Published in final edited form as:
Biometrics. 2011 March ; 67(1): 242–249. doi:10.1111/j.1541-0420.2010.01436.x.

Bent Line Quantile Regression with Application to an Allometric Study of Land Mammals' Speed and Mass

Chenxi Li[1,*], **Ying Wei**[2], **Rick Chappell**[3,1], and **Xuming He**[4]

[1]Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.

[2]Department of Biostatistics, Columbia University, New York, New York 10032, U.S.A.

[3]Departments of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin 53792, U.S.A.

[4]Departments of Statistics, University of Illinois, Champaign, Illinois 61820, U.S.A.

Summary

Quantile regression, which models the conditional quantiles of the response variable given covariates, usually assumes a linear model. However, this kind of linearity is often unrealistic in real life. One situation where linear quantile regression is not appropriate is when the response variable is piecewise linear but still continuous in covariates. To analyze such data, we propose a bent line quantile regression model. We derive its parameter estimates, prove that they are asymptotically valid given the existence of a change-point, and discuss several methods for testing the existence of a change-point in bent line quantile regression together with a power comparison by simulation. An example of land mammal maximal running speeds is given to illustrate an application of bent line quantile regression in which this model is theoretically justified and its parameters are of direct biological interests.

Keywords

Bahadur representation; Bootstrap; Change-point; Piecewise linear; Profile estimation

1. Introduction

Quantile regression, introduced by Koenker and Bassett (1978), models the conditional quantiles of the response variable Y given a set of covariates $\mathbf{X}$. Compared with ordinary least squares regression, quantile regression can provide a more complete picture of the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$, and it is particularly useful when upper or lower or all quantiles are of interest. It is more flexible for modeling data with heterogeneous conditional distributions and more robust to outliers in Y than least squares regression.

The usual linear regression methods including ordinary least squares and quantile regression assume a linear model between Y and $\mathbf{X}$. The linearity assumption puts some limitations on the pattern of data. The maximal running speed (MRS) data of land mammals in Garland (1983) are an example where the simple linear (quantile) regression is not appropriate. A scatter plot (Figure 2) shows that the natural logarithm of MRS of land mammals increases

* chenxili@stat.wisc.edu.

Web Appendix A, B, Web Figure 1 and Web Table 1 referenced in Sections 2.2.1, 2.2.2 and 4.1 are available under the Paper Information link at the *Biometrics* website http://www.biometrics.tibs.org.

with the natural logarithm of mass up to a certain point and then goes down as the mass rises. It implies that there may be linear relationships between MRS and mass on the log scale which are continuous but have slopes that differ on either side of a boundary called a change-point. Biologic considerations support the idea of a change-point: a quadratic model is unrealistic here because there is no reason to believe that the slope for smaller animals is related to that for larger ones—see Section 5.3 and Chappell (1989). To accommodate data in which a linear relationship is not constant, it is necessary to consider segmented regressions which have different analytical forms in different regions of the domain. A lot of work about parameter estimation, hypothesis testing, and asymptotic distribution has been done for segmented least squares regression (see Robinson, 1964; Quandt, 1958, 1960; Feder, 1975). However, there has not been much analogous work done for segmented quantile regression. When the upper or lower quantiles of the response variable are of interest, segmented quantile regression may be preferable. Upper quantiles are indeed of particular interest here because biologists are interested in the *potential* running speed as a function of mass. For example, the bottom-most point in Figure 2 (at 4 kg, 1.6 km/h) is the three-toed sloth. This animal, due to its arboreal lifestyle, is extremely slow. That fact is useful in some instances but not for the present estimation of the upper limits of speed given the constraints imposed by size. One can see many other slow outliers in Figure 2, but few fast ones (hopping animals, due to their different form of locomotion, are treated separately). Thus we are interested in the maximum rather than the mean in this instance. For reasons of estimability and numerical stability, we substitute a high quantile such as the 0.8th. In this article, we concentrate on an important special case of segmented quantile regression—bent line quantile regression, with two segments which are different straight lines which intersect at a change-point. We develop a systematic bent line quantile regression procedure, from modeling to asymptotics.

The rest of the article is organized as follows. Section 2 describes a bent line quantile regression model and develops parameter estimates, their asymptotic distribution, and methods for confidence intervals and linear hypothesis testing. Section 3 proposes three tests for the existence of a change-point. Section 4 investigates the performances of the tests and confidence intervals via simulation studies. Section 5 presents data analysis results for the maximal running speed of land mammals example. The last section discusses the advantages and limitations of the procedure developed in this paper and further work.

2. Methods

2.1 Model

Given a probability τ strictly between 0 and 1, we consider a continuous bent line quantile regression model

$$Y_i = \begin{cases} \alpha_1 + \beta_1 X_i + \mathbf{z}_i^\top \gamma + e_i & X_i \le t \\ \alpha_2 + \beta_2 X_i + \mathbf{z}_i^\top \gamma + e_i & X_i \ge t \end{cases} \quad i = 1, \cdots, n, \tag{1}$$

where Y_i is the ith response, X_i is the scalar covariate whose slope changes at the change-point t, $\mathbf{z}_i$ is a q-dimensional vector of linear covariates with constant slopes and e_i is the error term whose τ-th quantile is zero conditional on (X_i, z_i). Here, t is unknown and needs to be estimated. Because of continuity, we express one of the parameters in (1), α_2, as a function of the others: $\alpha_2 = \alpha_1 + (\beta_1 - \beta_2)t$. Let $\boldsymbol{\eta} = (\alpha_1, \beta_1, \beta_2, \boldsymbol{\gamma}^\top)^\top$. From Model (1), it is easy to show that given t the τ-th quantile of Y_i given X_i and $\mathbf{z}_i$ is

$$Q(X_i, z_i;\eta)=\begin{cases} \alpha_1+\beta_1 X_i+z_i^\top\gamma & X_i \le t \\ \alpha_1+\beta_1 t+\beta_2(X_i-t)+z_i^\top\gamma & X_i \ge t \end{cases} \quad i=1,\cdots,n. \tag{2}$$

2.2 Statistical Inference

2.2.1 Estimates of coefficients and change-point—If the change-point t were known, the estimates of coefficients $\boldsymbol{\eta}$ could be formulated as the solution to the optimization problem

$$\min_{\eta\in\mathbb{R}^{3+q}} \sum_{i=1}^{n} \rho_\tau(Y_i - Q(X_i, z_i;\eta)), \tag{3}$$

where $\rho_\tau(w) = w(\tau - I(w < 0))$, $0 < \tau < 1$. Here $I(\cdot)$ denotes the indicator function. We may now, via an extension of a method due to Quandt (1958) and Chappell (1989), find estimates by conditioning on t and then minimizing (3) over all values of t in the range of X_i's. Letting $u_i = \min(X_i, t)$ and $v_i = \max(0, X_i - t)$, the quantile conditional on t is linear in form, i.e.,

$$Q_i=Q(X_i, z_i;\eta)=\alpha_1+\beta_1 u_i+\beta_2 v_i+z_i^\top\boldsymbol{\gamma}, \quad i=1,\cdots,n \tag{4}$$

which is equivalent to Equation (2). Let $w_{i,t}=(1, u_i, v_i, z_i^\top)^\top$. Then, conditional on t, $\boldsymbol{\eta}$ can be estimated by solving

$$\widehat{\eta}=\operatorname{argmin}_{\eta\in\mathbb{R}^{3+q}} \sum_{i=1}^{n} \rho_\tau(Y_i - w_{i,t}^\top\eta). \tag{5}$$

Now we have the estimates of $\boldsymbol{\eta}$ conditional on t. Let

$$S(\widehat{\eta}|t)=\sum_{i=1}^{n} \rho_\tau(Y_i - w_{i,t}^\top\widehat{\eta}). \tag{6}$$

$S(\hat{\boldsymbol{\eta}}|\, t)$ is a function of t. One therefore can obtain the unconditional estimates of $\boldsymbol{\eta}$ by minimizing $S(\hat{\boldsymbol{\eta}}|t)$ over t. The value of t at which the minimum is realized is $\hat{t}$, the estimated change-point. This estimation method is needed as an alternative to the algorithm in Koenker (2005, Section 6.6) for estimating nonlinear regression quantiles, which requires differentiability of quantile functions. The computational aspects of bent line quantile regression are discussed in Web Appendix B.

2.2.2 Asymptotics—We consider a general bent line quantile regression model

$$Y_i=\alpha_\tau+(\beta_{1,\tau} I\{X_i \le t_\tau\}+\beta_{2,\tau} I\{X_i>t_\tau\})\,(X_i - t_\tau)+z_i^\top\gamma_\tau+e_{\tau,i}, \quad i=1,\cdots,n, \tag{7}$$

where Y_i is the ith response, X_i is the covariate whose slope changes at t_τ, $\mathbf{z}_i$ is a q-dimensional vector of linear covariates with constant slopes, and $e_{\tau,i}$ is the error term whose τ-th quantile is zero. We treat both the response and the covariates as random to derive the asymptotic properties of the parameter estimates in Section 2.2.1. The observations $\{(Y_i, X_i, z_i^\top)\}_{i=1}^n$ are assumed to be independent but not necessarily identically distributed.

To simplify the presentation, we denote the conditional quantile function

$$g(w_i;\theta_0(\tau))=\alpha_\tau+(\beta_{1,\tau}I\{X_i \le t_\tau\}+\beta_{2,\tau}I\{X_i>t_\tau\})\,(X_i - t_\tau)+z_i^\top\gamma_\tau,$$

where $w_i=(1, X_i, z_i^\top)^\top$ and $\theta_0(\tau)=(\alpha_\tau,\beta_{1,\tau},\beta_{2,\tau}, t_\tau, \gamma_\tau^\top)^\top$ is the true parameter vector. The estimator of $\boldsymbol{\theta}_0(\tau)$ can be expressed by

$$\widehat{\theta}_n(\tau)= \arg\min_\theta \sum_{i=1}^n \rho_\tau(Y_i - g(w_i;\theta)), \tag{8}$$

where $\rho_\tau(r) = r(\tau - I\{r < 0\})$ is the quantile regression loss function and $\boldsymbol{\theta} = (\alpha, \beta_1, \beta_2, t, \boldsymbol{\gamma}^\top)^\top$. Due to the non-convexity of the objective function, $\hat{\boldsymbol{\theta}}_n(\tau)$ is obtained via profile estimation. Recalling that $\boldsymbol{\eta}= (\alpha, \beta_1, \beta_2, \boldsymbol{\gamma}^\top)^\top$ is the vector of parameters excluding t, we can write our objective function with respect to $\boldsymbol{\eta}$ and t as

$$S_{n,\tau}(\eta, t)=\frac{1}{n}\sum_{i=1}^n(\rho_\tau(Y_i - g(w_i;\eta, t)) - \rho_\tau(Y_i)).$$

A profile estimate of $\boldsymbol{\eta}$ at a fixed t is given by

$$\widehat{\eta}_{n,\tau}(t)= \arg\min_\eta S_{n,\tau}(\eta, t).$$

An estimate of the change-point t_τ is given by

$$\widehat{t}_{n,\tau}= \arg\min_t S_{n,\tau}(\widehat{\eta}_{n,\tau}(t), t).$$

$\hat{\boldsymbol{\theta}}_n(\tau)$ is obtained from $\hat{\boldsymbol{\eta}}_{n,\tau}(\hat{t}_{n,\tau})$ and $\hat{t}_{n,\tau}$.

To derive the asymptotic properties of $\hat{\boldsymbol{\theta}}_n(\tau)$, we introduce the following notation:

$$
\begin{gathered}
f_{\tau,\mathrm{w}_i}\text{—the conditional density of } e_{\tau,i} \text{ given } \mathrm{w}_i,\\
h(\mathrm{w}_i;\theta)=(\mathrm{I}\{X_i \le t\}, X_i\mathrm{I}\{X_i \le t\}, \mathrm{I}\{X_i>t\}, X_i\mathrm{I}\{X_i>t\}, \mathrm{z}_i^\top)^\top,\\
g_1(\mathrm{w}_i;\theta)=\alpha+\beta_1(X_i-t)+\mathrm{z}_i^\top\gamma,\quad g_2(\mathrm{w}_i;\theta)=\alpha+\beta_2(X_i-t)+\mathrm{z}_i^\top\gamma,\\
\tilde{g}\,(\mathrm{w}_i;\theta_0(\tau))=\frac{\partial g_1(\mathrm{w}_i;\theta)}{\partial\theta}|\theta=\theta_0(\tau)\, I\{X_i<=t_\tau\}+\frac{\partial g_2(\mathrm{w}_i;\theta)}{\partial\theta}|\theta=\theta_0(\tau)\, I\{X_i>t_\tau\},\\
C_{n,\tau}=n^{-1}\sum_{i=1}^{n}E[\,h(\mathrm{w}_i;\theta_0(\tau))h(\mathrm{w}_i;\theta_0(\tau))^\top],\\
D_{n,\tau}=n^{-1}\frac{\partial E\sum_{i=1}^{n}\psi_\tau(Y_i-g(\mathrm{w}_i;\theta))h(\mathrm{w}_i;\theta)}{\partial\theta}|\theta=\theta_0(\tau),\\
C_{0,\tau}=\lim_{n\to\infty}C_{n,\tau},\quad\text{and}\quad D_{0,\tau}=\lim_{n\to\infty}D_{n,\tau}.
\end{gathered}
$$

The regularity conditions needed to derive the asymptotic properties are given in Web Appendix A. We now present the main results of the asymptotic properties of $\hat{\boldsymbol{\theta}}_n(\tau)$ in the following theorem but defer the proof to Web Appendix A:

Theorem 1: Under the regularity conditions and the condition that $\beta_{1,\tau} \neq \beta_{2,\tau}$, $\hat{\boldsymbol{\theta}}_n(\tau)$ has the following Bahadur representation:

$$
(\widehat{\theta}_n(\tau)-\theta_0(\tau))=-n^{-1}D_{n,\tau}^{-1}\sum_{i=1}^{n}\psi_\tau(e_{\tau,i})h(\mathrm{w}_i;\theta_0(\tau))+o_p(n^{-1/2}), \tag{9}
$$

where $\psi_\tau(r) = \tau - \mathrm{I}\{r < 0\}$ is the first derivative of $\rho_\tau(\cdot)$. The representation (9) implies that

$$
n^{1/2}(\widehat{\theta}_n(\tau)-\theta_0(\tau)) \to N(\mathbf{0}, \tau(1-\tau)D_{0,\tau}^{-1}C_{0,\tau}D_{0,\tau}^{-\top}). \tag{10}
$$

Remark 1: When $\beta_{1,\tau} = \beta_{2,\tau}$, i.e., the change-point does not exist and the estimation is ill-conditioned.

Remark 2: The optimization (8) is equivalent to solving the estimation equation

$$
\sum\psi_\tau(Y_i-g(\mathrm{w}_i;\theta))h(\mathrm{w}_i;\theta)=0, \tag{11}
$$

This estimation equation provides a nice interpretation of the estimation: the quantile gradient condition holds on both sides of t at $\hat{\boldsymbol{\theta}}_n(\tau)$.

Moreover, letting

$$
\widehat{D}_{n,\tau}=-n^{-1}\sum_{i=1}^{n}\widehat{f}_{\tau,\mathrm{w}_i}(0)h(\mathrm{w}_i;\widehat{\theta}_n(\tau))\,\tilde{g}\,(\mathrm{w}_i;\widehat{\theta}_n(\tau))^\top,\quad \widehat{C}_{n,\tau}=n^{-1}\sum_{i=1}^{n}h(\mathrm{w}_i;\widehat{\theta}_n(\tau))h(\mathrm{w}_i;\widehat{\theta}_n(\tau))^\top \tag{12}
$$

where $\hat{f}_{\tau,\mathbf{w}_i}$ is some consistent estimate of $f_{\tau,\mathbf{w}_i}$, then we have

$$\widehat{D}_{n,\tau} \to D_{0,\tau} \text{ and } \widehat{C}_{n,\tau} \to C_{0,\tau} \text{ in probability.}$$

Therefore, we could estimate the asymptotic variance-covariance matrix by $\hat{D}_{n,\tau}$ and $\hat{C}_{n,\tau}$.

We can easily extend the foregoing argument to consider the asymptotic joint distribution of distinct vectors of bent line quantile regression parameters. For $0 < \tau_1 < \tau_2 < \cdots < \tau_m < 1$, $m \in \mathbb{Z}_+$, we set $\theta_0(\tau_j)=(\alpha_{\tau_j},\beta_{1,\tau_j},\beta_{2,\tau_j}, t_{\tau_j}, \gamma_{\tau_j}^\top)^\top$, $1 \le j \le m$. We assume that Model (7) is true for all the τ_j's, i.e.

$$Y_i=g(\mathrm{w}_i;\theta_0(\tau_j))+e_{\tau_j,i}, \quad i=1,\cdots,n, \quad j=1,\cdots,m. \tag{13}$$

Let $\hat{\boldsymbol{\vartheta}}_n = (\hat{\boldsymbol{\theta}}_n(\tau_1)\top, \hat{\boldsymbol{\theta}}_n(\tau_2)\top,\cdots \hat{\boldsymbol{\theta}}_n(\tau_m)\top)\top$ be the estimated vector of parameters with $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0(\tau_1)\top,\boldsymbol{\theta}_0(\tau_2)\top, \cdots, \boldsymbol{\theta}_0(\tau_m)\top)\top$. To describe the asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$, we introduce three more notations: $H_n = \mathrm{diag}[D_{n,\tau_1}, \cdots, D_{n,\tau_m}]$, $H_0 = \mathrm{diag}[D_{0,\tau_1}, \cdots, D_{0,\tau_m}]$ and

$$J_0(\tau_k,\tau_l)=\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E[h(\mathrm{w}_i,\theta_0(\tau_k))h(\mathrm{w}_i,\theta_0(\tau_l))^\top], \quad k,l=1,\cdots,m.$$

The regularity conditions needed to derive the asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$ are also given in Web Appendix A. The asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$ are presented in the following theorem, and its proof is again deferred to Web Appendix A.

Theorem 2: Under the regularity conditions and the condition that $\beta_{1,\tau_j} \neq \beta_{2,\tau_j}$ for $j = 1, \cdots, m$, $\hat{\boldsymbol{\vartheta}}_n$ has the following Bahadur representation:

$$n^{1/2}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)=-n^{-1/2}H_n^{-1}\sum_{i=1}^{n}\begin{pmatrix} \psi_{\tau_1}(e_{\tau_1,i})h(\mathrm{w}_i,\theta_0(\tau_1)) \\ \vdots \\ \psi_{\tau_m}(e_{\tau_m,i})h(\mathrm{w}_i,\theta_0(\tau_m)) \end{pmatrix}+o_p(1). \tag{14}$$

The representation (14) implies that

$$n^{1/2}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \to N(\mathbf{0}, H_0^{-1}\Lambda_0 H_0^{-\top}), \tag{15}$$

where Λ_0 is a $m(4 + q) \times m(4 + q)$ matrix with klth block

$$\Lambda_0(\tau_k,\tau_l)=(\tau_k \wedge \tau_l - \tau_k\tau_l)J_0(\tau_k,\tau_l), \quad k,l=1,\cdots,m.$$

2.2.3 Confidence intervals for the change-point and slopes—The change-point t and slopes β_1, β_2 in Model (1) are, in general, of great interest and so we construct their confidence intervals. Theorem 1 provides an asymptotic confidence interval for every parameter, which involves density estimation. The density estimation method we use in this

paper is called *The Hendricks-Koenker Sandwich* described in Koenker (2005, Section 3.4.2). Bootstrap techniques are also capable of constructing confidence intervals. We resample n triples of variables with replacement from $\{(Y_i, X_i, \mathbf{z}_i)|i = 1,\cdots,n\}$ and use this bootstrap sample to reestimate t, β_1 and β_2. We repeat this procedure B times so that B estimates of the parameters can be obtained. The confidence intervals for t, β_1 and β_2 are computed using Efron's percentile method. This bootstrap method for confidence intervals is valid because the asymptotic joint distribution of the parameter estimates is normal, as discussed in Section 2.2.2.

Note that when the change-point doesn't exist, estimation and confidence interval construction for bent line quantile regression are ill-conditioned. So we recommend testing for a change-point (Section 3) before estimation and constructing confidence intervals.

2.2.4 General linear hypothesis testing—We consider a general hypothesis testing on the vector $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0(\tau_1)^\top, \boldsymbol{\theta}_0(\tau_2)^\top,\cdots, \boldsymbol{\theta}_0(\tau_m)^\top)^\top$ of the form

$$H_0:\ R\boldsymbol{\vartheta}_0=\mathrm{r}.$$

This kind of hypothesis can be accommodated by the Wald approach since we have the asymptotic normality of $\hat{\boldsymbol{\vartheta}}_n$ as discussed in Section 2.2.2. The test statistic is

$$W_n=n(R\hat{\boldsymbol{\vartheta}}_n-\mathrm{r})^\top[RV_nR^\top]^{-1}(R\hat{\boldsymbol{\vartheta}}_n-\mathrm{r}), \tag{16}$$

where V_n is a $m(4+q)\times m(4+q)$ matrix with klth block

$$V_n(\tau_k,\tau_l)=(\tau_k\wedge\tau_l-\tau_k\tau_l)\hat{D}_{n,\tau_k}^{-1}\hat{J}_n(\tau_k,\tau_l)\hat{D}_{n,\tau_l}^{-\top},$$

where

$$\hat{J}_n(\tau_k,\tau_l)=n^{-1}\sum_{i=1}^{n}h(\mathrm{w}_i,\hat{\theta}_n(\tau_k))h(\mathrm{w}_i,\hat{\theta}_n(\tau_l))^\top.$$

The test statistic W_n is asymptotically χ^2_ν under the null hypothesis, where ν is the rank of the matrix R.

This formulation accommodates a wide variety of testing situations. For example, we could test for the equality of several slope coefficients across several quantiles. Also, we might test for the increasing trend in the slope coefficient as the quantile level rises by constructing an appropriate R for this purpose.

3. Three Tests for the Change-point Existence

One may question whether there is a bend at all. So we are interested in testing simple versus continuous two-phase linear quantile regression.

Three methods of testing the change-point existence are discussed below.

1. **An omnibus lack of fit test.** We apply the lack of fit test in He and Zhu (2003) to test H_0: the simple linear quantile regression model (17) is adequate for the data.

$$Y_i = \mathrm{w}_i^\top \xi + e_i, \quad i = 1, \cdots, n, \tag{17}$$

where $\mathrm{w}_i = (1, X_i, \mathrm{z}_i^\top)^\top$ is a $(2+q)$-dimensional vector of linear covariates, $\xi = (\alpha, \beta, \gamma^\top)^\top$ and e_i is the error term whose τ-th quantile equals zero.

2. **Testing the significance of a quadratic term in the model.** A general test for nonlinearity (bent line in our case) is testing the significance of a quadratic term of X. This can be done by a Wald test as described in Bassett and Koenker (1982). The reason we test for a quadratic term is not because we believe a quadratic model fits the data but because it is an easy test against a wide range of nonlinear alternatives.

3. **A bootstrap-based test using the continuous change-point as an alternative.** We are interested in testing

   H_0: no change-point in (c, d) vs. H_A: there exists a continuous change-point in (c, d), where (c, d) is in the support of the density of the covariate which might have non-constant slope. In other words, we want to test the adequacy of the simple model (17) with Model (1), where $t \in (c, d)$, as the alternative. Conditional on t, one may test whether β_1 and β_2 in (4) are equal with a Wald test as described in Bassett and Koenker (1982), which can test the hypothesis that the nested smaller model is adequate relative to the larger model. Inspired by this fact, we can construct an unconditional test. For the observed Wald test statistic T_{obs} computed conditional on $\hat{t}$, the following bootstrap method would give an approximate level of significance.

First, fit Model (1) and obtain $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\gamma}$, and the residuals $\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_n$. Next generate the null data $(Y_i^0, X_i, \mathbf{z}_i)$ using

$$Y_i^0 = \widehat{\alpha}_1 + \widehat{\beta}_1 X_i + \mathrm{z}_i^\top \widehat{\gamma} + \widehat{e}_i, \quad i = 1, \cdots, n. \tag{18}$$

Then generate bootstrap samples $(Y_i^*, X_i^*, \mathrm{z}_i^*)$ by sampling with replacement from $\{(Y_i^0, X_i, \mathrm{z}_i) | i = 1, \cdots, n\}$. For every bootstrap sample $\{(Y_i^*, X_i^*, \mathrm{z}_i^*) | i = 1, \cdots, n\}$, get the estimate of change-point, $\hat{t}^*$, and conditional on $\hat{t}^*$, compute the Wald test statistic T^* for H_0:$\beta_2 - \beta_1 = \hat{\beta}_2^{0,*} - \hat{\beta}_1^{0,*}$, where $\hat{\beta}_1^{0,*}$ and $\hat{\beta}_2^{0,*}$ are the estimates of β_1 and β_2 in (4) with the null data $\{(Y_i^0, X_i, \mathrm{z}_i) | i = 1, \cdots, n\}$ and $t = \hat{t}^*$. In this way, we obtain the empirical null hypothesis distribution of T^*. The approximate p-value is then the proportion of T^*'s exceeding T_{obs}, because large values of T_{obs} are indicative of (1).

Similar bootstrap-based tests for change-points are discussed by Hinkley et al. (1980, Section 3.4) and Gijbels and Goderniaux (2004) for linear regression and nonparametric regression, respectively.

4. Simulation

4.1 Power Simulation

A power comparison among the three tests proposed in Section 3 is performed through simulation.

The data for simulation are generated from the following two sets of scenarios:

1. Symmetric scenarios:
   - A straight line:

$$y=1+0.5x+v_A(x)e \quad 0 \le x \le 10 \tag{19}$$

   - Symmetric bent lines with β_1 = 0.125, 0.25, 0.375, and 0.5:

$$y=5\beta_1+(\beta_1 \mathbf{I}\{x \le 5\} - \beta_1 \mathbf{I}\{x>5\})(x-5)+v_A(x)e \quad 0 \le x \le 10 \tag{20}$$

2. Asymmetric scenarios:
   - A straight line:

$$y=5-0.5x+v_B(x)e \quad 0 \le x \le 10 \tag{21}$$

   - Asymmetric bent lines with $\beta_1 = -0.25$, 0, 0.25, and 0.5:

$$y=4.5+(\beta_1 \mathbf{I}\{x \le 1\} - 0.5\mathbf{I}\{x>1\})(x-1)+v_B(x)e \quad 0 \le x \le 10 \tag{22}$$

where $v_A(x) = 1 + 0.2x$, $v_B(x) = 0.5 + 0.02x$ and $e \sim N(0, 1)$. In the symmetric scenarios, x is uniformly distributed on [0, 10]. In the asymmetric scenarios, x has a mixture distribution that is uniform on [0, 1] with probability 0.1 and uniform on [1, 10] with probability 0.9. The change-point in the symmetric scenarios is the median of x, but its counterpart in the asymmetric scenarios is the 10th percentile of x.

We run the power simulation for these ten models. The median functions for the ten models are plotted in Web Figure 1. Under each model, for every single run of simulation, we generate a sample of (x, y) of size n from the model. The three tests are performed on the same sample with τ = 0.5, 0.7, and 0.8. The change-point search range for the bootstrap-based test is (1, 9) for the symmetric scenarios, and (0.75, 9.25) for the asymmetric scenarios. The significance level of every test is chosen to be 0.05.

The simulation results for sample size n = 200 are summarized in Table 1.

From Table 1, we can see that the test for the quadratic term outperforms the other two at all the quantiles under all the models in the symmetric scenarios in terms of power. The lack of fit test is the second best in the symmetric scenarios. The bootstrap-based test is the most powerful in the asymmetric scenarios. The lack of fit test beats the test for the quadratic term at the median, but is less powerful at the upper quantiles in the asymmetric scenarios. It is not difficult to explain the different relative performances of the three tests in the two sets of scenarios. The lack of fit test assesses the goodness of fit of the linear quantile regression model and so in the asymmetric scenarios where 90% of the data are from a linear model, it

retains the null hypothesis of goodness of fit very often. This degrades its performance as a test for the change-point existence. In order for the quadratic test to be powerful for testing the change-point existence, the data points need to be approximated well by a quadratic function, which is not the case in the asymmetric scenarios since 90% of the data points come from a linear model. Fig. 1 graphically shows the pattern of power change with $|\beta_2 - \beta_1|$ and across the three tests for $\tau = 0.7$. The power plots for $\tau = 0.5$ and 0.8 are omitted here to save space.

In sum, all three tests have pros and cons. The lack of fit test, which has power against the widest range of alternatives among the three, has good power to detect a bent line when the change-point is not at a tail of the covariate's distribution, but has poor power when it is. The test against quadratic alternatives is powerful only to detect a bent line that can be approximated well by a quadratic function. The bootstrap-based test performs well in terms of power in general, but is less powerful than the other two tests when the data are from their favorite scenarios like what we described above. The test against quadratic alternatives is the least computationally intensive and very fast, the lack of fit test has a medium computational cost, and the bootstrap-based test is the most computationally intensive. Computing times are reported for the example in Section 5 as an illustration (see Web Table 1).

4.2 Simulation for Confidence Intervals

Besides power simulation, we also investigate the performance of the confidence intervals for slopes and the change-point proposed in Section 2.2.3 through simulation. The simulation scenario is the same as the symmetric scenario (20) with $|\beta_2 - \beta_1| = 1$ in power simulation. The asymptotic confidence intervals and the bootstrapped confidence intervals are computed for one sample in every single run of simulation. The number of bootstrap replications is 500. The number of simulation runs is 1000. The simulation results are in Table 2. In all cases, we report results for confidence intervals that are intended to have coverage 0.95.

Table 2 shows that the bootstrapped confidence interval has a much better coverage than the asymptotic confidence interval in all cases. A sample size of 200 is large enough for the bootstrapped confidence interval to have good coverage, but a larger sample size than 500 is needed for an asymptotic confidence interval, especially for the change-point interval, to be accurate. An investigation on Wald confidence intervals for linear quantile regression in Koenker (2005, Section 3.10) has a similar finding that the interval's coverage is remarkably smaller than intended, even for a sample size of 500.

5. An Example

This section presents an application of bent line quantile regression to the maximal running speed example mentioned in Section 1.

5.1 An allometric model

Maximal running speed (MRS) obviously depends on size for land mammals. It is apparent that the dependency is non-monotonic; the fastest mammals are neither the largest nor the smallest. Huxley (1932) introduced an allometric equation with which MRS may be modeled to be proportional to a power of mass, i.e.

$$\text{MRS} = \exp(\alpha) \times \text{mass}^{\beta} \tag{23}$$

where α and β may vary after the mass exceeds some change-point. The logarithmic transformation renders the allometric relation linear; the exponent to which mass is raised may now be viewed as the slope of a line, with log-linear effects of covariates easily estimated by the inclusion of extra terms. So bent line quantile regression may be applied to analyze the MRS/mass data.

5.2 The data

One hundred and seven land mammal (MRS(km/h), mass(kg)) pairs were collected by Garland (given in Garland, 1983) from many sources, covering a wide range of body sizes and types. A plot of the data set on the log scale is given in Figure 2. Points belonging to animals which ambulate by hopping are labeled by 'h'. Figure 2 clearly shows the need for accommodating nonlinearity of log(MRS) in log(mass).

5.3 Analysis

There is no scientific interest in how slowly mammals can run. Indeed, there are environmental niches which are filled by animals, e.g., sloths, which give the points in the bottom center of Figure 2, for which high running speed does not give a selective advantage. The theoretical minimum obviously is near 0. However, biologists are definitely interested in the determinants of MRS for organisms in niches for which speed is important (see discussion in Garland, 1983); thus we estimate the median and higher quantiles. In particular, the hypothesis of isometry, that biological properties (speed in this case) are invariant with respect to size, is of great and general interest. Here isometry means speed is proportional to length, which in turn is proportional to mass$^{1/3}$, implying that (23) holds with $\beta = 1/3$. Other values of β, such as 0.25 for elastic stress similarity and 0.40 for static stress similarity, have been proposed (see Garland, 1983). Isometry and other models implying positive β may not hold for larger animals in which structural and metabolic constraints may apply, in which case speed would be expected to stabilize or decrease with mass.

The bent line quantile regression model was fitted to Garland's data under three specific values of τ: 0.5, 0.7 and 0.8. Y_i is log(MRS), X_i is log(mass), and z_i accounts for the hopping effect, i.e., $z_i = 1$ if the i-th animal is a hopper; otherwise, $z_i = 0$. The three tests for change-point existence all showed that the bent line patterns in Figure 2 are highly statistically significant (the p-values for all the three tests at all the three quantiles are 0, and the bootstrap-based test uses 400 bootstrap replications and searches for change-point in (-3.04, 6.15)). The data analysis results are tabulated in Table 3.

Since all three tests indicate that for the median and higher quantiles the slope with respect to log(mass) does change in the bent line quantile regression model, the estimates and the confidence intervals in Table 3 are valid. The estimated coefficients show that the maximal running speed increases first and then gradually drops as the body mass gets bigger. The change-point is about 34kg for the 0.5th and 0.7th quantiles, and 37kg for the 0.8th quantile. Hopping has a positive effect on running speed. These estimates are consistent (or nearly so for $\tau = 0.8$) with the hypothesis of isometry in small mammals and indicate negative allometry for large ones.

We performed a statistical test for the equality of the change points across the three quantile levels using the test statistic (16). The resulting p-value is 0.96, so there is no statistically significant difference in the change point across the three quantile levels.

The three bent quantile regression lines excluding hoppers are plotted with the data in Figure 2. The 0.8th quantile line approaches the 0.7th quantile line at the right tail. This might be due to not enough data in that region to distinguish the two quantiles well.

6. Discussion

Bent line quantile regression as a special case of nonlinear quantile regression retains many of quantile regression's good properties. For example, it is robust to any response outlier in the data, and especially useful when high or low quantiles instead of the mean are of particular interest. More importantly, bent line quantile regression enables us to use quantile regression to analyze data for which the linear relationship is not constant. Such data can be found in allometry, nutrition, medicine, etc. (Pawitan, 2005), which means the area where bent line quantile regression could be applied is large. One caution about the bent line quantile estimation here is quantile crossing. Crossing is theoretically possible due to the virtues of independently estimating a family of conditional bent line quantile functions. In such a case, we can restrict the bent line estimates to prevent this, similar to He's (1997) results for the linear quantile regression case, but a price to pay for using the method of He (1997) in a bent line case is the assumption that the change-point is the same at different quantile levels. The authors have implememented this algorithm and can be contacted for details.

In real applications, the problem of how to identify which covariates have slope changes would arise in the presence of at least two covariates. As a general model-building and evaluation strategy we could examine a vector **X** of covariates for nonlinearity. One possibility suggested by a referee is for the testing procedures proposed above except the lack of fit test to be applied separately to each element of **X**. This would be appropriate when we knew nonlinearity to be restricted to at most a single element. A computationally tractable simultaneous test is desirable in the general case. One practical solution would be to conduct simultaneous tests of all quadratic terms. After we decide which subset of **X** has bends, if any, then we can extend the computational methods of Section 2.2.1 to estimate potentially multivariate **t** and coefficients β_1, β_2. This actually extends the line of research in bent line quantile regression in the direction of multiple covariates, each having its own break point. In this case, the dimension of the change-point is greater than one, and we have a folded hyperplane instead of a bent line. The parameter estimates and their asymptotic properties can be derived using similar arguments as in the one-dimensional change-point case. The test for quadratic terms can be used to test for the existence of each change-point simultaneously.

Another direction to extend the current research is to consider more than one break point for a covariate so that bent line quantile regression is generalized to multi-phase quantile regression. Estimation and inference for multi-phase quantile regression are not so analogous to our work as the previous case because the change-points are ordered. Optimization with inequality constraints and the asymptotics of constrained M-estimation are involved in this case. The tests for the existence of change-point or change-points would be complex and need much innovative work. Instead of multi-phase quantile regression, one could refer to Koenker et al. (1994) for quantile estimation in a general nonlinear situation using smoothing splines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Ying Wei's research was supported by the National Science Foundation (DMS-0906568) and a career award from NIEHS Center for Environmental Health in Northern Manhattan (ES009089). Rick Chappell's research was supported by NIH grant P50 AG033514. Xuming He's research was partially supported by NSF awards DMS-0604229, DMS-0724752, and NIH Grant 1R01GM08050301-A2. The authors thank Professors S. Portnoy

and R. Koenker for some helpful discussions, and also thank the associate editor, the two referees and the editor G. Molenberghs for their useful comments that improved the presentation of the paper.

References

Bassett G, Koenker R. Tests of linear hypotheses and l_1 estimation. Econometrica. 1982; 50:1577–1584.

Chappell R. Fitting bent lines to data, with applications to allometry. Journal of Theoretical Biology. 1989; 138:235–256. [PubMed: 2607772]

Feder PI. On asymptotic distribution theory in segmented regression problems—identified case. Annals of Statistics. 1975; 3:49–83.

Garland T. The relation between maximal running speed and body mass in terrestrial mammals. Journal of Theoretical Biology. 1983; 199:157–170.

Gijbels I, Goderniaux AC. Bootstrap test for change-points in nonparametric regression. Journal of Nonparametric Statistics. 2004; 16:591–611.

He X. Quantile curves without crossing. The American Statistician. 1997; 51:186–192.

He X, Zhu L. A lack-of-fit test for quantile regression. Journal of American Statistical Association. 2003; 98:1013–1022.

Hinkley, DV.; Chapman, P.; Runger, G. Technical Report No 382. University of Minnesota; 1980. Change-point problems.

Huxley, JS. Problems of Relative Growth. Methuen; London: 1932.

Koenker, R. Quantile Regression. Cambridge; New York: 2005.

Koenker R, Bassett G. Regression quantiles. Econometrica. 1978; 46:33–50.

Koenker R, Ng P, Portnoy S. Quantile smoothing splines. Biometrika. 1994; 81:673–680.

Pawitan, Y. Change-point problem. In: Armitage, P.; Colton, T., editors. Encyclopedia of Biostatistics. 2nd. John Wiley & Sons; 2005. Published online: July 15, 2005

Quandt RE. The estimation of the parameters of a linear regression system obeying two separate regimes. Journal of American Statistical Association. 1958; 53:873–880.

Quandt RE. Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of American Statistical Association. 1960; 55:324–330.

Robinson D. Estimates for the points of intersection of two polynomial regressions. Journal of American Statistical Association. 1964; 59:214–224.

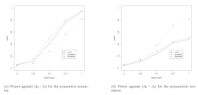

Figure 1.
Power plots for the three tests in the two sets of scenarios given $\tau = 0.7$.

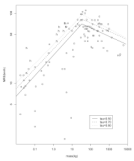

Figure 2.
MRS vs. mass of land mammals, on the logarithm scale (hoppers denoted “h”), fitted with bent line quantile regression.

Table 1

Power simulation results of the three tests for the change-point existence

			$\lvert\beta_2-\beta_1\rvert$				
Scenario	τ	**Test**	**0**	**0.25**	**0.5**	**0.75**	**1**
Symmetric	0.5	L.O.F.	0.049	0.131	0.487	0.816	0.963
		Quadratic	0.059	0.143	0.518	0.843	0.966
		Bootstrap	0.045	0.100	0.332	0.676	0.869
	0.7	L.O.F.	0.056	0.110	0.407	0.759	0.946
		Quadratic	0.053	0.153	0.488	0.798	0.955
		Bootstrap	0.034	0.074	0.275	0.559	0.817
	0.8	L.O.F.	0.062	0.095	0.341	0.657	0.877
		Quadratic	0.049	0.143	0.432	0.710	0.916
		Bootstrap	0.032	0.077	0.208	0.440	0.688
Asymmetric	0.5	L.O.F.	0.052	0.161	0.463	0.724	0.829
		Quadratic	0.049	0.124	0.381	0.590	0.694
		Bootstrap	0.036	0.158	0.538	0.825	0.909
	0.7	L.O.F.	0.044	0.119	0.262	0.415	0.483
		Quadratic	0.051	0.143	0.312	0.436	0.516
		Bootstrap	0.044	0.157	0.414	0.711	0.818
	0.8	L.O.F.	0.053	0.084	0.137	0.221	0.246
		Quadratic	0.045	0.137	0.248	0.343	0.380
		Bootstrap	0.036	0.129	0.331	0.523	0.682

NOTE: L.O.F.: lack of fit test; Quadratic: test for the quadratic term; Bootstrap: bootstrap-based test. The number of runs per simulation is 1000. The sample size is is 200. The number of bootstrap samples for the bootstrap-based test is 400.

Table 2

Simulation results of the confidence intervals' performance

			Mean Interval Width			Coverage		
Sample Size	**Interval Type**	τ	β_1	β_2	t	β_1	β_2	t
n = 200	Asymptotic	0.5	0.542	1.042	2.752	0.897	0.902	0.812
		0.7	0.573	1.149	2.989	0.905	0.906	0.810
		0.8	0.631	1.296	3.289	0.900	0.893	0.811
	Bootstrapped	0.5	0.811	2.011	4.212	0.977	0.968	0.965
		0.7	0.838	2.052	4.274	0.982	0.968	0.971
		0.8	0.923	2.254	4.510	0.982	0.966	0.974
n = 500	Asymptotic	0.5	0.326	0.568	1.748	0.905	0.914	0.853
		0.7	0.344	0.614	1.854	0.904	0.911	0.853
		0.8	0.379	0.664	2.024	0.901	0.904	0.838
	Bootstrapped	0.5	0.438	0.908	2.775	0.972	0.964	0.954
		0.7	0.467	0.989	2.930	0.971	0.975	0.964
		0.8	0.522	1.112	3.181	0.980	0.966	0.972

Table 3
Parameter estimates and confidence intervals for change-point and slopes

	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.8$
$\hat{\alpha_1}$	3.23	3.43	3.58
$\hat{\alpha_2}$	4.68	4.84	5.07
β_1	0.29	0.28	0.27
95% Boot. C.I. for β_1	[0.21, 0.34]	[0.20, 0.34]	[0.23, 0.34]
95% Asymp. C.I. for β_1	[0.24, 0.34]	[0.23, 0.33]	[0.22, 0.32]
β_2	-0.12	-0.12	-0.14
95% Boot. C.I. for β_2	[-0.48, -0.03]	[-0.41, -0.02]	[-0.26, -0.05]
95% Asymp. C.I. for β_2	[-0.25, 0]	[-0.19, -0.04]	[-0.20, -0.08]
Change-point $\hat{t}$	3.53	3.53	3.61
95% Boot. C.I. for t	[3.00, 5.85]	[2.62, 5.41]	[2.43, 4.83]
95% Asymp. C.I. for t	[2.68, 4.38]	[2.78, 4.27]	[2.96, 4.26]
$\hat{\gamma}$	0.61	0.43	0.32

NOTE: Boot. C.I.: bootstrap confidence inteval; Asymp. C.I.: asymptotic confidence interval. 500 bootstrap samples are used for constructing a confidence interval.