TUTORIAL ON QUANTILE REGRESSION

Xuming He
Department of Statistics, University of Illinois at Urbana-Champaign

Ying Wei
Department of Biostatistics, Columbia University

ENAR 2005
Motivation for quantile regression

- Regression — to obtain a summary of the relationship between a response variable $y$ and a set of covariates $x$.

- Least squares regression captures how the mean of $y$ changes with $x$.

- Sometimes, a single *mean* curve is not informative enough.

- Conditional quantile functions provide a more complete view.
Motivation for quantile regression

The distribution of $y$ may be asymmetric around the mean.

Heteroscedasticity may exist in the data.
Example 1 – Risk factors for low birthweight

- Low birthweight is known to be associated with
  - higher infant mortality (Abreveya, 2001).
  - higher health-care cost (Lewit et al., 1995).
  - a wide range of subsequent health problems (Hack et al., 1995).
  - long-term educational attainment and even labor market outcomes (Corman and Chaikind, 1998, and Currie and Hyson, 1999).

- Investigate the factors influencing birthweight, especially the ones that may help reduce the incidence of low birthweight infants.
Example 1 – Risk factors for low birthweight

• The research question can be rephrased as exploring the covariate effects on the lower quantiles of birthweight.

• Potential covariates include
  ◦ Mother’s education
  ◦ Mother’s prenatal care
  ◦ Mother’s age
  ◦ Mother’s weight gain
  ◦ …

• Covariate effects on lower quantiles may differ from those on the mean or median birthweight.

Example 2 – Growth charts

- Display the distribution of a certain measurement within a certain range of time for a certain population.
Example 2 – Growth charts

- Used to screen the measurements from an individual subject in the context of population values.
- Focused on the entire distribution, especially the extreme quantiles.
Example 2 – Growth charts

- Widely used in medical sciences:

  **Pediatrics:**
  to monitor an infant’s growth status by screening his/her height, weight and head circumference. (Hamill et al., 1979)

  **HIV research:**
  to monitor CD4 lymphocyte counts in uninfected children born to HIV-1 infected women.

  **Genetics:**
  to help determine the gene frequency of the most common mutations of the HFE gene causing hereditary hemochromatosis. (Koziol et al., 2001)
Outlines

• Quantile regression
  ◦ Estimation
  ◦ Properties
  ◦ Computational aspects

• Software in SAS and R

• Examples (continue)

• Inference in quantile regression

• Additional topics
  ◦ Beyond the linear models
  ◦ Quantile regression with censored data
Univariate Sample Quantile via Optimization

A random sample \( \{y_1, y_2, \ldots, y_n\} \),

Median \( = \arg\min_\xi \sum |y_i - \xi|; \)

The \( \tau \)th sample quantile \( = \arg\min_\xi \sum \rho_\tau(y_i - \xi) \),

where

\( \rho_\tau(u) = u(\tau - 1_{(u<0)}) \).
Linear Regression Quantile (Koenker and Bassett, 1978)

- A linear model for the $\tau$th quantile

$$ y_i = x_i^\top \beta_\tau + e_i, \quad i = 1, \ldots, n, $$

(1.1)

where the $\tau$th quantile of $e_i$ is zero.

- Underlying assumption of model (1.1):

$$ Q_\tau(Y | X) = X^\top \beta_\tau $$

- Estimator of $\beta_\tau$

$$ \hat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta). $$

- Regression Quantile: $\hat{Q}_\tau(Y | X) = X^\top \hat{\beta}_\tau.$
Basic properties of regression quantiles

- Scale equivariant: for any $a > 0$,
  \[
  \hat{\beta}_\tau(a y, X) = a \hat{\beta}_\tau(y, X),
  \]
  \[
  \hat{\beta}_\tau(-a y, X) = a \hat{\beta}_{1-\tau}(y, X).
  \]

- Regression Shift: for any $\gamma \in \mathbb{R}^p$,
  \[
  \hat{\beta}_\tau(y + X\gamma, X) = \hat{\beta}_\tau(y, X) + \gamma.
  \]

- Reparameterization of design: for any $|A| \neq 0$,
  \[
  \hat{\beta}_\tau(y, AX) = A^{-1} \hat{\beta}_\tau(y, X).
  \]

- Equivariant to monotone transformation,
  \[
  Q_\tau(h(Y)|X) = h(Q_\tau(Y|X)).
  \]
Asymptotic properties of regression quantiles

- Asymptotics for the univariate sample quantile.

Let \( \{y_1, y_2, \ldots, y_n\} \) be an i.i.d. random sample with distribution \( F \), and

\[
\hat{\xi}_\tau = \arg\min_{\xi} \sum_i \rho_\tau(y_i - \xi),
\]

then

\[
\sqrt{n}(\hat{\xi}_\tau - \xi_\tau) \sim N \left( 0, \frac{\tau(1 - \tau)}{f^2(F^{-1}(\tau))} \right),
\]

where \( f = F' \).
Asymptotic properties of regression quantiles

- Asymptotics for linear models.

\[ y_i = x_i^\top \beta_\tau + e_i, \quad \text{where } e_i \sim F_i \text{ and } F_i^{-1}(\tau) = 0 \]

Then the joint distribution

\[ \sqrt{n} V_n^{-1/2} (\hat{\beta}_{\tau_k} - \beta_{\tau_k})_{k=1}^m \sim N(0, I_n), \quad (1.2) \]

where

- \( H_n(\tau) = n^{-1} \sum x_i x_i^\top f_i(0) \), and \( D_n = n^{-1} \sum x_i x_i^\top \);

- \( V_n = [(\tau_k \wedge \tau_l - \tau_k \tau_l) H_n(\tau_k)^{-1} D_n H_n(\tau_l)^{-1}]_{k,l=1}^m. \)

Computational aspects

• $\hat{\beta}_\tau$ can be calculated by means of linear programming, which is efficient in computation.

  ◦ Primal formulation as a Linear Program

  \[
  \min \{ \tau \mathbf{1}^\top u + (1 - \tau) \mathbf{1}^\top v \mid y = Xb + u - v, \ (b, u, v) \in \mathbb{R}^p \times \mathbb{R}^{2n}_+ \}.
  \]

  ◦ Dual formulation as a Linear Program

  \[
  \max \{ y^\top d \mid X^\top d = (1 - \tau) X^\top \mathbf{1}, \ d \in [0, 1]^n \}.
  \]

  ◦ Solutions are characterized by an exact fit to $p$ observations.
Existing algorithms and practical recommendations

• Simplex method (Koenker and D’Orey, 1987 and 1993)
  ◦ Start at a random vertex, and search around the edge of a circumscribed polygon, until the optimum is found.
  ◦ For moderate data size.

• Interior Point method (Portnoy and Koenker, 1997)
  ◦ Start from the interior of the circumscribed polygon to avoid crossing the boundary.
  ◦ Computationally efficient for large data size.
Existing algorithms and practical recommendations

- Interior point method with preprocessing (Portnoy and Koenker, 1997)
  - For very large datasets, e.g., \( n > 10^5 \).

- Smoothing method (Chen, 2004)
  - Search the optimized solution by smoothing the objective function \( \rho_\tau(\cdot) \).
Estimation in SAS

• Package: SAS/STAT PROC QUANTREG
  ◦ Available as an experimental procedure for SAS 9.1 (Windows only);
  ◦ Downloadable from http://www.sas.com/statistics

• Basic syntax

PROC QUANTREG DATA = sas-data-set <options>;
   BY variables;
   CLASS variables;
   MODEL response = independents </ options>;
RUN;
Estimation in SAS

• To specify the quantile level:
  - Use the option QUANTILE in the MODEL statement

\[
\text{MODEL } Y = X \ / \ \text{QUANTILE} = \text{<number list | ALL>};
\]

<table>
<thead>
<tr>
<th>choice of quantile(s)</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a single quantile</td>
<td>QUANTILE = 0.25</td>
</tr>
<tr>
<td>several quantiles</td>
<td>QUANTILE = 0.25 0.5 0.75</td>
</tr>
<tr>
<td>entire quantile process</td>
<td>QUANTILE = ALL</td>
</tr>
</tbody>
</table>

- Default value is 0.5, the median.
Estimation in SAS

- To specify the algorithm:
  - Use the option ALGORITHM in the PROC QUANTREG statement

<table>
<thead>
<tr>
<th>method</th>
<th>option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplex</td>
<td>ALGORITHM = SIMPLEX</td>
</tr>
<tr>
<td>Interior point</td>
<td>ALGORITHM = INTERIOR</td>
</tr>
<tr>
<td>Interior point with preprocessing</td>
<td>ALGORITHM = INTERIOR PP</td>
</tr>
<tr>
<td>Smoothing</td>
<td>ALGORITHM = SMOOTHING</td>
</tr>
</tbody>
</table>

- By default, the Simplex method is used.
Estimation in SAS

• Default Output

  **Model information**  — report the name of the data set and the response variable, the number of covariates, the number of observations, algorithm of optimization and the method for confidence intervals.

  **Summary statistics**  — report the sample mean and standard deviation, sample median, MAD and interquartile range for each variable included in the MODEL statement.

  **Quantile objective function**  — report the quantile level to be estimated and the optimized objective function.

  **Parameter Estimates**  — report the estimated coefficients and their 95% confidence intervals.
Estimation in R

- **R** package *quantreg* (contributed by Koenker).
  - Official releases of **R** and the install package of *quantreg* are available at
    
    http://lib.stat.cmu.edu/R/CRAN/

- The syntax of *rq*().

  ```r
  library(quantreg)

  rq(formula, tau=.5, data, weights, na.action, method="br", ...)
  ```
Estimation in \( \mathbb{R} \)

- Arguments

**formula** model statement, e.g.

\[ y \sim x_1 + x_2 \]

**tau** the quantile(s) to be estimated:

- If \( \tau \) is a single number \( \in (0,1) \), return a single quantile function.
- If \( \tau \) is a vector \( \in (0,1) \), return several quantile functions.
- If \( \tau \not\in (0,1) \), return the entire quantile process.
Estimation in R

**method**

- `method="br"` Simplex method (default).
- `method="fn"` the Frisch–Newton interior point method.
- `method="pfn"` the Frisch–Newton approach with preprocessing.
Example 1: Exploring the risk factors of low birthweight

- Data (NATALITY1997):
  - National Center for Health Statistics (NCHS).
  - Provide information on live births in the United States during calendar year 1997, including infant birthweight.
  - Also provide:
    - socioeconomic and demographic characteristics of mothers;
    - Mother’s prenatal care and pregnancy history
Example 1: Exploring the risk factors of low birthweight

• Data (NATALITY1997) used in our example:
  ◦ Subset: live singleton births; mothers between the ages of 18 and 45, either black or white, residing in the U.S.
  ◦ Sample size: 198,377 observations.
  ◦ More detailed description of the data is available at http://www.cdc.gov/nchs/births.htm
A quantile regression model for birthweight

**Model:**

\[ Q_\tau(\text{Birthweight}) = \]
- Boy + Married + Black
- + High School + Some College + College
- + No Prenatal + Prenatal Second + Prenatal Third
- + Smoker + Ciga Per Day
- + Mother’s Age + Mother’s Age \(^2\)
- + Mother’s weight gain + Mother’s weight gain \(^2\)

**Choice of quantile levels**

\[ \tau = (0.05, 0.1, 0.15, 0.2, 0.25, ..., 0.8, 0.85, 0.9, 0.95) \]
SAS codes for the birthweight model

ods graphics on;

PROC QUANTREG DATA=LIBRARY.natality;

MODEL weight = Boy Married Black ... Mother_Wt_Gain2/QUANTILE = 0.05 0.1 ... 0.9 0.95 PLOT = QUANTPLOT;

RUN;

ods graphics off;
Output from SAS / PROC QUANTREG
Some conclusions for Example 1

- Heteroscedasticity exists in the Natality data.
- Mother’s race, education and prenatal care have much more significant impact on lower quantiles of birthweight than the upper quantiles.
- These effects will be underestimated by least squares regression.
Example 2: Growth charts

- Data: Berkeley data of Tuddenham and Snyder (1954)
- Number of subjects: 136
- Measurements: longitudinal measurements of height and weight collected
  - quarterly between ages 0 and 3;
  - yearly between ages 2 and 8;
  - semi-yearly between ages 8 and 21.
Growth charts

Unconditional Reference Centiles — Boy’s weight, 0–18 years

Weight (kg)

Age (years)
Conditional growth chart

- Screening an individual subject given his/her prior path or other information (Wei, Pere, Koenker and He, 2004).

- e.g., the rapid weight gain during infancy could lead to obesity later in the childhood (Stettler et al., 2002).

- A simple AR(2) model

\[ Q_\tau(W_t) = \beta_{\tau 0} + \beta_{\tau 1}W_{t-1} + \beta_{\tau 2}W_{t-2} + \beta_{\tau 3}H_t, \]

where

- \( W_t \) is the current weight at time \( t \).
- \( W_{t-1} \) and \( W_{t-2} \) are two prior weights at time \( t - 1 \) and \( t - 2 \), respectively.
- \( H_t \) is the current height at time \( t \).
An illustrative example

- $W_{t-2} = 5.4\text{kg}$
- $W_{t-1} = 7.2\text{kg}$
- $W_t = 9.8\text{kg}$
- $H_t = 70.6\text{cm}$
Example of conditional growth chart (cont.)

- **Model:**

\[ Q_\tau(W_{0.75}) = \beta_0 + \beta_1 W_{0.5} + \beta_2 W_{0.25} + \beta_3 H_{0.75} \]

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.05</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\tau}$</td>
<td>8.04</td>
<td>8.18</td>
<td>8.38</td>
<td>8.74</td>
<td>9.05</td>
<td>9.36</td>
<td>9.55</td>
</tr>
</tbody>
</table>

- **Conclusion:**

His current weight = 9.8 > 9.55 = 0.95th conditional quantile.

\[ \Rightarrow \]

The subject gained an unusual amount of weight at age 0.75 given his prior weights and current height.
Comparison with Least Squares Analysis

- Gaussian model (Cole, 1994)

\[ Z_t = \beta_0 + \beta_1 Z_{t-1} + U_t \quad \text{where } U_t \sim N(0, \sigma^2). \]

where \( Z_t \) is the Box-Cox transformed weight, i.e.,

\[ Z_t = \begin{cases} 
\frac{W_t^{\lambda-1}}{\lambda}, & \lambda \neq 0; \\
\ln(W_t), & \lambda = 0.
\end{cases} \]

- The conditional quantiles

\[ \tilde{Q}_\tau(Z_t) = \hat{\beta}_0 + \hat{\beta}_1 Z_{t-1} + \sigma \Phi^{-1}(\tau). \]

- Marginal normality \( \implies \) Joint normality ?
Inference in quantile regression

1. Direct estimation

- Based on the asymptotic normality of the estimators and the direct estimation of their variance-covariance matrix.
- In an i.i.d. error model,

\[ y_i = x_i^\top \beta + e_i, \quad e_i \sim F. \]

- The variance covariance matrix of \( \hat{\beta}_\tau \) is

\[ \frac{\tau(1 - \tau)}{f(F^{-1}(\tau))^2} \left( X^\top X \right)^{-1}, \]

where \( f(F^{-1}(\tau)) \) is the common error density evaluated at \( F^{-1}(\tau) \).
Inference in quantile regression

1. Direct estimation
   - In a non \(i.i.d\.) error model,
     \[
y_i = x_i^\top \beta + e_i \quad e_i \sim F_i.
     \]
   - The variance covariance matrix of \(\hat{\beta}_\tau\) is
     \[
     (\tau(1 - \tau))(X^\top FX)^{-1}(X^\top X)(X^\top FX)^{-1},
     \]
     where
     \[
     F = diag\{f_1(F_1^{-1}(\tau)), f_2(F_2^{-1}(\tau)), \ldots, f_n(F_n^{-1}(\tau))\},
     \]
     and \(f_i\) is the density function of \(e_i\) evaluated at its \(\tau\)th quantile \(F_i^{-1}(\tau)\).
Inference in quantile regression

2. Rank score method

- Avoid direct estimation of the error densities.
- Introduced by Gutenbrunner, Jurčeková, Koenker, and Portnoy (1993) for an i.i.d error model

\[ y_i = x_i^T \beta + e_i, \quad e_i \text{ i.i.d.} \]

- Extended by Koenker and Machado (1999) to location-scale regression models

\[ y_i = x_i^T \beta + (x_i^T \gamma)e_i. \]

- Generalization of sign tests.
Inference in quantile regression

3. Resampling method

- To avoid direct estimation of variance-covariance matrix.
- Resampling methods
  - Pairwise bootstrap (Efron and Tibshirani, 1998).
  - Bootstrapping the estimating equations (Parzen, Wei and Ying, 1993).
  - Markov chain marginal bootstrap (MCMB) (He and Hu, 2002).
3. Resampling method (cont.)

- Confidence interval based on bootstrap sample
  \( \{ \hat{\beta}^*_1, \ldots, \hat{\beta}^*_m \} \)

  - SD-based
    \[
    [\hat{\beta}_\tau \pm z_{\alpha/2} SD^*(\hat{\beta}^*_\tau)].
    \]

  - Percentile based
    \[
    [2\hat{\beta}_\tau - Q^*_{1-\alpha/2}, \ 2\hat{\beta}_\tau + Q^*_\alpha/2].
    \]
4. Summary

• Direct estimation
  ◦ The asymptotic validity holds generally.
  ◦ The performance in finite samples is sensitive to the choice of smoothing.
  ◦ Computationally efficient.

• Rank score method
  ◦ The asymptotic validity holds for certain models.
  ◦ The performance is stable, and robust against common deviations from the model assumptions.
  ◦ Time consuming for large data sets.
Inference in quantile regression

4. Summary

- Resampling
  - The asymptotic validity holds generally.
  - Could be time consuming for large data sets.
  - MCMB considerably relieves the computation burden by reducing a high-dimension problem to several one-dimensional problems.
5. Practical recommendations

- Use rank score method for relatively small problems, e.g. \( n \leq 1000 \) and \( p \leq 10 \).

- Use MCMB for moderately large problem, e.g. \( 1 \times 10^4 < np < 2 \times 10^6 \).

- Use direct estimation with non \( i.i.d. \) error assumption for very large problems.

- Defaults in SAS PROC QUANTREG are similar.

Inference in quantile regression

6. Computation packages

- **SAS** — specify at PROC QUANTREG Statement

  \[ \text{PROC QUANTREG CI= <NONE|RANK>|...> ALPHA = value} \]

  ◦ By default, the rank score method is used for small data sets \((n < 5000 \text{ and } p < 20)\). Otherwise, the resampling method is used.

- **R** — use function `summary.rq()`.

  \[ \text{summary.rq(object, se = “nid”, ...)} \]

  ◦ By default, the rank score method is used. If \(rq(..., iid = F, )\), then the direct method with non i.i.d. error model is used.
# Inference in quantile regression

## List of inference options in the computation packages

<table>
<thead>
<tr>
<th>Methods</th>
<th>subcategory</th>
<th>Options</th>
<th>R</th>
<th>SAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>$i.i.d.$ model</td>
<td></td>
<td>se = &quot;iid&quot;</td>
<td>CI = SPARCITY/IID</td>
</tr>
<tr>
<td></td>
<td>$n.i.d.$ model</td>
<td></td>
<td>se = &quot;nid&quot;</td>
<td>CI = SPARCITY</td>
</tr>
<tr>
<td>Rank Score</td>
<td></td>
<td></td>
<td>se = &quot;rank&quot;</td>
<td>CI = RANK</td>
</tr>
<tr>
<td>Resampling</td>
<td>Pairwise</td>
<td></td>
<td>se = &quot;boot&quot;, bsmethod=&quot;xy&quot;</td>
<td>Not available</td>
</tr>
<tr>
<td></td>
<td>Parzen, Wei and Ying</td>
<td></td>
<td>se = &quot;boot&quot;, bsmethod=&quot;pxy&quot;</td>
<td>Not available</td>
</tr>
<tr>
<td></td>
<td>MCMB</td>
<td>use R package $rqmemb2$</td>
<td>CI = RESAMPLING</td>
<td></td>
</tr>
</tbody>
</table>
Examples (cont.)

- Birthweight example:
  - Using the MCMB method implemented in SAS, construct confidence band of the estimated coefficients.
  - Comparison of computing times with the other methods.

- Growth chart example:
  - Using the Rank Score method implemented in R, to test the significance of prior weights and current height.
  - Comparison of confidence intervals using different methods.
Comparison of inference methods

• Comparison of computing time for the birthweight example.

List of total computing time for $\tau = 0.5$.

<table>
<thead>
<tr>
<th>Methods</th>
<th>CPU time (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct with $n.i.d.$ model</td>
<td>0.4 0.5</td>
</tr>
<tr>
<td>Rank Score</td>
<td>&gt; 30 &gt; 30</td>
</tr>
<tr>
<td>Resampling with MCMB</td>
<td>3.6 3.0</td>
</tr>
</tbody>
</table>

◦ The CPU time includes both estimation (using interior point algorithm with preprocessing) and confidence interval calculation.

◦ CPU 1.4 GHz and Memory 512 M.
### Growth chart example

\[ Q_\tau(W_{0.75}) = \beta_{\tau0} + \beta_{\tau1}W_{0.5} + \beta_{\tau2}W_{0.25} + \beta_{\tau3}H_{0.75}, \]

| \( \tau \) | \( \beta_{\tau1} \) & 90% CI | \( \beta_{\tau2} \) & 90% CI | \( \beta_{\tau3} \) & 90% CI |
|---------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 0.05    | 0.89 (0.42, 1.41) | 0.07 (-1.20, 0.26) | 0.07 (-0.06, 0.10) |
| 0.10    | 1.00 (0.68, 1.34) | -0.05 (-0.48, 0.20) | 0.05 (-0.01, 0.10) |
| 0.25    | 1.18 (0.73, 1.32) | -0.27 (-0.43, 0.05) | 0.02 (0.00, 0.11) |
| 0.50    | 0.90 (0.77, 1.11) | -0.05 (-0.45, 0.05) | 0.04 (0.03, 0.06) |
| 0.75    | 1.02 (0.94, 1.13) | -0.21 (-0.46, 0.10) | 0.04 (-0.01, 0.09) |
| 0.90    | 1.12 (0.58, 1.30) | -0.34 (-0.39, 0.56) | 0.01 (-0.05, 0.17) |
| 0.95    | 0.62 (0.42, 1.56) | 0.31 (-0.68, 0.66) | 0.08 (-0.10, 0.19) |

* (\( n = 54 \) boys.)
Comparison of inference methods

- Using growth chart example

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\hat{\beta}_{\tau_1}$</th>
<th>Rank method</th>
<th>Direct method, nid</th>
<th>Paired bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>90% CI</td>
<td>90% CI</td>
<td>90% CI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>length</td>
<td>length</td>
<td>length</td>
</tr>
<tr>
<td>0.05</td>
<td>0.89</td>
<td>(0.42,1.41)</td>
<td>0.99</td>
<td>(0.66,1.13)</td>
</tr>
<tr>
<td>0.10</td>
<td>1.00</td>
<td>(0.68,1.34)</td>
<td>0.66</td>
<td>(0.67,1.33)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.19</td>
<td>(0.73,1.32)</td>
<td>0.59</td>
<td>(0.95,1.43)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.90</td>
<td>(0.77,1.11)</td>
<td>0.34</td>
<td>(0.66,1.14)</td>
</tr>
<tr>
<td>0.75</td>
<td>1.02</td>
<td>(0.94,1.13)</td>
<td>0.20</td>
<td>(0.70,1.34)</td>
</tr>
<tr>
<td>0.90</td>
<td>1.12</td>
<td>(0.58,1.30)</td>
<td>0.72</td>
<td>(0.78,1.47)</td>
</tr>
<tr>
<td>0.95</td>
<td>0.62</td>
<td>(0.42,1.56)</td>
<td>1.14</td>
<td>(-0.18,1.42)</td>
</tr>
</tbody>
</table>
Additional topics

• Nonparametric quantile regression model

\[ Q_\tau(Y_t) = g_\tau(t) \]

◦ A simple idea: approximate \( g_\tau(t) \) by a linear combination of spline basis functions, and then use PROC QUANTREG with the basis functions as "covariates".

◦ Reference: He and Shi (1994) and Koenker, Ng and Portnoy (1994).
• Quantile regression with censored data.

The proportional hazard model $h(t, x; \beta) = h_0(t) \exp\{x^T \beta\}$ uses one global coefficient for the effect of each covariate on all quantile functions.

Available software on censored quantile regression:
(a) the censoring time known for all observations; or
(b) right censoring all above one quantile function; or
(c) censoring time independent of covariates.


References


References


References


References


