf{x}) = \beta_0 + \beta_2 x_2$. Koenker and Machado (1999) addressed this issue.

We shall motivate our discussion of the second issue by using the immunoglobulin G data as an example. This data set comprises the serum concentration (grams per litre) of immunoglobulin G in 298 children aged from 6 months to 6 years and is discussed in detail by Isaacs *et al.* (1983). We take the response variable Y to be the concentration of immunoglobulin G and use a quadratic model in age x to fit the quantile regression:

$$q_p(x) = \beta_0 + \beta_1 x + \beta_2 x^2,$$

for various values of p. We used the parametric method of Section 3.1 to fit the model. Fig. 8(a) shows the plot of the data along with the quantile regression lines for $p = 0.05, 0.25, 0.5, 0.75, 0.95$. All the curves except the 75% curve show a clear quadratic form.

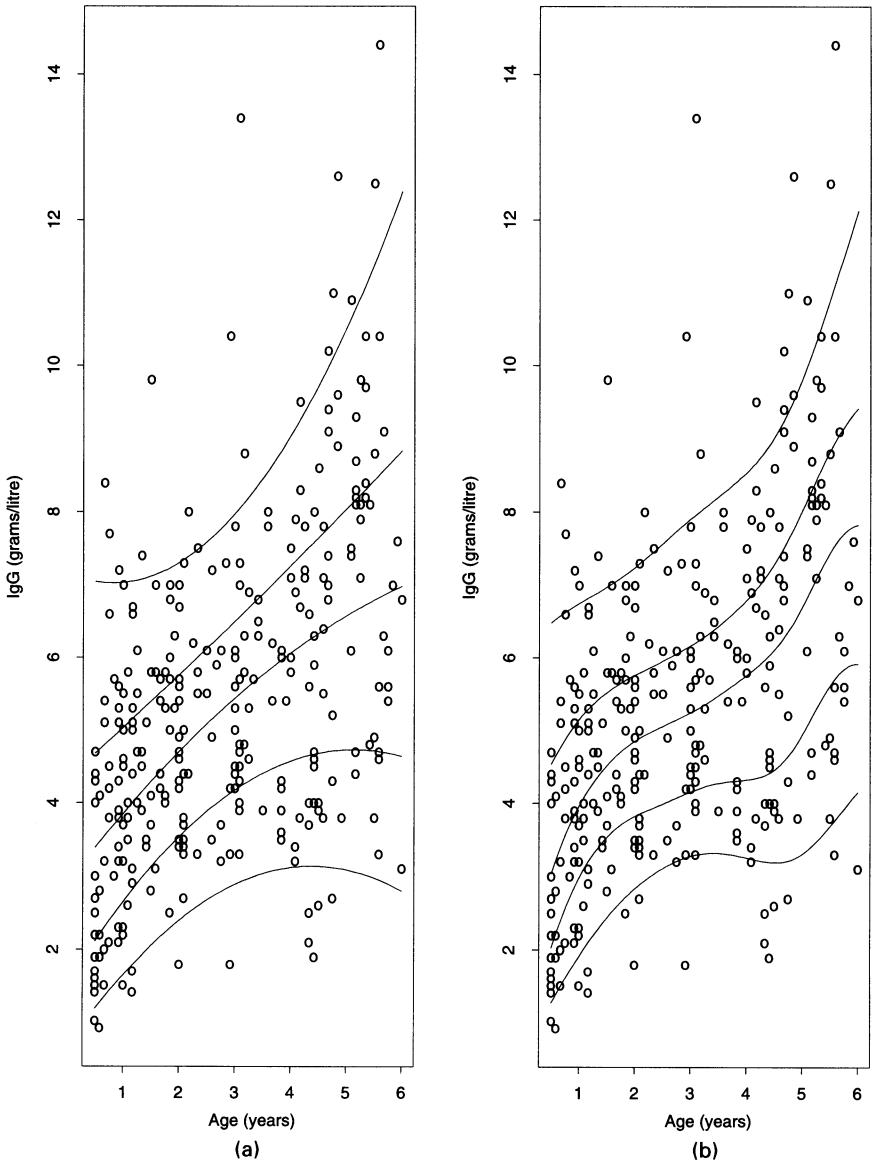

Fig. 8. Immunoglobulin G data with five quantile regression curves with $p = 0.05, 0.25, 0.5, 0.75, 0.95$ fitted by using (a) a quadratic model and (b) a nonparametric model

As an alternative, the kernel-based nonparametric smoothing method described in Section 3.3 estimates the underlying quantile curves through expression (3). The estimated quantile curves are illustrated in Fig. 8(b).

By comparing these different parametric and nonparametric quantile regression curves for the immunoglobulin G data, we see that the problem of assessing the goodness of fit of a quantile regression relationship is very important. Although there is much research on assessing the goodness of fit of mean regression models, this work is not really suitable for quantile regression model fitting since the definitions and estimating methods of the quantile regression model are different from those of the mean regression model. Almost all mean regression testing statistics

are based on a quadratic form such as the R^2-statistic, or on a residual sum of squares, which corresponds to the quadratic loss function.

In general, we need to be able to assess whether a nonparametric quantile regression model can be simplified to a linear or a parametric quantile regression model. This issue is the subject of current research by the authors. We also need to be able to assess the additivity of a quantile regression model with a high dimensional predictor. The additive structure is important in terms of interpretability.

4.3. Bayesian quantile regression

The use of Bayesian inference for generalized linear models is quite standard these days; see Besag and Higdon (1999). Two obvious advantages of the Bayesian approach over classical inference are that the Bayesian approach may lead to exact inference as opposed to the asymptotic inference of the classical approach, and that Bayesian inference deals in a better way with parameter uncertainty. The relative ease with which Markov chain Monte Carlo methods (see Chen *et al.* (2001)) may be used for obtaining the posterior distribution, even in complex situations, has made Bayesian inference much more accessible and attractive. Markov chain Monte Carlo methods make available the entire posterior distribution of parameters β of interest.

Recently, Yu and Moyeed (2001) developed a fully Bayesian modelling approach for quantile regression by employing the asymmetric Laplace likelihood function, as discussed in Section 1.5. Their method is as follows.

Given the observations $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, the posterior density $\pi(\beta|\mathbf{y})$ of β given $\mathbf{y}$, $\pi(\beta|\mathbf{y})$, is provided by

$$\pi(\beta|\mathbf{y}) \propto L(\mathbf{y}|\beta)\,\pi(\beta),$$

where $\pi(\beta)$ is the prior distribution of β and $L(\mathbf{y}|\beta)$ is the likelihood function based on the asymmetric Laplace density $\rho_p(\cdot)$ of Section 1.5. In their method, Yu and Moyeed (2001) showed that choosing an improper uniform prior for β results in a proper joint posterior distribution, i.e., if $\pi(\beta) \propto 1$, then the posterior density $\pi(\beta|\mathbf{y})$ will be proper. Other distributions could be used in the presence of more substantial prior information.

Kottas and Gelfand (2001) proposed an alternative approach that employs a mixture model for errors in a median regression model. Their likelihood is based on a parametric family of skewed distributions that they introduced.

5. Conclusions

Quantile regression is emerging as a comprehensive approach to the statistical analysis of linear and non-linear response models, partly because classical linear theory is essentially a theory just for models of conditional expectations. In this paper we have presented a general review of quantile regression models and associated algorithms. We have illustrated that quantile regression has strong links to three very useful statistical concepts: regression, robustness and extreme value theory. We have also demonstrated that quantile regression is widely used in many important application areas, such as medicine and survival analysis, financial and economic statistics and environmental modelling.

As we believe that quantile regression has a bright future, we have outlined some current research areas, making suggestions for further work. Whereas parametric, semiparametric and nonparametric quantile regression models have been reported for many years, Bayesian quantile

regression and additive quantile regression have attracted much recent attention. Some research or potential research areas have not been described in this review. For example, the estimation and application of quantile regression derivative functions is discussed by Chaudhuri *et al.* (1997). Bivariate and high dimensional nonparametric fitting algorithms for quantile regression models are studied by He *et al.* (1998). A reversible jump Markov chain Monte Carlo algorithm for fitting quantile regression is derived by Yu (2002). Extending the quantile regression approach to spatial and random-effects models is another area of interest and has briefly been discussed by Yu and Moyeed (2001).

Acknowledgements

The authors thank the Joint Editor and referees for their valuable comments and constructive suggestions. This work was partially supported by the National Nature Science Foundation China (grants 70271003 and 79930900).

References

Altman, N. S. (1990) Kernel smoothing of data with correlated errors. *J. Am. Statist. Ass.*, **85**, 749–759.
Bailar, B. (1991) Salary survey of U.S. colleges and universities offering degrees in statistics. *Amstat News*, **182**, 3–10.
Bassett, G. W. and Chen, H. (2001) Portfolio style: return-based attribution using quantile regression. *Emp. Econ.*, **26**, 293–305.
Besag, J. and Higdon, D. (1999) Bayesian analysis of agricultural field experiments (with discussion). *J. R. Statist. Soc.* B, **61**, 691–746.
Bhattacharya, P. K. and Gangopadhyay, A. K. (1990) Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.*, **18**, 1400–1415.
Buchinsky, M. (1995) Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987. *J. Econometr.*, **65**, 109–154.
Cai, Z. (2002) Regression quantiles for time series. *Econometr. Theory*, **18**, 169–192.
Chaudhuri, P., Doksum, K. and Samarov, A. (1997) On average derivative quantile regression. *Ann. Statist.*, **25**, 715–744.
Chen, M.-H., Shao, Q.-M. and Ibrahim, J. G. (2001) *Monte Carlo Methods in Bayesian Computation*. New York: Springer.
Cole, T. J. (1988) Fitting smoothed centile curves to reference data (with discussion). *J. R. Statist. Soc.* A, **151**, 385–418.
Cole, T. J. and Green, P. J. (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. *Statist. Med.*, **11**, 1305–1319.
Crowley, J. and Hu, M. (1977) Covariance analysis of heart transplant data. *J. Am. Statist. Ass.*, **72**, 27–36.
Gannoun, A., Saracco, J. and Yu, K. (2003) Nonparametric prediction by conditional median and quantiles. *J. Statist. Planng Inf.*, to be published.
Hall, P., Wolff, R. C. L. and Yao, Q. (1999) Methods for estimating a conditional distribution. *J. Am. Statist. Ass.*, **94**, 154–163.
He, X., Ng, P. and Portnoy, S. (1998) Bivariate quantile smoothing splines. *J. R. Statist. Soc.* B, **60**, 537–550.
Heagerty, P. J. and Pepe, M. S. (1999) Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Appl. Statist.*, **48**, 533–551.
Hendricks, W. and Koenker, R. (1992) Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Am. Statist. Ass.*, **93**, 58–68.
Hull, J. and White, A. (1998) Value at risk when daily changes in market variables are not normally distributed. *J. Deriv.*, **5**, 9–19.
Isaacs, D., Altman, D. G., Tidmarsh, C. E., Valman, H. B. and Webster, A. D. B. (1983) Serum Immunoglobin concentrations in preschool children measured by laser nephelometry: reference ranges for IgG, IgA and IgM. *J. Clin. Path.*, **36**, 1193–1196.
Koenker, R. W. and D'Orey, V. (1987) Algorithm AS 229: Computing regression quantiles. *Appl. Statist.*, **36**, 383–393.
Koenker, R. and Geling, R. (2001) Reappraising medfly longevity: a quantile regression survival analysis. *J. Am. Statist. Ass.*, **96**, 458–468.
Koenker, R. and Machado, J. (1999) Goodness of fit and related inference processes for quantile regression. *J. Am. Statist. Ass.*, **94**, 1296–1309.

Koenker, R. and Park, B. J. (1996) An interior point algorithm for nonlinear quantile regression. *J. Econometr.*, **71**, 265–283.

Koenker, R., Portnoy, S. and Ng, P. (1992) *Nonparametric Estimation of Conditional Quantile Functions: L_1 Statistical Analysis and Related Methods* (ed. Y. Dodge), pp. 217–229. Amsterdam: Elsevier.

Kottas, A. and Gelfand, A. E. (2001) Bayesian semiparametric median regression model. *J. Am. Statist. Ass.*, **96**, 1458–1468.

Lauridsen, S. (2000) Estimation of value of risk by extreme value methods. *Extremes*, **3**, 107–144.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. and Zhao, L. P. (1997) Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Appl. Statist.*, **46**, 463–476.

Pandey, G. R. and Nguyen, V. T. (1999) A comparative study of regression based methods in regional flood frequency analysis. *J. Hydrol.*, **225**, 92–101.

Portnoy, S. and Koenker, R. (1997) The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators (with discussion). *Statist. Sci.*, **12**, 279–300.

Royston, P. and Altman, D. G. (1994) Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). *Appl. Statist.*, **43**, 429–467.

Royston, P. and Wright, E. M. (2000) Goodness-of-fit statistics for age-specific reference intervals. *Statist. Med.*, **19**, 2943–2962.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995) An effective bandwidth selector for local least squares regression. *J. Am. Statist. Ass.*, **90**, 1257–1270.

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

Taylor, J. (1999) A quantile regression approach to estimating the distribution of multiperiod returns. *J. Deriv.*, **24**, 64–78.

Yang, S. (1999) Censored median regression using weighted empirical survival and hazard functions. *J. Am. Statist. Ass.*, **94**, 137–145.

Yu, K. (1999) Smoothing regression quantile by combining k-NN with local linear fitting. *Statist. Sin.*, **9**, 759–771.

Yu, K. (2002) Quantile regression using RJMCMC algorithm. *Comput. Statist. Data Anal.*, **40**, 303–315.

Yu, K. and Jones, M. C. (1998) Local linear regression quantile estimation. *J. Am. Statist. Ass.*, **93**, 228–238.

Yu, K. and Moyeed, R. A. (2001) Bayesian quantile regression. *Statist. Probab. Lett.*, **54**, 437–447.