Introduction to Quantile Regression

CHUNG-MING KUAN

Department of Finance

National Taiwan University

May 31, 2010
1. Introduction

2. Quantile Regression
   - Quantiles
   - Quantile Regression (QR) Method
   - QR Models

3. Algebraic Properties
   - Equivariance
   - Gooness of Fit

4. Asymptotic Properties
   - Heuristics
   - QR Estimator as a GMM Estimator
   - Asymptotic Distribution
Estimation of Asymptotic Covariance Matrix

Hypothesis Testing
- Wald Tests
- Likelihood Ratio Tests

Quantile Treatment Effect
- Digression: Average Treatment Effect
- Quantile Treatment Effect
The behavior of a random variable is governed by its distribution.

**Moment** or summary measures:
- Location measures: mean, median
- Dispersion measures: variance, range
- Other moments: skewness, kurtosis, etc.

**Quantiles**: quartiles, deciles, percentiles

Except in some special cases, a distribution can **not** be completely characterized by its moments or by a few quantiles.

Mean and median characterize the “average” and “center” of $y$ but may provide little info about the tails.
Conventional Methods

For the behavior of $y$ conditional on $x$, consider regression $y_t = x'_t \beta + e_t$.

- Least squares (LS): Legendre (1805)
  - Minimizing $\sum_{t=1}^{T} (y_t - x'_t \beta)^2$ to obtain $\hat{\beta}_T$.
  - $x'_T \hat{\beta}_T$ approximates the conditional mean of $y$ given $x$.

- Least absolute deviation (LAD): Boscovich (1755)
  - Minimizing $\sum_{t=1}^{T} |y_t - x'_t \beta|$ to obtain $\tilde{\beta}_T$.
  - $x'_T \tilde{\beta}_T$ approximates the conditional median of $y$ given $x$.

- Both the LS and LAD methods provide only partial description of the conditional distribution of $y$. 
Mosteller F. and J. Tukey, *Data Analysis and Regression*:

“What the regression curve does is (to) give a grand summary for the averages of the distributions corresponding to the set of xs. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions.”
Quantiles

- The $\theta$th $(0 < \theta < 1)$ quantile of $F_Y$ is

$$q_Y(\theta) := F_Y^{-1}(\theta) = \inf\{y : F_Y(y) \geq \theta\}.$$ 

- $q_Y(\theta)$ is an order statistic, and it can also be obtained by minimizing an asymmetric (linear) loss function:

$$\theta \int_{y > q} |y - q| \, dF_Y(y) + (1 - \theta) \int_{y < q} |y - q| \, dF_Y(y).$$

The first order condition of this minimization problem is

$$0 = -\theta \int_{y > q} dF_Y(y) + (1 - \theta) \int_{y < q} dF_Y(y)$$

$$= -\theta [1 - F_Y(q)] + (1 - \theta) F_Y(q) = -\theta + F_Y(q).$$
Sample Quantiles

- The sample counterpart of the asymmetric linear loss function is

\[
\frac{1}{T} \sum_{t=1}^{T} \rho_\theta(y_t - q) = \frac{1}{T} \left[ \theta \sum_{t:y_t \geq q} |y_t - q| + (1 - \theta) \sum_{t:y_t < q} |y_t - q| \right],
\]

where \( \rho_\theta(u) = (\theta - 1 \{u < 0\})u \) is known as the check function.

- Given \( \theta \), minimizing this function yields the \( \theta \)th sample quantile of \( y \).

- Key point: Other than sorting the data, a sample quantile can also be found via an optimization program.

- Given various \( \theta \) values, we can compute a collection of sample quantiles, from which the distribution can be traced out.
Figure: Check function \( \rho_\theta(u) = (\theta - 1_{u<0})u \).
Quantile Regression (QR) Method

Koenker and Basset (1978)

Given $y_t = x'_t \beta + e_t$, the $\theta$ th QR estimator $\hat{\beta}(\theta)$ minimizes

$$V_T(\beta; \theta) = \frac{1}{T} \sum_{t=1}^{T} \rho_\theta(y_t - x'_t \beta)$$

where $\rho_\theta(e) = (\theta - 1\{e < 0\})e$.

- For $\theta = 0.5$, $V_T$ is symmetric, and $\hat{\beta}(0.5)$ is the LAD estimator.
- $x'_\hat{\beta}(\theta)$ approximates the $\theta$ th conditional quantile function $Q_{y|x}(\theta)$, with $\hat{\beta}_i(\theta)$ the estimated marginal effect of the $i$ th regressor on $Q_{y|x}(\theta)$. 
Finding the Solution to $V_T$

- **Difficulties in estimation:**
  - The QR estimator $\hat{\beta}(\theta)$ does not have a closed form.
  - $V_T$ is not everywhere differentiable, so that standard numerical algorithms do not work.

- A minimizer of $V_T(\beta; \theta)$ is such that the directional derivatives at that point are non-negative in all directions $w$:

$$\frac{d}{d\delta} V_T(\beta + \delta w; \theta) \bigg|_{\delta=0} = -\frac{1}{T} \sum_{t=1}^{T} \psi^*_\theta(y_t - x_t'\beta, -x_t'w)x_t'w,$$

$$\psi^*_\theta(a, b) = \theta - 1_{\{a<0\}} \text{ if } a \neq 0, \quad \psi^*_\theta(a, b) = \theta - 1_{\{b<0\}} \text{ if } a = 0.$$
Let \( \mathbf{b} \) be the point such that \( y_t = \mathbf{x}'_t \mathbf{b} \) for \( t = 1, \ldots, k \). This is a minimizer of \( V_k \) because the directional derivative is

\[
\frac{-1}{k} \sum_{t=1}^{k} \left( \theta - 1_{\{x'_t \mathbf{w} < 0\}} \right) x'_t \mathbf{w},
\]

which must be non-negative for any \( \mathbf{w} \). Thus, \( \mathbf{b} \) a basic solution to the minimization of \( V_T \).

Other basic solutions: \( \mathbf{b}(\kappa) = \mathbf{X}(\kappa)^{-1} \mathbf{y}(\kappa) \), each yielding a perfect fit of \( k \) observations.

The desired estimator \( \hat{\beta}(\theta) \) can be obtained by searching among those basic solutions, for which a linear programming algorithm will do.
- \mathbf{y} = \mathbf{X}\beta + \mathbf{e} \text{ can be expressed as}

\[
\mathbf{y} = \mathbf{X}(\beta^+ - \beta^-) + (\mathbf{e}^+ - \mathbf{e}^-) = \mathbf{A}\mathbf{z},
\]

where \(\mathbf{A} = [\mathbf{X}, -\mathbf{X}, \mathbf{I}_T, -\mathbf{I}_T]\) and \(\mathbf{z} = [\beta^+, \beta^-, \mathbf{e}^+, \mathbf{e}^-]'\), with \(\beta^+\) and \(\beta^-\) the positive and negative parts of \(\beta\), respectively.

- Let \(\mathbf{c} = [0', 0', \theta\mathbf{1}', (1 - \theta)\mathbf{1}']'\). Minimizing \(V_T(\beta; \theta)\) with respect to \(\beta\) is equivalent to the following linear program:

\[
\min_{\mathbf{z}} \frac{1}{T} \mathbf{c}'\mathbf{z}, \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{z}, \mathbf{z} \geq 0.
\]
Some Remarks

- \( \hat{\beta}(\theta) \) is also the QMLE based on an asymmetric Laplace density:

  \[
  f_\theta(e) = \theta(1 - \theta) \exp[-\rho_\theta(e)].
  \]

- Due to linear loss function, \( \hat{\beta}(\theta) \) is more robust to outliers than the LS estimator.

- The estimated \( \theta \)th quantile regression hyperplane must interpolate \( k \) observations in the sample. (Why?)

- QR is not the same as the regressions based on split samples because every quantile regression utilizes all sample data (with different weights). Thus, QR also avoids the sample selection problem arising from sample splitting.
DGP: \( y_t = x_t' \beta_0 + \varepsilon_t = \beta_0 + \tilde{x}_t' \beta_1 + \varepsilon_t \), where \( \varepsilon_t \) are i.i.d. with the common distribution function \( F_\varepsilon \).

- The \( \theta \)-th quantile function of \( y \) is

\[
Q_{y|x}(\theta) = \beta_0 + \tilde{x}' \beta_1 + F_{\varepsilon}^{-1}(\theta),
\]

and hence quantile functions differ only by the “intercept” term and are a vertical displacement of one another.

- The model can also be expressed as

\[
y_t = [\beta_0 + F_{\varepsilon}^{-1}(\theta)] + \tilde{x}_t' \beta_1 + \varepsilon_{t, \theta},
\]

where \( Q_{\varepsilon_\theta|x}(\theta) = 0 \).
DGP: $y_t = x_t' \beta_o + (x_t' \gamma_o) \varepsilon_t$, where $\varepsilon_t$ are i.i.d. with the df $F_\varepsilon$.

- The $\theta$th quantile function of $y$ is

$$Q_{y|x}(\theta) = x_t' \beta_o + (x_t' \gamma_o) F^{-1}_\varepsilon(\theta),$$

and hence quantile functions differ not only by the “intercept” but also the “slope” term.

- The model can also be expressed as

$$y_t = x_t' [\beta_o + \gamma_o F^{-1}_\varepsilon(\theta)] + \varepsilon_{t,\theta},$$

where $Q_{\varepsilon\theta|x}(\theta) = 0$.

- The QR estimator $\hat{\beta}(\theta)$ converges to $\beta(\theta)$, and $x_t' \hat{\beta}(\theta)$ approximates the $\theta$th quantile function of $y$ given $x$, $Q_{y|x}(\theta)$. 
Let $\hat{\beta}(\theta)$ be the qR estimator of the quantile regression of $y_t$ on $x_t$.

- **Scale equivariance**: For $y^*_t = cy_t$, let $\hat{\beta}^*(\theta)$ be the QR estimator of the quantile regression of $y^*_t$ on $x_t$.
  - For $c > 0$, $\hat{\beta}^*(\theta) = c \hat{\beta}(\theta)$.
  - For $c < 0$, $\hat{\beta}^*(1 - \theta) = c \hat{\beta}(\theta)$.
  - $\hat{\beta}^*(0.5) = c \hat{\beta}(0.5)$, regardless of the sign of $c$.

- **Shift equivariance**: For $y^*_t = y_t + x'_t \gamma$, let $\hat{\beta}^*(\theta)$ be the QR estimator of the quantile regression of $y^*_t$ on $x_t$. Then, $\hat{\beta}^*(\theta) = \hat{\beta}(\theta) + \gamma$.

- **Equivariance to reparameterization of design**: Given $X^* = XA$ for some nonsingular $A$, $\hat{\beta}^*(\theta) = A^{-1}\hat{\beta}(\theta)$. 
• **Equivariance to monotonic transformations:** For a nondecreasing function $h$,

$$\mathbb{I} \{ y \leq a \} = \mathbb{I} \{ h(y) \leq h(a) \},$$

so that

$$Q_{h(y) | \theta} = h(Q_{y | \theta}).$$

Note that the expectation operator does not have this property because $\mathbb{E}[h(y)] \neq h(\mathbb{E}(y))$ in general, except when $h$ is linear.

• **Example:** If $\mathbf{x}' \beta$ is the $\theta$th conditional quantile of $\ln y$, then $\exp(\mathbf{x}' \beta)$ is the $\theta$th conditional quantile of $y$. 
Goodness of Fit

Specification: \( y_t = x_{1t}\beta_1 + x_{2t}\beta_2 + e_t \).

- A measure of the relative contribution of additional regressors \( x_{2t} \) is
  \[
  1 - \frac{V_T(\hat{\beta}_1(\theta), \hat{\beta}_2(\theta); \theta)}{V_T(\tilde{\beta}_1(\theta), 0; \theta)},
  \]
  where \( V_T(\tilde{\beta}_1(\theta), 0; \theta) \) is computed under the constraint \( \beta_2 = 0 \).

- A measure of the goodness-of-fit of a specification is thus
  \[
  R^1(\theta) = 1 - \frac{V_T(\hat{\beta}(\theta); \theta)}{V_T(\hat{q}(\theta), 0; \theta)}.
  \]
  where \( \hat{q}(\theta) \) is the sample quantile and \( V_T(\hat{q}(\theta), 0; \theta) \) is obtained from
  the model with the constant term only. Clearly, \( 0 < R^1(\theta) < 1 \).
Asymptotic Properties: Heuristics

- Ignoring \( y_t = q \), the “FOC” of minimizing \( T^{-1} \sum_{t=1}^{T} \rho_\theta(y_t - q) \) is

\[
g_T(q) := \frac{1}{T} \sum_{t=1}^{T} (1_{\{y_t < q\}} - \theta).
\]

- Clearly, \( g_T(q) \) is non-decreasing in \( q \) (why?), so that \( \hat{q}(\theta) > q \) iff \( g_T(q) < 0 \). Thus,

\[
\text{IP} \left[ \sqrt{T}(\hat{q}(\theta) - q(\theta)) > c \right] = \text{IP} \left[ g_T(q(\theta) + c/\sqrt{T}) < 0 \right].
\]

We have

\[
\mathbb{E} \left[ g_T(q(\theta) + c/\sqrt{T}) \right] = F \left( q(\theta) + \frac{c}{\sqrt{T}} \right) - \theta \approx f(q(\theta)) \frac{c}{\sqrt{T}}
\]

\[
\text{var} \left[ g_T(q(\theta) + c/\sqrt{T}) \right] = \frac{1}{T} F(1 - F) \approx \frac{1}{T} \theta(1 - \theta).
\]
Setting $\lambda^2 = \theta(1 - \theta)/f^2(q(\theta))$,

$$\text{IP} \left[ \sqrt{T}(\hat{q}(\theta) - q(\theta)) > c \right]$$

$$= \text{IP} \left[ \frac{g_T(q(\theta) + c/\sqrt{T})}{\sqrt{\theta(1 - \theta)/T}} < 0 \right]$$

$$= \text{IP} \left[ \frac{g_T(q(\theta) + c/\sqrt{T})}{\sqrt{\theta(1 - \theta)/T}} - \frac{c}{\lambda} < -\frac{c}{\lambda} \right]$$

$$= \text{IP} \left[ \frac{g_T(q(\theta) + c/\sqrt{T}) - f(q(\theta))c/\sqrt{T}}{\sqrt{\theta(1 - \theta)/T}} < -\frac{c}{\lambda} \right]$$

$$\rightarrow 1 - \Phi(c/\lambda),$$

by a CLT. This implies

$$\sqrt{T}(\hat{q}(\theta) - q(\theta)) \overset{D}{\rightarrow} \mathcal{N}(0, \lambda^2).$$
Given $q$ moment conditions $\mathbb{E}[m(w_t; \beta_0)] = 0$, $\beta_0$ ($k \times 1$) is exactly identified if $q = k$ and over-identified if $q > k$. When $\beta_0$ is exactly identified, the GMM estimator $\hat{\beta}$ of $\beta_0$ solves $T^{-1} \sum_{t=1}^{T} m(w_t; \beta) = 0$.

Asymptotic Distribution of the GMM Estimator

Given the GMM estimator $\hat{\beta}$ of $\beta_0$,

$$\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{A} \mathcal{N}(0, G_o^{-1} \Sigma_o G_o^{-1}),$$

with $\Sigma_o = \mathbb{E}[m(w_t; \beta_0)m(w_t; \beta_0)']$, and

$$\frac{1}{T} \sum_{t=1}^{T} \nabla_{\beta} m(w_t; \beta_0) \xrightarrow{P} G_o := \mathbb{E}[\nabla_{\beta} m(w_t; \beta_0)].$$
QR Estimator as a GMM Estimator

- The QR estimator $\hat{\beta}(\theta)$ satisfies the “asymptotic FOC”:
  \[
  \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varphi_{\theta}(y_t - x_t' \hat{\beta}(\theta)) := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t (\theta - 1_{\{y_t - x_t' \hat{\beta}(\theta) < 0\}}) = o_P(1).
  \]

- The (approximate) estimating function is thus
  \[
  \frac{1}{T} \sum_{t=1}^{T} x_t (\theta - 1_{\{y_t - x_t' \beta < 0\}}).
  \]

- The expectation of the estimating function is
  \[
  \mathbb{E}\{x_t [\theta - \mathbb{E}(1_{\{y_t - x_t' \beta < 0\}} | x_t)]\} = \mathbb{E}\{x_t [\theta - F_{y|x}(x_t' \beta)]\}.
  \]

- When $\beta$ is evaluated at $\beta(\theta)$, $F_{y|x}(x_t' \beta)$ must be $\theta$ so that the moment conditions are $\mathbb{E}[\varphi_{\theta}(y_t - x_t' \beta(\theta))] = 0$. 

When integration and differentiation can be interchanged,

\[ G(\beta) = IE \left[ \nabla_\beta \varphi_\theta(y_t - x'_t \beta) \right] \]

\[ = \nabla_\beta IE \left\{ x_t \left[ \theta - F_{y|X}(x'_t \beta) \right] \right\} = -IE \left[ x_t x'_t f_{e_\theta|x}(x'_t \beta) \right]. \]

Then, \( G(\beta(\theta)) = -IE \left[ x_t x'_t f_{e_\theta|x}(0) \right]. \)

\[ 1_{\{y_t - x'_t \beta(\theta) < 0\}} \] is Bernoulli with mean \( \theta \) and variance \( \theta(1 - \theta) \), so that

\[ \Sigma(\beta) = IE \left( x_t x'_t IE \left[ (\theta - 1_{\{y_t - x'_t \beta < 0\}})^2 \mid x_t \right] \right). \]

Then, \( \Sigma(\beta(\theta)) = \theta(1 - \theta) IE(x_t x'_t) =: \theta(1 - \theta)M_{xx}. \)
Asymptotic Normality of the QR Estimator

\[
\sqrt{T} [\hat{\beta}(\theta) - \beta(\theta)] \xrightarrow{D} \mathcal{N} \left(0, \theta(1 - \theta) G(\beta(\theta))^{-1} M_{xx} G(\beta(\theta))^{-1} \right),
\]

where \( M_{xx} = \mathbb{E}(x_t x_t') \) and \( G(\beta(\theta)) = -\mathbb{E} \left[ x_t x_t' f_{e_\theta|x(0)} \right]. \)

- Conditional heterogeneity is characterized by the conditional density \( f_{e_\theta|x(0)} \) in \( G(\beta(\theta)) \), which is not limited to heteroskedasticity.
- If \( f_{e_\theta|x(0)} = f_{e_\theta}(0) \), i.e., conditional homogeneity,

\[
\sqrt{T} [\hat{\beta}(\theta) - \beta(\theta)] \xrightarrow{D} \mathcal{N} \left(0, \frac{\theta(1 - \theta)}{[f_{e_\theta}(0)]^2} M_{xx}^{-1} \right).
\]
Estimation of Asymptotic Covariance Matrix

Consistent estimation of $D(\beta(\theta)) = G(\beta(\theta))^{-1} M_{xx} G(\beta(\theta))^{-1}$.

- Estimation of $M_{xx}: M_T = T^{-1} \sum_{t=1}^{T} x_t x'_t$.
- Digression: Differentiating both sides of $F(F^{-1}(\theta)) = \theta$:
  \[
  \frac{d F^{-1}(\theta)}{d \theta} = \frac{1}{f(F^{-1}(\theta))} =: s(\theta),
  \]
differentiating a quantile function yields a sparsity function.

- Estimating the sparsity function:
  - Using a difference quotient of empirical quantiles $\hat{F}_T^{-1}(\theta)$:
    \[
    \hat{s}_T(\theta) = \left[ \hat{F}_T^{-1}(\theta + h_T) - \hat{F}_T^{-1}(\theta - h_T) \right] / (2h_T).
    \]
  - Letting $\hat{e}_{(i)}$ be the $i$th order statistic of QR residuals $\hat{e}_t$,
    \[
    \hat{F}_T^{-1}(\tau) = \hat{e}_{(i)}, \quad \tau \in \left[ (i-1)/T, i/T \right).
    \]
Hendricks and Koenker (1991): Estimating \( f_{e(\theta)|x}(0) \) in \( G(\beta(\theta)) \) by

\[
\hat{f}_t = \frac{2h_T}{x'_t[\hat{\beta}(\theta + h_T) - \hat{\beta}(\theta - h_T)]},
\]

and estimating \(-G\) by \(-\hat{G}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t x_t x'_t\).

Powell (1991): Estimating \(-G(\beta(\theta))\) by

\[
-\hat{G}_T = \frac{1}{2Tc_T} \sum_{t=1}^{T} 1_{\{|\hat{e}_t(\theta)| < c_T\}} x_t x'_t,
\]

where \(c_T \to 0\) and \(T^{1/2}c_T \to \infty\) as \(T \to \infty\).

STATA: Bootstrap
Standard Wald Test

\( H_0: \ R\beta(\theta) = r, \) where \( R \) is \( q \times k \) and \( r \) is \( q \times 1 \).

- \( \sqrt{T}[\hat{\beta}(\theta) - \beta(\theta)] \overset{D}{\rightarrow} \mathcal{N}(0, \theta(1 - \theta)D(\beta(\theta))) \).

- Under the null,

\[
\sqrt{T}R(\hat{\beta}(\theta) - \beta(\theta)) = \sqrt{T}(R\hat{\beta}(\theta) - r) \overset{D}{\rightarrow} \mathcal{N}(0, \theta(1 - \theta)\Gamma(\beta(\theta))),
\]

where \( \Gamma(\beta(\theta)) = RD(\beta(\theta))R' \).

The Null Distribution of the Wald Test

\( \mathcal{W}_T(\theta) = T[R\hat{\beta}(\theta) - r]'\hat{\Gamma}(\theta)^{-1}[R\hat{\beta}(\theta) - r]/[\theta(1 - \theta)] \overset{D}{\rightarrow} \chi^2(q), \)

where \( \hat{\Gamma}(\theta) = R\hat{D}(\theta)R' \), with \( \hat{D}(\theta) \) a consistent estimator of \( D(\beta(\theta)) \).
H₀: \( R\beta(\theta) = r \) for all \( \theta \in S \subset (0, 1) \).

The Brownian bridge: \( B_q(\theta) \overset{d}{=} [\theta(1 - \theta)]^{1/2}N(0, I_q) \), and hence

\[
\hat{\Gamma}(\theta)^{-1/2}\sqrt{T}[R\hat{\beta}(\theta) - r] \xrightarrow{D} B_q(\theta).
\]

Thus,

\[
\mathcal{W}_T(\theta) \xrightarrow{D} \left\| B_q(\theta)/\sqrt{\theta(1-\theta)} \right\|^2,
\]

uniformly in \( \theta \).

The Null Distribution of the Sup-Wald Test

\[
\sup_{\theta \in S} \mathcal{W}_T(\theta) \Rightarrow \sup_{\theta \in S} \left\| \frac{B_q(\theta)}{\sqrt{\theta(1-\theta)}} \right\|^2,
\]

where \( S \) is a compact set in \((0, 1)\).
To test $R\beta(\theta) = r$, $\theta \in [a, b]$, set $a = \theta_1 < \ldots < \theta_n = b$ and compute

$$\sup -\mathcal{W}_T = \sup_{i=1,\ldots,n} \mathcal{W}_T(\theta_i).$$

Koenker and Machado (1999): $[a, b] = [\epsilon, 1 - \epsilon]$ with $\epsilon$ small.

For $s = \theta/(1 - \theta)$, $B(\theta)/\sqrt{\theta(1 - \theta)} \overset{d}{=} W(s)/\sqrt{s}$, so that

$$\mathbb{P}\left\{ \sup_{\theta \in [a,b]} \left\| \frac{B_q(\theta)}{\sqrt{\theta(1 - \theta)}} \right\|^2 < c \right\} = \mathbb{P}\left\{ \sup_{s \in [1, s_2/s_1]} \left\| \frac{W_q(s)}{\sqrt{s}} \right\|^2 < c \right\},$$

with $s_1 = a/(1 - a)$, $s_2 = b/(1 - b)$.

Some critical values were tabulated in DeLong (1981) and Andrews (1993); the other can be obtained via simulations.
Likelihood Ratio Tests

- Let $\hat{\beta}(\theta)$ and $\tilde{\beta}(\theta)$ be the constrained and unconstrained estimators and $\hat{V}_T(\theta) = V_T(\hat{\beta}(\theta); \theta)$ and $\tilde{V}_T(\theta) = V_T(\tilde{\beta}(\theta); \theta)$ be the corresponding objective functions.

- Given the asymmetric Laplace density: $f_\theta(u) = \theta(1 - \theta) \exp[-\rho_\theta(u)]$, the log-likelihood is

$$L_T(\beta; \theta) = T \log(\theta(1 - \theta)) - \sum_{t=1}^{T} \rho_\theta(y_t - x'_t \beta).$$

- $-2$ times the log-likelihood ratio is

$$2 \left[ L_T(\hat{\beta}(\theta); \theta) - L_T(\tilde{\beta}(\theta); \theta) \right] = 2 \left[ \tilde{V}_T(\theta) - \hat{V}_T(\theta) \right].$$
Koenker and Machado (1999):

\[
\mathcal{L}\mathcal{R}_T(\theta) = \frac{2 [\tilde{V}_T(\theta) - \hat{V}_T(\theta)]}{\theta(1 - \theta)[f_{e \theta}(0)]^{-1}} \xrightarrow{D} \chi^2(q).
\]

This test is also known as the quantile \( \rho \) test.

Koenker and Bassett (1982): For median regression,

\[
\mathcal{L}\mathcal{R}_T(0.5) = \frac{8 [\tilde{V}_T(0.5) - \hat{V}_T(0.5)]}{[f_{e_{0.5}}(0)]^{-1}} = 2 [\tilde{V}_T(0.5) - \hat{V}_T(0.5)],
\]

because \( f_{e_{0.5}}(0) = 1/4 \).
Digression: Average Treatment Effect

- Evaluating the impact of a treatment (program, policy, intervention).
- Let $D$ be the binary indicator of treatment and $X$ be covariates.
  - $Y_1 (Y_0)$ is the potential outcome when an agent is (is not) exposed to the treatment.
  - The observed outcome is $Y = DY_1 + (1 - D)Y_0$.
- We observe only one potential outcome ($Y_{1i}$ or $Y_{0i}$) and hence can not identify the individual treatment effect, $Y_{1i} - Y_{0i}$. We may estimate the average treatment effect (ATE): $\mathbb{E}(Y_1 - Y_0)$.
- Under conditional independence: $(Y_1, Y_0) \perp D \mid X$,
  \[ \mathbb{E}(Y \mid D = 1, X) - \mathbb{E}(Y \mid D = 0, X) = \mathbb{E}(Y_1 - Y_0 \mid X), \]
  so that the ATE is $\mathbb{E}(Y_1 - Y_0) = \mathbb{E}[\mathbb{E}(Y_1 - Y_0 \mid X)]$. 
- Using the sample counterpart of $\mathbb{E}(Y|D = 1, X) - \mathbb{E}(Y|D = 0, X)$ we have
  
  $$\hat{ATE} = \frac{1}{N} \sum_{i=1}^{N} [\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)].$$

- For the dummy-variable regression:

  $$Y_i = \alpha + D_i \gamma + X_i' \beta + \mu_D + e_i, \quad i = 1, \ldots, n,$$

  the LS estimate of $\gamma$ is $\hat{ATE}$.

- Other estimators: Kernel matching, nearest neighbor matching, propensity score matching (based on $p(x) = \mathbb{P}(D = 1|X = x)$), etc.
Quantile Treatment Effect

Let $F_0$ and $F_1$ be, resp., the distributions of control and treatment responses. Let $\Delta(\eta)$ be the “horizontal shift” from $F_0$ to $F_1$:

$$F_0(\eta) = F_1(\eta + \Delta(\eta)).$$

Then, $\Delta(\eta) = F_1^{-1}(F_0(\eta)) - \eta$, and the $\theta$ th quantile treatment effect (QTE) is, for $F_0(\eta) = \theta$,

$$\text{QTE}(\theta) = F_1^{-1}(\theta) - F_0^{-1}(\theta) = q_{Y_1}(\theta) - q_{Y_0}(\theta),$$

the difference between the quantiles of two distributions.

We may apply the QR method to

$$Y_i = \alpha + D_i \gamma + X_i' \beta + e_i,$$

the resulting QR estimate $\hat{\gamma}(\theta)$ is the estimated $\theta$ th QTE.

Other: A weighting estimator based on the propensity score.
The impact of a program (policy) may be observed after a certain period of time. To identify the “true” treatment effect, the potential change due to time (other factors) must be excluded first.

Define the following dummy variables:

(i) $D_{i,\tau} = 1$ if the $i$th individual receives the treatment;
(ii) $D_{i,a} = 1$ if the $i$th individual is in the post-program period;
(iii) $D_{i,a\tau} = D_{i,\tau} \times D_{i,a}$.

Model: $Y_i = \alpha + \alpha_1 D_{i,\tau} + \alpha_2 D_{i,a} + \alpha_3 D_{i,a\tau} + X_i'\beta + e_i$.

- For the treatment group in pre- and post-program periods, the time effect is $\alpha_2 + \alpha_3$.
- For the control group in pre- and post-program periods, the time effect is $\alpha_2$.
- The treatment effect is the difference between these two effects: $\alpha_3$. 