Web-based Supplementary Materials for Bent Line Quantile Regression with Application to an Allometric Study of Land Mammals Speed and Mass by Chenxi Li, Ying Wei, Rick Chappell and Xuming He
Web Appendix A: Proof of Theorem 1 and 2

To prove Theorem 1 and 2, we introduce two more notations in addition to the ones used in
the two theorems:

\[ S_\tau(\eta, t) = \lim_{n \to \infty} S_{n,\tau}(\eta, t) \text{ in probability and } \]

\[ \eta_{0,\tau}(t) = \arg \min_{\eta} S_\tau(\eta, t), \]

and make the following assumptions:

(A0) The limiting objective function \( S_\tau(\eta, t) \) reaches its unique global minimum at \( \theta_0(\tau) \).

(A1) \( 0 < \inf_\tau f_{\tau,\omega_i}(0) < \sup_\tau f_{\tau,\omega_i}(0) < \infty \) for any sequence of values of \( \omega_i \).

(A2) \( X_i \) has continuous density function \( p(x) \) on a bounded support \([-M, M]\), where \( M \) is a positive constant.

(A3) \( \max_{1 \leq i \leq n} \| z_i \| = o_p(n^{1/2}) \) and \( E(\| z \|^3) < \infty \).

(A4) Given \( \beta_{1,\tau} \neq \beta_{2,\tau} \), there exists a nonnegative definite matrix \( C_{0,\tau} \), such that,

\[ C_{n,\tau} \to C_{0,\tau}. \]

(A5) Given \( \beta_{1,\tau} \neq \beta_{2,\tau} \), there exists a full rank matrix \( D_{0,\tau} \), such that

\[ D_{n,\tau} \to D_{0,\tau}. \]

Under Assumptions (A1) to (A3), \( S_\tau(\eta, t) \) is well defined by the weak law of large numbers. Assumption (A0) is the identifiability condition of the estimation, which also implies the density of \( X, p(x) \), is nonzero around \( t_\tau \). In other words, \( t_\tau \) is in the interior of \( x \)-space such that the numbers of \( X_i \)’s on both sides of \( t \) go to \( \infty \) as \( n \to \infty \). Assumptions (A1) and (A3) are standard regularity conditions in quantile regression. Assumptions (A0) to (A3) together suffice for the consistency of \( \hat{\theta}_n(\tau) \), and Assumptions (A4) to (A5) are used for the asymptotic normality.
Lemma 1: Under Assumptions (A0) to (A3), \( \hat{\theta}_n(\tau) \) is a consistent estimator of \( \theta_0(\tau) \).

Proof: With fixed \( t \), the bent line model is linear in form. Using the similar augments for the consistency of the linear quantile regression estimates given by Koenker (2005), we can show that, under Assumptions (A1) to (A3),

\[
\sup_{|t|<M} \| \hat{\eta}_{n,\tau}(t) - \eta_{0,\tau}(t) \| = o_p(1).
\]

(1)

On the other hand, Assumption (A2) implies that \( \hat{t}_{n,\tau} \) is bounded by \( M \), and \( \eta_{0,\tau}(t) \) is a continuous function of \( t \). Notice that \( |\hat{\eta}_{n,\tau}(\hat{t}_{n,\tau}) - \eta_{0,\tau}(\hat{t}_{n,\tau})| \leq |\hat{\eta}_{n,\tau}(\hat{t}_{n,\tau}) - \eta_{0,\tau}(\hat{t}_{n,\tau})| + |\eta_{0,\tau}(\hat{t}_{n,\tau}) - \eta_{0,\tau}(t)| \), based on (1), we only need to show \( |\hat{t}_{n,\tau} - t_{\tau}| = o_p(1) \) for the consistency of \( \hat{\theta}_n(\tau) \).

Based on Assumptions (A0) to (A2), the limiting objective function \( S_{\tau}(\eta_{0,\tau}(t), t) \) is a continuous function of \( t \), and uniquely minimized at \( t_{\tau} \). Since \( \hat{t}_{n,\tau} \) is bounded by \( M \), the following condition suffices the consistency of \( \hat{t}_{n,\tau} \),

\[
\sup_{|t|<M} |S_{n,\tau}(\hat{\eta}_{n,\tau}(t), t) - S_{\tau}(\eta_{0,\tau}(t), t)| = o_p(1)
\]

(2)

Following the chaining argument in Huber (1967), one can show that \( \sup_{|t|<M; |\eta|<M'} |S_{n,\tau}(\eta, t) - S_{\tau}(\eta, t)| = o_p(1) \), which, together with the uniform convergence in (1), implies that

\[ \sup_{|t|<M} |S_{n,\tau}(\hat{\eta}_{n,\tau}(t), t) - S_{\tau}(\hat{\eta}_{n,\tau}(t), t)| = o_p(1). \]

On the other hand, since \( S_{\tau}(\cdot, t) \) is a continuous function of \( \eta \), the uniform convergence of \( \hat{\eta}_{n,\tau}(t) \) also implies \( \sup_{|t|<M} |S_{\tau}(\hat{\eta}_{n,\tau}(t), t) - S_{\tau}(\eta_{0,\tau}(t), t)| = o_p(1) \). Finally, notice that the left side of (2) is bounded by \( \sup_{|t|<M} |S_{n,\tau}(\hat{\eta}_{n,\tau}(t), t) - S_{\tau}(\hat{\eta}_{n,\tau}(t), t)| + \sup_{|t|<M} |S_{\tau}(\hat{\eta}_{n,\tau}(t), t) - S_{\tau}(\eta_{0,\tau}(t), t)| \), (2) holds immediately, and the proof of Lemma 1 is hence complete.

Lemma 2: Define

\[
u_i(\theta, \theta_0(\tau)) = \psi_{\tau}(Y_i - g(w_i; \theta))h(w_i; \theta) - \psi_{\tau}(Y_i - g(w_i; \theta_0(\tau)))h(w_i; \theta_0(\tau)),
\]

then, under Assumptions (A1) to (A3), for any positive sequence \( d_n \) converging to zero, we
have
\[ \sup_{\|\theta - \theta_0(\tau)\| \leq d_n} n^{-1/2} \left\| \sum_{i=1}^{n} \{ u_i(\theta, \theta_0(\tau)) - E(Y_i, w_i)[u_i(\theta, \theta_0(\tau))] \} \right\| = o_p(1). \] (3)

**Proof:** We first partition \( u_i(\theta, \theta_0(\tau)) \) based on the value of \( X_i \) such that
\[
u_i(\theta, \theta_0(\tau)) = u_i(\theta, \theta_0(\tau)) \mathbb{I}\{X_i > \max(t, \tau)\} + u_i(\theta, \theta_0(\tau)) \mathbb{I}\{X_i \leq \min(t, \tau)\} + u_i(\theta, \theta_0(\tau)) \mathbb{I}\{t < X_i < \tau\} + u_i(\theta, \theta_0(\tau)) \mathbb{I}\{\tau < X_i < t\}
\]
\[ = u_i,1(\theta, \theta_0(\tau)) + u_i,2(\theta, \theta_0(\tau)) + u_i,3(\theta, \theta_0(\tau)) + u_i,4(\theta, \theta_0(\tau)). \] (4)

Since \( \|u_i,1(\theta, \theta_0(\tau))\| \leq \|w_i\| \mathbb{I}\{|e_{i,1}| \leq |(\alpha - \alpha_r) + (\beta_2 - \beta_2, \tau)X_i - (\beta_2 t - \beta_2, \tau) + z_i^T (\gamma - \gamma_r)|\} \), based on the mean value theorem, we have
\[ E_Y[\| u_i,1(\theta, \theta_0(\tau)) \|^2 | w_i] \leq L d_f n \| w_i \|^3, \] (5)
where \( L \) is some positive constant and \( f_n^{*, \tau, w_i} \) is some intermediate density with \( f_n^{*, \tau, w_i} \rightarrow f^{*, \tau, w_i}(0) \) almost surely when \( n \rightarrow \infty \). Let \( A_n = L \sum_{i} f_n^{*, \tau, w_i} \| w_i \|^3 \). Under Assumptions (A1) to (A3), \( A_n = O_p(n) \). Under the same assumptions, we also have
\[ \max_{1 \leq i \leq n} \| u_i,1(\theta, \theta_0(\tau)) \| = O_p(n^{1/2}). \] (6)

Similarly, the equations (5) and (6) also hold for \( u_i,2 \) to \( u_i,4 \). Following the similar arguments used in Lemma 4.6 of He and Shao (1996), the uniform convergence (3) is implied by (6) and (5) for \( u_i,1 \) to \( u_i,4 \), Lemma 2 is hence proved.

**Proof of Theorem 1:** Due to the consistency of \( \hat{\theta}_n(\tau) \) as shown in Lemma 1, a direct consequence of Lemma 2 is
\[ n^{-1/2} \left\| \sum_{i=1}^{n} \psi_\tau(Y_i - g(w_i; \hat{\theta}_n(\tau))) h(w_i; \hat{\theta}_n(\tau)) - \sum_{i=1}^{n} \psi_\tau(e_{i,1}) h(w_i; \theta_0(\tau)) - \sum_{i=1}^{n} E(Y_i, w_i)[\psi_\tau(Y_i - g(w_i; \hat{\theta}_n(\tau))) h(w_i; \hat{\theta}_n(\tau))] \right\| = o_p(1). \] (7)

We then Taylor expand \( \sum_{i=1}^{n} E(Y_i, w_i)[\psi_\tau(Y_i - g(w_i; \hat{\theta}_n(\tau))) h(w_i; \hat{\theta}_n(\tau))] \) around \( \theta_0(\tau) \), which, together with the fact that \( \sum_{i=1}^{n} E(Y_i, w_i)[\psi_\tau(Y_i - g(w_i; \hat{\theta}_n(\tau))) h(w_i; \theta_0(\tau))] = 0 \), leads to the
following equations,

\[
E \sum_{i=1}^{n} \psi_{\tau}(Y_i - g(w_i; \hat{\theta}_n(\tau)))h(w_i; \hat{\theta}_n(\tau))
= \frac{\partial E \sum_{i=1}^{n} \psi_{\tau}(Y_i - g(w_i; \theta))h(w_i; \theta)}{\partial \theta}|_{\theta=\theta_0(\tau)} \left( \hat{\theta}_n(\tau) - \theta_0(\tau) \right) + R_n
\approx nD_{n,\tau} \left( \hat{\theta}_n(\tau) - \theta_0(\tau) \right) + R_n,
\]

where \(R_n = o_p(n^{1/2})\).

Since both \(g(w_i; \theta)\) and \(h(w_i; \theta)\) are differentiable with respect to \(\alpha, \beta_1, \beta_2\) and \(\gamma\), together with the assumption (A1), the expectation \(E \sum_{i=1}^{n} \psi_{\tau}(Y_i - g(w_i; \theta))h(w_i; \theta)\) is consequently differentiable with respect to those parameters. In what follows, we show that, under Assumptions (A1) and (A2), the expectation \(E \sum_{i=1}^{n} \psi_{\tau}(Y_i - g(w_i; \theta))h(w_i; \theta)\) is also differentiable with respect to \(t\). Hence the matrix \(D_{n,\tau}\) is well defined. \(g_1(w_i; \theta)\) and \(g_2(w_i; \theta)\) are the conditional quantile functions respectively on the left and right side of \(t_\tau\). It is easy to see that \(g_1()\) and \(g_2()\) are differentiable respective to \(t\). Letting \(p_{X_i|z_i}(x)\) be the conditional density function of \(X_i\) given \(z_i\), we then decompose

\[
E \psi_{\tau}(Y_i - g(w_i; \theta))h(z_i, X_i; \theta)
= \sum_{i=1}^{n} E_{z_i} \left[ \int_{-\infty}^{t} E_{Y_i} \left\{ \psi_{\tau}(Y_i - g_1(z_i, x; \theta))(1, x, 0, 0, z_i^\top) | w_i \right\} p_{X_i|z_i}(x) dx \right]
+ E_{z_i} \left[ \int_{t}^{\infty} E_{Y_i} \left\{ \psi_{\tau}(Y_i - g_2(z_i, x; \theta))(0, 0, 1, x, z_i^\top) | w_i \right\} p_{X_i|z_i}(x) dx \right]
\approx P_1(t) + P_2(t).
\]

The first derivative of \(P_1(t)\) with respect to \(t\) is then

\[
\frac{\partial P_1(t)}{\partial t} = E_{z_i} \left[ E_{Y_i} \left\{ \psi_{\tau}(Y_i - g_1(z_i, t; \theta))(1, t, 0, 0, z_i^\top) | w_i = (1, t, z_i^\top) \right\} p_{X_i|z_i}(t) \right]
+ E_{z_i} \left[ \int_{-\infty}^{t} \frac{\partial E_{Y_i} \left\{ \psi_{\tau}(Y_i - g_1(z_i, x; \theta))(1, x, 0, 0, z_i^\top) | w_i \right\}}{\partial t} p_{X_i|z_i}(x) dx \right]
= E_{z_i} \left[ E_{Y_i} \left\{ \psi_{\tau}(Y_i - g_1(z_i, t; \theta)) | w_i = (1, t, z_i^\top) \right\} p_{X_i|z_i}(t)(1, t, 0, 0, z_i^\top) \right]
- E_{z_i} \left[ \int_{-\infty}^{t} f_{\tau,w_i}(g_1(z_i, x; \theta) - g(z_i, x; \theta(\tau))) \frac{\partial g_1(z_i, x; \theta)}{\partial t} (1, x, 0, 0, z_i^\top) p_{X_i|z_i}(x) dx \right],
\]
Since and similarly, 

\[
\frac{\partial P_2(t)}{\partial t} = E_{z_i} \left[ -E_Y \left\{ \psi_Y(Y_i - g_2(z_i, t; \theta)) | w_i = (1, t, z_i^\top) \right\} p_{X_i|z_i}(0, 0, 1, t, z_i^\top) \right] 
\]

\[= -E_{z_i} \int_t^M f_{\tau, w_i}(g_2(z_i, x; \theta)) \frac{\partial g_2(z_i, x; \theta)}{\partial t} \left(0, 0, 1, x, z_i^\top\right) p_{X_i|z_i}(x) dx \]

Since \(p_{X_i|z_i}(x)\) is a continuous function, and both \(g_1()\) and \(g_2()\) are differentiable, the differentiability of \(E\psi_Y(Y_i - g(w_i; \theta))h(w_i; \theta)\) respective to \(t\) follows immediately.

Given the defined \(D_{n, \tau}\), and combing (7) and (8), we have

\[n^{1/2}(\hat{\theta}_n(\tau) - \theta_0(\tau)) = -n^{-1/2}D_{n, \tau}^{-1} \sum_{i=1}^n \psi_Y(e_{\tau,i})h(w_i; \theta_0(\tau)) + o_p(1). \quad (9)\]

Based on the central limit theorem, \(n^{-1/2}D_{n, \tau}^{-1} \sum_{i=1}^n \psi_Y(e_{\tau,i})h(w_i; \theta_0(\tau))\) is asymptotically normal with limiting variance covariance matrix \(\tau(1 - \tau)D_{0, \tau}^{-1}C_{0, \tau}D_{0, \tau}^{-1}\), hence, the rest of Theorem 1 holds immediately.

The proof of Theorem 1 can be easily extended to Theorem 2. To prove Theorem 2, we assume that Assumptions (A0) - (A3) and (A5) hold for every \(\tau_j, j = 1, \cdots, m\), and as a result (A5) implies that \(H_n \to H_0\). Besides, we need to make one more assumption, which generalizes Assumption (A4) to multiple quantiles:

(A4') Given \(\beta_{1, \tau_j} \neq \beta_{2, \tau_j}, j = 1, \cdots, m\), there exists a nonnegative definite matrix \(J_0\), such that

\[
J_n = n^{-1} \sum_{i=1}^n E \begin{pmatrix}
  h(w_i, \theta_0(\tau_1)) \\
  \vdots \\
  h(w_i, \theta_0(\tau_m))
\end{pmatrix} \to J_0.
\]

**Proof of Theorem 2:** Under Assumptions (A0) - (A3), (A4'), (A5) and the condition that \(\beta_{1, \tau_j} \neq \beta_{2, \tau_j}\) for \(j = 1, \cdots, m\), each component of \(\hat{\theta}_n\) has a Bahadur representation according to Theorem 1:

\[n^{1/2}(\hat{\theta}_n(\tau_j) - \theta_0(\tau_j)) = -n^{-1/2}D_{n, \tau_j}^{-1} \sum_{i=1}^n \psi_{\tau_j}(e_{\tau_j,i})h(w_i; \theta_0(\tau_j)) + o_p(1), \quad j = 1, \cdots, m, \quad (10)\]
from which it is easy to show that

\[ n^{1/2}(\hat{\theta}_n - \theta_0) = -n^{-1/2}H_n^{-1} \sum_{i=1}^{n} \begin{pmatrix} \psi_{\tau_1}(e_{\tau_1,i})h(w_i, \theta_0(\tau_1)) \\ \vdots \\ \psi_{\tau_m}(e_{\tau_m,i})h(w_i, \theta_0(\tau_m)) \end{pmatrix} + o_p(1). \]  

(11)

Based on the central limit theorem, \( n^{-1/2}H_n^{-1} \sum_{i=1}^{n} \begin{pmatrix} \psi_{\tau_1}(e_{\tau_1,i})h(w_i, \theta_0(\tau_1)) \\ \vdots \\ \psi_{\tau_m}(e_{\tau_m,i})h(w_i, \theta_0(\tau_m)) \end{pmatrix} \) is asymptotically normal with limiting variance covariance matrix \( H_0^{-1}\Lambda_0H_0^{-\top} \), and thus the rest of Theorem 2 holds immediately.

Web Appendix B: Computational Aspects of Bent line Quantile Regression

The optimization method we use for minimizing \( S(\hat{\eta}|t) \) defined in (8) in the paper is a combination of golden section search and successive parabolic interpolation, which is implemented by the function ‘optimize’ in R package ‘stats’. It requires no more than about

\[ 2K \left[ \log_2 \left( \frac{3(b-a)}{tol} \right) \right]^2 \]

function evaluations, where

\[ K = \frac{1}{\log_2[(1 + \sqrt{5})/2]} \approx 1.44, \]

\( a \) and \( b \) are the lower and upper bounds of the interval over which the change-point is searched for, and \( tol \) is the desired accuracy (see Brent, 1973, Section 5.5). Therefore, it is easily deduced that the computational cost of the parameter estimation for the bent quantile regression with \( 1 + q \) covariates does not exceed \( 2K \left[ \log_2 \left( \frac{3(b-a)}{tol} \right) \right]^2 \) times the cost of the standard linear quantile regression with \( 2 + q \) covariates.

We compare the computational costs between the bent line quantile regression and the standard linear quantile regression on the land mammals’ speed and mass data. These two procedures were run in R version 2.9.1 on a 64-bit linux computer (Intel(R) Pentium(R) 4
CPU 3.20GHz, 3G RAM). The CPU times recorded are in Web Table 1 which shows that parameter estimation and asymptotic covariance matrix estimation in both of the procedures take less than 1 second. The bent line quantile regression takes remarkably longer than the linear quantile regression for estimating parameters and their covariance matrix, which is due to the extra step of searching for change-point as well as the resulting higher dimension of parameter space in bent line quantile regression. As to the tests for the change-point existence, the test for the quadratic term, the lack of fit test and the bootstrap-based test take about 0.014, 15 and 68 seconds, respectively.

References


Web Figure 1. The median functions $Q(x)$ in the two sets of simulation scenarios.

Web Table 1

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Component</th>
<th>$\tau = 0.5$</th>
<th>$\tau = 0.7$</th>
<th>$\tau = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bent Line</td>
<td>Parameter Estimation</td>
<td>0.139</td>
<td>0.105</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>Asymptotic Covariance Matrix Estimation</td>
<td>0.263</td>
<td>0.285</td>
<td>0.252</td>
</tr>
<tr>
<td></td>
<td>L.O.F.</td>
<td>15.327</td>
<td>14.620</td>
<td>14.778</td>
</tr>
<tr>
<td></td>
<td>Quadratic</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Bootstrap</td>
<td>68.243</td>
<td>67.452</td>
<td>67.303</td>
</tr>
<tr>
<td>Linear</td>
<td>Parameter Estimation</td>
<td>0.005</td>
<td>0.006</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Asymptotic Covariance Matrix Estimation</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>

NOTE: L.O.F.—lack of fit test, Quadratic—test for the quadratic term and Bootstrap—bootstrap-based test are the three tests for the existence of change-point. The linear quantile regression assumes that $\log(\text{MRS})$ is linear in $\log(\text{mass})$. 