Introduction to Quantile Regression

CHUNG-MING KUAN

Department of Finance

National Taiwan University

May 31, 2010

Lecture Outline

Lecture Outline (cont'd)

5. Estimation of Asymptotic Covariance Matrix

6. Hypothesis Testing
   - Wald Tests
   - Likelihood Ratio Tests

7. Quantile Treatment Effect
   - Digression: Average Treatment Effect
   - Quantile Treatment Effect

Introduction

- 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 qunatiles.
- 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 $\mathbf{x}$, consider regression $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$.

- Least squares (LS): Legendre (1805)
  - Minimizing $\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2$ to obtain $\hat{\boldsymbol{\beta}}_T$.
  - $\mathbf{x}'\hat{\boldsymbol{\beta}}_T$ approximates the conditional mean of y given $\mathbf{x}$.
- Least absolute deviation (LAD): Boscovich (1755)
  - Minimizing $\sum_{t=1}^{T}|y_t - \mathbf{x}_t'\boldsymbol{\beta}|$ to obtain $\check{\boldsymbol{\beta}}_T$.
  - $\mathbf{x}'\check{\boldsymbol{\beta}}_T$ approximates the conditional median of y given $\mathbf{x}$.
- 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 θth $(0 < \theta < 1)$ 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:

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

The first order condition of this minimization problem is

$$\begin{aligned} 0 &= -\theta \int_{y>q} \mathrm{d}F_Y(y) + (1-\theta) \int_{y<q} \mathrm{d}F_Y(y) \\ &= -\theta[1 - F_Y(q)] + (1-\theta)F_Y(q) = -\theta + F_Y(q). \end{aligned}$$

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 θ, minimizing this function yields the θ 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 θ values, we can compute a collection of sample quantiles, from which the distribution can be traced out.

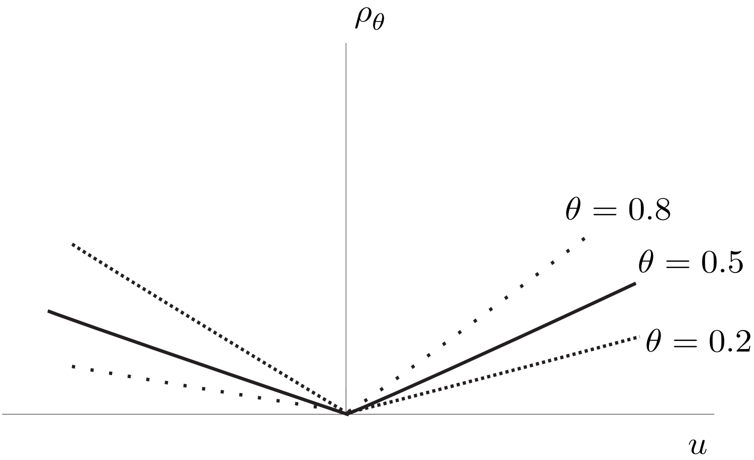

Figure: Check function $\rho_\theta(u) = (\theta - 1_{\{u<0\}})u$.

Quantile Regression (QR) Method

Koenker and Basset (1978)

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

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

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

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

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 $V_T(\boldsymbol{\beta};\theta)$ is such that the directional derivatives at that point are non-negative in all directions $\mathbf{w}$:

$$\frac{\mathrm{d}}{\mathrm{d}\,\delta} V_T(\boldsymbol{\beta} + \delta \mathbf{w}; \theta)\Big|_{\delta=0} = \frac{-1}{T} \sum_{t=1}^{T} \psi_\theta^*(y_t - \mathbf{x}_t' \boldsymbol{\beta}, \; -\mathbf{x}_t' \mathbf{w}) \mathbf{x}_t' \mathbf{w},$$

$\psi_\theta^*(a, b) = \theta - \mathbf{1}_{\{a<0\}}$ if $a \neq 0$, $\psi_\theta^*(a, b) = \theta - \mathbf{1}_{\{b<0\}}$ 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

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

  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.

Linear Programming

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ can be expressed as

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

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

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

$$\min_{\mathbf{z}} \frac{1}{T}\mathbf{c}'\mathbf{z}, \qquad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{z}, \ \ \mathbf{z} \geq 0.$$

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 θ 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.

QR: Location Shift Model

DGP: $y_t = \mathbf{x}_t'\boldsymbol{\beta}_o + \varepsilon_t = \beta_0 + \tilde{\mathbf{x}}_t'\boldsymbol{\beta}_1 + \varepsilon_t$, where ε_t are i.i.d. with the common distribution function F_ε.

- The θ-th quantile function of y is

$$Q_{y|\mathbf{x}}(\theta) = \beta_0 + \tilde{\mathbf{x}}'\boldsymbol{\beta}_1 + F_\varepsilon^{-1}(\theta),$$

  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

$$y_t = \underbrace{[\beta_0 + F_\varepsilon^{-1}(\theta)]}_{\beta_0(\theta)} + \tilde{\mathbf{x}}_t'\boldsymbol{\beta}_1 + \varepsilon_{t,\theta},$$

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

QR: Location-Scale Shift Model

DGP: $y_t = \mathbf{x}_t'\boldsymbol{\beta}_o + (\mathbf{x}_t'\boldsymbol{\gamma}_o)\varepsilon_t$, where ε_t are i.i.d. with the df F_ε.

- The θ th quantile function of y is

$$Q_{y|\mathbf{x}}(\theta) = \mathbf{x}_t'\boldsymbol{\beta}_o + (\mathbf{x}_t'\boldsymbol{\gamma}_o)F_\varepsilon^{-1}(\theta),$$

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

- The model can also be expressed as

$$y_t = \mathbf{x}_t'[\underbrace{\boldsymbol{\beta}_o + \boldsymbol{\gamma}_o F_\varepsilon^{-1}(\theta)}_{\boldsymbol{\beta}(\theta)}] + \varepsilon_{t,\theta},$$

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

- The QR estimator $\hat{\boldsymbol{\beta}}(\theta)$ converges to $\boldsymbol{\beta}(\theta)$, and $\mathbf{x}'\hat{\boldsymbol{\beta}}(\theta)$ approximates the θ th quantile function of y given $\mathbf{x}$, $Q_{y|\mathbf{x}}(\theta)$.

Algebraic Properties: Equivariance

Let $\hat{\beta}(\theta)$ be the qR estimator of the quantile regression of y_t on $\mathbf{x}_t$.

- Scale equivariance: For $y_t^* = c\, y_t$, let $\hat{\beta}^*(\theta)$ be the QR estimator of the quantile regression of y_t^* on $\mathbf{x}_t$.
  - 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 $y_t^* = y_t + \mathbf{x}_t'\gamma$, let $\hat{\beta}^*(\theta)$ be the QR estimator of the quantile regression of y_t^* on $\mathbf{x}_t$. Then, $\hat{\beta}^*(\theta) = \hat{\beta}(\theta) + \gamma$.
- Equivariance to reparameterization of design: Given $\mathbf{X}^* = \mathbf{X}\mathbf{A}$ for some nonsingular $\mathbf{A}$, $\hat{\beta}^*(\theta) = \mathbf{A}^{-1}\hat{\beta}(\theta)$.

- Equivariance to monotonic transformations: For a nondecreasing function h,

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

so that

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

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 $\mathbf{x}'\boldsymbol{\beta}$ is the θ th conditional quantile of $\ln y$, then $\exp(\mathbf{x}'\boldsymbol{\beta})$ is the θ th conditional quantile of y.

Goodness of Fit

Specification: $y_t = \mathbf{x}_{1t}\boldsymbol{\beta}_1 + \mathbf{x}_{2t}\boldsymbol{\beta}_2 + e_t$.

- A measure of the relative contribution of additional regressors $\mathbf{x}_{2t}$ is

$$1 - \frac{V_T\big(\hat{\boldsymbol{\beta}}_1(\theta), \hat{\boldsymbol{\beta}}_2(\theta); \theta\big)}{V_T\big(\tilde{\boldsymbol{\beta}}_1(\theta), \mathbf{0}; \theta\big)},$$

  where $V_T\big(\tilde{\boldsymbol{\beta}}_1(\theta), \mathbf{0}; \theta\big)$ is computed under the constraint $\boldsymbol{\beta}_2 = \mathbf{0}$.

- A measure of the goodness-of-fit of a specification is thus

$$R^1(\theta) = 1 - \frac{V_T\big(\hat{\boldsymbol{\beta}}(\theta); \theta\big)}{V_T\big(\hat{q}(\theta), \mathbf{0}; \theta\big)}.$$

  where $\hat{q}(\theta)$ is the sample quantile and $V_T\big(\hat{q}(\theta), \mathbf{0}; \theta\big)$ 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

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

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

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

We have

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

$$\text{var}\left[g_T\left(q(\theta) + \frac{c}{\sqrt{T}}\right)\right] = \frac{1}{T}F(1-F) \approx \frac{1}{T}\theta(1-\theta).$$

- Setting $\lambda^2 = \theta(1-\theta)/f^2(q(\theta))$,

$$\begin{aligned}
&\mathbb{P}\left[\sqrt{T}\big(\hat{q}(\theta) - q(\theta)\big) > c\right] \\
&\quad = \mathbb{P}\left[\frac{g_T\big(q(\theta) + c/\sqrt{T}\big)}{\sqrt{\theta(1-\theta)/T}} < 0\right] \\
&\quad = \mathbb{P}\left[\frac{g_T\big(q(\theta) + c/\sqrt{T}\big)}{\sqrt{\theta(1-\theta)/T}} - \frac{c}{\lambda} < -\frac{c}{\lambda}\right] \\
&\quad = \mathbb{P}\left[\frac{g_T\big(q(\theta) + c/\sqrt{T}\big) - f(q(\theta))c/\sqrt{T}}{\sqrt{\theta(1-\theta)/T}} < -\frac{c}{\lambda}\right] \\
&\quad \to 1 - \Phi(c/\lambda),
\end{aligned}$$

by a CLT. This implies

$$\sqrt{T}\big(\hat{q}(\theta) - q(\theta)\big) \xrightarrow{D} \mathcal{N}(0, \lambda^2).$$

GMM Estimation

Given q moment conditions $\mathbb{E}[\mathbf{m}(\mathbf{w}_t; \boldsymbol{\beta}_o)] = \mathbf{0}$, $\boldsymbol{\beta}_o$ $(k \times 1)$ is exactly identified if $q = k$ and over-identified if $q > k$. When $\boldsymbol{\beta}_o$ is exactly identified, the GMM estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_o$ solves $T^{-1} \sum_{t=1}^{T} \mathbf{m}(\mathbf{w}_t; \boldsymbol{\beta}) = \mathbf{0}$.

Asymptotic Distribution of the GMM Estimator

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

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \overset{A}{\sim} \mathcal{N}\big(\mathbf{0},\ \mathbf{G}_o^{-1} \boldsymbol{\Sigma}_o \mathbf{G}_o^{-1}\big),$$

with $\boldsymbol{\Sigma}_o = \mathbb{E}[\mathbf{m}(\mathbf{w}_t; \boldsymbol{\beta}_o)\mathbf{m}(\mathbf{w}_t; \boldsymbol{\beta}_o)']$, and

$$\frac{1}{T} \sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}} \mathbf{m}(\mathbf{w}_t; \boldsymbol{\beta}_o) \xrightarrow{\mathbb{P}} \mathbf{G}_o := \mathbb{E}[\nabla_{\boldsymbol{\beta}} \mathbf{m}(\mathbf{w}_t; \boldsymbol{\beta}_o)].$$

QR Estimator as a GMM Estimator

- The QR estimator $\hat{\boldsymbol{\beta}}(\theta)$ satisfies the "asymptotic FOC":

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varphi_\theta(y_t-\mathbf{x}_t'\hat{\boldsymbol{\beta}}(\theta)) := \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\big(\theta-\mathbf{1}_{\{y_t-\mathbf{x}_t'\hat{\boldsymbol{\beta}}(\theta)<0\}}\big) = o_{\mathbb{P}}(1).$$

- The (approximate) estimating function is thus

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\big(\theta-\mathbf{1}_{\{y_t-\mathbf{x}_t'\boldsymbol{\beta}<0\}}\big).$$

- The expectation of the estimating function is

$$\mathbb{E}\left\{\mathbf{x}_t\left[\theta-\mathbb{E}\left(\mathbf{1}_{\{y_t-\mathbf{x}_t'\boldsymbol{\beta}<0\}}\mid\mathbf{x}_t\right)\right]\right\} = \mathbb{E}\left\{\mathbf{x}_t\left[\theta-F_{y|x}(\mathbf{x}_t'\boldsymbol{\beta})\right]\right\}.$$

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

Asymptotic Distribution

- When integration and differentiation can be interchanged,

$$\begin{aligned}\mathbf{G}(\boldsymbol{\beta}) &= \mathbb{E}\left[\nabla_{\boldsymbol{\beta}}\varphi_\theta(y_t - \mathbf{x}_t'\boldsymbol{\beta})\right] \\ &= \nabla_{\boldsymbol{\beta}}\,\mathbb{E}\left\{\mathbf{x}_t\left[\theta - F_{y|x}(\mathbf{x}_t'\boldsymbol{\beta})\right]\right\} = -\,\mathbb{E}\left[\mathbf{x}_t\mathbf{x}_t' f_{y|x}(\mathbf{x}_t'\boldsymbol{\beta})\right].\end{aligned}$$

  Then, $\mathbf{G}(\boldsymbol{\beta}(\theta)) = -\,\mathbb{E}\left[\mathbf{x}_t\mathbf{x}_t' f_{e_\theta|x}(0)\right]$.

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

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}) = \mathbb{E}\left(\mathbf{x}_t\mathbf{x}_t'\,\mathbb{E}\left[\left(\theta - \mathbf{1}_{\{y_t - \mathbf{x}_t'\boldsymbol{\beta} < 0\}}\right)^2 \mid \mathbf{x}_t\right]\right).$$

  Then, $\boldsymbol{\Sigma}(\boldsymbol{\beta}(\theta)) = \theta(1-\theta)\,\mathbb{E}(\mathbf{x}_t\mathbf{x}_t') =: \theta(1-\theta)\mathbf{M}_{xx}$.

Asymptotic Normality of the QR Estimator

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

where $\mathbf{M}_{xx} = \mathbb{E}(\mathbf{x}_t\mathbf{x}_t')$ and $\mathbf{G}(\beta(\theta)) = -\,\mathbb{E}\left[\mathbf{x}_t\mathbf{x}_t' f_{e_\theta|\mathbf{x}}(0)\right]$.

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

$$\sqrt{T}\left[\hat{\beta}(\theta) - \beta(\theta)\right] \xrightarrow{D} \mathcal{N}\left(\mathbf{0},\ \frac{\theta(1-\theta)}{[f_{e_\theta}(0)]^2}\mathbf{M}_{xx}^{-1}\right).$$

Estimation of Asymptotic Covariance Matrix

Consistent estimation of $\mathbf{D}(\beta(\theta)) = \mathbf{G}\big(\beta(\theta)\big)^{-1}\mathbf{M}_{xx}\mathbf{G}\big(\beta(\theta)\big)^{-1}$.

- Estimation of $\mathbf{M}_{xx}$: $\mathbf{M}_T = T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$.
- Digression: Differentiating both sides of $F(F^{-1}(\theta)) = \theta$:

$$\frac{\mathrm{d}\,F^{-1}(\theta)}{\mathrm{d}\theta} = \frac{1}{f(F^{-1}(\theta))} =: s(\theta),$$

  differentiating a quantile function yields a sparsity function.
- Estimating the sparsity function:
  - Using a difference quotient of empirical quantiles $\widehat{F}_T^{-1}(\theta)$:

$$\hat{s}_T(\theta) = \big[\widehat{F}_T^{-1}(\theta + h_T) - \widehat{F}_T^{-1}(\theta - h_T)\big]/(2h_T).$$

  - Letting $\hat{e}_{(i)}$ be the i^{th} order statistic of QR residuals $\hat{e}_t$,

$$\widehat{F}_T^{-1}(\tau) = \hat{e}_{(i)}, \quad \tau \in [(i-1)/T, i/T).$$

- Hendricks and Koenker (1991): Estimating $f_{e(\theta)|x}(0)$ in $\mathbf{G}(\beta(\theta))$ by

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

  and estimating $-\mathbf{G}$ by $-\widehat{\mathbf{G}}_T = \frac{1}{T}\sum_{t=1}^{T} \hat{f}_t \mathbf{x}_t \mathbf{x}_t'$.

- Powell (1991): Estimating $-\mathbf{G}(\beta(\theta))$ by

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

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

- STATA: Bootstrap

Standard Wald Test

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

- $\sqrt{T}\left[\hat{\beta}(\theta) - \beta(\theta)\right] \xrightarrow{D} \mathcal{N}\big(\mathbf{0},\, \theta(1-\theta)\mathbf{D}(\beta(\theta))\big)$.
- Under the null,

$$\sqrt{T}\mathbf{R}\big(\hat{\beta}(\theta) - \beta(\theta)\big) = \sqrt{T}\big(\mathbf{R}\hat{\beta}(\theta) - \mathbf{r}\big) \xrightarrow{D} \mathcal{N}\big(\mathbf{0},\, \theta(1-\theta)\boldsymbol{\Gamma}(\beta(\theta))\big),$$

where $\boldsymbol{\Gamma}(\beta(\theta)) = \mathbf{R}\mathbf{D}(\beta(\theta))\mathbf{R}'$.

The Null Distribution of the Wald Test

$$\mathcal{W}_T(\theta) = T\left[\mathbf{R}\hat{\beta}(\theta) - \mathbf{r}\right]'\widehat{\boldsymbol{\Gamma}}(\theta)^{-1}\left[\mathbf{R}\hat{\beta}(\theta) - \mathbf{r}\right]/[\theta(1-\theta)] \xrightarrow{D} \chi^2(q),$$

where $\widehat{\boldsymbol{\Gamma}}(\theta) = \mathbf{R}\widehat{\mathbf{D}}(\theta)\mathbf{R}'$, with $\widehat{\mathbf{D}}(\theta)$ a consistent estimator of $\mathbf{D}(\beta(\theta))$.

Sup-Wald Test

- H_0: $\mathbf{R}\beta(\theta) = \mathbf{r}$ for all $\theta \in \mathcal{S} \subset (0,1)$.
- The Brownian bridge: $\mathbf{B}_q(\theta) \stackrel{d}{=} [\theta(1-\theta)]^{1/2}\mathcal{N}(\mathbf{0},\ \mathbf{I}_q)$, and hence

$$\widehat{\mathbf{\Gamma}}(\theta)^{-1/2}\sqrt{T}\left[\mathbf{R}\hat{\beta}(\theta) - \mathbf{r}\right] \xrightarrow{D} \mathbf{B}_q(\theta).$$

Thus, $\mathcal{W}_T(\theta) \xrightarrow{D} \left\|\mathbf{B}_q(\theta)/\sqrt{\theta(1-\theta)}\right\|^2$, uniformly in θ.

The Null Distribution of the Sup-Wald Test

$$\sup_{\theta\in\mathcal{S}} \mathcal{W}_T(\theta) \Rightarrow \sup_{\theta\in\mathcal{S}} \left\| \frac{\mathbf{B}_q(\theta)}{\sqrt{\theta(1-\theta)}} \right\|^2,$$

where $\mathcal{S}$ is a compact set in $(0,1)$.

- To test $\mathbf{R}\beta(\theta) = \mathbf{r}$, $\theta \in [a, b]$, set $a = \theta_1 < \ldots < \theta_n = b$ and compute

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

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

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

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

  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

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

- -2 times the log-likelihood ratio is

$$2\big[L_T(\hat{\beta}(\theta);\theta) - L_T(\tilde{\beta}(\theta);\theta)\big] = 2\big[\tilde{V}_T(\theta) - \hat{V}_T(\theta)\big].$$

- Koenker and Machado (1999):

$$\mathcal{LR}_T(\theta) = \frac{2\big[\tilde{V}_T(\theta) - \hat{V}_T(\theta)\big]}{\theta(1-\theta)[f_{e_\theta}(0)]^{-1}} \xrightarrow{D} \chi^2(q).$$

This test is also known as the quantile ρ test.

- Koenker and Bassett (1982): For median regression,

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

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: $(Y_1, Y_0) \perp D \mid X$,

$$\mathbb{E}(Y|D=1,X) - \mathbb{E}(Y|D=0,X) = \mathbb{E}(Y_1 - Y_0|X),$$

so that the ATE is $\mathbb{E}(Y_1 - Y_0) = \mathbb{E}[\mathbb{E}(Y_1 - Y_0|X)]$.

- Using the sample counterpart of $\mathbb{E}(Y|D=1,X) - \mathbb{E}(Y|D=0,X)$ we have

$$\widehat{\text{ATE}} = \frac{1}{N}\sum_{i=1|}^{N}\big[\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)\big].$$

- For the dummy-variable regression:

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

the LS estimate of γ is $\widehat{\text{ATE}}$.

- 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 θ 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 θ th QTE.
- Other: A weighting estimator based on the propensity score.

Difference in Differences

- The impact of a program (policy) may be observed after 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 s 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 α_2.
  - The treatment effect is the difference between these two effects: α_3.