Quantile Regression: Primary Techniques

Shih-Kang Chao

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

Outline

1. Introduction
    - Tail event analysis
    - Theoretical quantiles and sample quantiles
2. Quantile Regression (QR)
    - Definition
    - Digression: treatment effect
    - Numerical properties
    - DGP and quantile function
    - Algebraic properties
    - Illustration: Engel's curve
3. Asmptotic Properties
    - Consistency
    - Asymptotic normality
4. Nonparametric QR
5. CAViaR model

Motivation

- Randomness is subject to certain law (distribution)
- Descriptive measures:
  - Moments:
    - 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: "average" and "center" of the distribution, but may provide little info about the tails

Tail Event Analysis

- Tail analysis is useful in many fields:
  - Banking: Value at Risk
  - Meteorology/Agriculture: temperature, rainfall; climate change
  - Energy Economy: electricity Demand

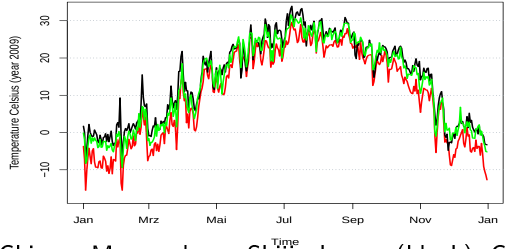


Figure 1: Chinese Meteorology: Shijiazhuang (black), Chengde (red) and Huailai (green) temperature data in 2009.

Regression Analysis

Consider regression $y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + e_t$, where $y \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^p$,

- ⊡ Least squares (LS): Legendre (1805)
  - ▸ $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} (y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta})^2$
  - ▸ $\boldsymbol{x}_t^\top \hat{\boldsymbol{\beta}}$ estimates the conditional mean of y given $\boldsymbol{x}$
- ⊡ Least squares (LAD): Boscovich (1755)
  - ▸ $\check{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} |y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta}|$
  - ▸ $\boldsymbol{x}_t^\top \check{\boldsymbol{\beta}}$ estimates the conditional median of y given $\boldsymbol{x}$
- ⊡ Both measures only the central tendency of the conditional distribution

Quantiles

- The τth $(0 < \tau < 1)$ quantile of F_Y is

$$q_Y(\tau) \overset{\text{def}}{=} F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\}.$$

- $q_Y(\tau)$ is an order statistics
- Define an asymmetric (linear) loss function:

$$\rho_\tau(u) \overset{\text{def}}{=} |\tau - \mathbf{1}(u \leq 0)||u|$$

- Given that Y is a random variable,

$$\mathsf{E}\,\rho_\tau(Y-\theta) = \tau \int_{y>\theta} |y-\theta| dF_Y(y) + (1-\tau)\int_{y<\theta} |y-\theta| dF_Y(y).$$

Quantiles

$q_Y(\tau)$ can be obtained via minimizing the expected asymmetric loss function. By first order condition:

$$\begin{aligned} 0 &\overset{!}{=} \frac{\partial}{\partial\theta} \operatorname{E} \rho_\tau(\theta) \\ \Rightarrow 0 &= -\tau \int_{y>\theta} dF_Y(y) + (1-\tau) \int_{y<\theta} dF_Y(y) \\ &= -\tau(1 - F_Y(\theta)) + (1-\tau) F_Y(\theta) \\ &= -\tau + F_Y(\theta), \end{aligned}$$

so $\theta = F_Y^{-1}(\tau) = q_Y(\tau)$ is the minimizer of $\operatorname{E} \rho_\tau(\theta)$, given the true distribution of Y is known.

Sample Quantiles

- F_Y is in general unknown, the sample counterpart of the expected asy. linear loss function is

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

  $\rho_\tau(y_t-\theta)$ is also known as the check function.
- Sample quantile $\hat{q}(\tau)$ can be found via an optimization scheme
- $F_T(\hat{q}(\tau))=\tau$, where F_T is the empirical dist. function and $\hat{q}(\tau)$ is the minimizer of (1)
- Specifying a function $f(\boldsymbol{x}_t)$ for the distribution quantile, we can use the optimization scheme to estimate the function

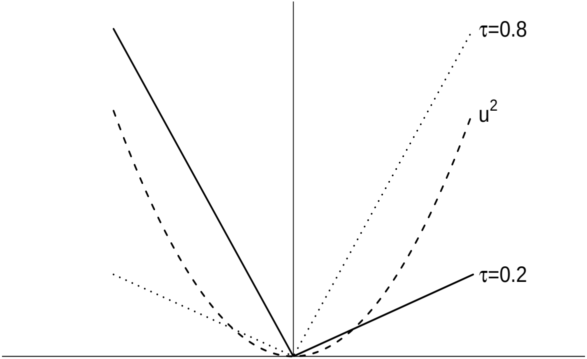


Figure 2: Plots of $\rho_\tau(u) = \big|\tau - \mathbf{1}(u \le 0)\big||u|$.

Quantile Regression

Definition (Koenker and Basset (1978))

Given the model $y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + \varepsilon_t$, the τth QR estimator $\hat{\boldsymbol{\beta}}(\tau)$ minimizes

$$V_T(\boldsymbol{\beta};\tau) = \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta}),$$

where $\rho_\tau(u) = \big|\tau - \mathbf{1}(u \leq 0)\big||u|$.

- For $\tau = 0.5$, $V_T(\boldsymbol{\beta};\tau)$ is symmetric, and $\hat{\boldsymbol{\beta}}(0.5)$ is the LAD estimator
- $\boldsymbol{x}_t^\top \boldsymbol{\beta}(\tau)$ gives the estimate for τth conditional quantile function $q_{Y|\boldsymbol{X}}(\tau)$. $\hat{\beta}_i(\tau)$ can be viewed as the estimated marginal effect of the ith regressor on $q_{Y|\boldsymbol{X}}(\tau)$

Digression: Treatment Effect

- Measuring the impact of a treatment (program, policy, intervention) in the distribution (e.g. income, age)
- Let $Y_1(Y_0)$ be the treatment(control) group, D is a dummy varible, the The effect can be measured through
    - Mean: the average treatment effect $\Delta_m = \mathsf{E}[Y_1 - Y_0]$
    - Quantile: $\Delta_\tau = \hat{F}_{1,n}^{-1}(\tau) - \hat{F}_{0,n}^{-1}(\tau)$, where $F_{1,n}, F_{0,n}$ are empirical distribution function of treatment and control group, for $0 < \tau < 1$
- If the experiment is randomized: $\mathsf{E}[Y_1 - Y_0] = \mathsf{E}[Y_1 | D = 1] - \mathsf{E}[Y_0 | D = 0]$
- To measure Δ_m, one can run a dummy-variable regression

$$Y_i = \alpha + D_i \gamma + \boldsymbol{X}_i^\top \beta + e_i, \quad i = 1, ..., n,$$

the LS estimate of γ is the estimated average treatment effect

Quantile Treatment Effect (QTE)

Doksum (1974): if we define $\Delta(x)$ as the "horizontal distance" between F_0 and F_1 at x so that

$$F_1(x) = F_0\left(x + \Delta(x)\right),$$

then $\Delta(x)$ can be expressed as

$$\Delta(x) = F_1^{-1}\left(F_0(x)\right) - x,$$

changing variable with $\tau = F_0(x)$, one gets the quantile treatment effect:

$$\Delta_\tau = \Delta(F_0^{-1}(\tau)) = F_1^{-1}(\tau) - F_0^{-1}(\tau).$$

Quantile Treatment Effect (QTE)

- One can run a quantile regression on $(D_i, \boldsymbol{X}_i)$ where $D_i = 0$ if i is in control group and $D_i = 1$ otherwise,

$$Y_i = \alpha + D_i \gamma + \boldsymbol{X}_i^\top \boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n,$$

the $\hat{\gamma}(\tau)$ is the estimated τ-th QTE (location shift)

-

$$Y_i = \alpha + \boldsymbol{X}_i^\top (\boldsymbol{\beta} + D_i \gamma) + e_i, \quad i = 1, \ldots, n,$$

the $\hat{\gamma}(\tau)$ is the estimated τ-th QTE (scaling)

- Quantile regression gives a complete look at the change in the distribution from treatment

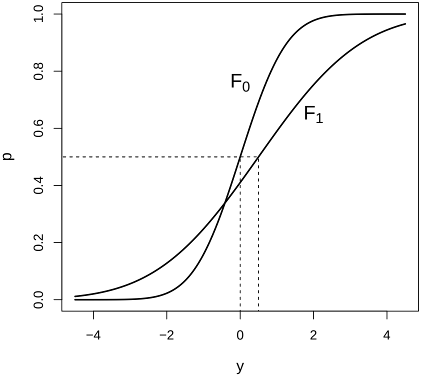


Figure 3: Horizontal distance between the treatment and control distribution functions.

Minimizing V_T

- Two main difficulties:
  - The QR estimator $\hat{\beta}(\tau)$ does not have a closed form
  - V_T is not everywhere differentiable, so that derivative-based algorithm (Newton method) does not work
- Two ways to solve:
  - approximate $\rho_\tau(u)$ with smooth function, and apply the Newton method
  - One can also apply linear programming
- Linear programming: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ can be expressed as

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

where

$$\boldsymbol{A} = \left[\boldsymbol{X}, -\boldsymbol{X}, I_T, -I_T\right],$$

$$\boldsymbol{z} = \left[(\boldsymbol{\beta}^+)^\top, (\boldsymbol{\beta}^-)^\top, (\boldsymbol{e}^+)^\top, (\boldsymbol{e}^-)^\top\right]^\top$$

Minimizing V_T

- Let $\boldsymbol{c} = \left[0^\top, 0^\top, \tau 1^\top, (1-\tau)1^\top\right]^\top$. Minimizing V_T is equivalent to the following constrained linear program (cLP):

$$\min_{\boldsymbol{z} \in \mathbb{R}_+^{2p} \times \mathbb{R}_+^{2T}} \frac{1}{T} \boldsymbol{c}^\top \boldsymbol{z}, \quad \text{s.t. } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}, \ \boldsymbol{z} \geq 0.$$

- This cLP solves on the extreme points(vertices) of the polyhedral convex set defined by the constraint $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}$ (Koenker (2005) Sec. 6.2)
- Define basic solution:

$$\beta(H) = \boldsymbol{X}(H)^{-1}\boldsymbol{y}(H)$$

where $H \subset \{1, ..., T\}$ with $|H| = \dim(\beta)$, and $\boldsymbol{y}(H)/\boldsymbol{X}(H)$ are the subvector/submatrix of $\boldsymbol{y}/\boldsymbol{X}$ with the corresponding elements/rows identified by index set H

Minimizing V_T

- $z(H) = \left[(\boldsymbol{\beta}(H)^+)^\top, (\boldsymbol{\beta}(H)^-)^\top, (\boldsymbol{e}(H)^+)^\top, (\boldsymbol{e}(H)^-)^\top\right]^\top$ are the vertices of the constrain set
- $V_T(\boldsymbol{\beta};\tau)$ is a convex function in $\boldsymbol{\beta}$. The minimizer $\hat{\boldsymbol{\beta}}(H)$ of $V_T(\boldsymbol{\beta};\tau)$ makes the directional derivative of V_T in direction $\boldsymbol{w}$ satisfy

$$\nabla V_T(\boldsymbol{\beta}) \overset{\text{def}}{=} -\frac{1}{T}\sum_{t=1}^{T} \psi_\tau^*(y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\beta}}(H), -\boldsymbol{x}_t^\top \boldsymbol{w})\boldsymbol{x}_t^\top \boldsymbol{w} \geq 0, \tag{2}$$

for all $\boldsymbol{w} \in \mathbb{R}^p$ with $\|\boldsymbol{w}\| = 1$, and

$$\psi_\tau^*(u,v) = \begin{cases} \tau - \mathbf{1}(u < 0), & \text{if } u \neq 0 \\ \tau - \mathbf{1}(v < 0), & \text{if } u = 0. \end{cases}$$

Minimizing V_T

Under the assumption that $\boldsymbol{X}$ is orthonormal, and reparametrize $\boldsymbol{w} = \boldsymbol{X}^{-1}\boldsymbol{v}$, where $\boldsymbol{v}$: elementary vectors. Given that $y_t - \boldsymbol{x}_t^\top \beta(H) \neq 0$ for any $t \notin H$. We have the following interesting result:

Theorem (Koenker (2005), Theorem 2.2.)

Let P, N, and Z denote the proportion of positive, negative, and zero elements of the residual vector $\boldsymbol{y} - \boldsymbol{X}\hat{\beta}(\tau)$. If $\boldsymbol{X}$ contains an intercept, that is, if there exists $\alpha \in \mathbb{R}^p$ such that $\boldsymbol{X}\alpha = 1_n$, then for any $\hat{\beta}(\tau)$ minimizing $V_T(\beta)$, we have

$$\frac{N}{n} \leq \tau \leq \frac{N+Z}{n},$$

and

$$\frac{P}{n} \leq 1 - \tau \leq \frac{P+Z}{n}.$$

Remarks

- MLE interpretation: $\beta(\tau)$ is the QMLE based on an asymmetric Laplace density:

$$f_\tau(u) = \tau(1-\tau)\exp\big[-\rho_\tau(u)\big].$$

- $\beta(\tau)$ is more robust to outliers
- The estimated hyperplane with normal $\hat{\beta}(\tau)$ must interpolate $\dim(\hat{\beta})$ observations in the sample
- QR utilizes all sample data (with different weights)

DGP and its quantile function

Let ε_t be i.i.d. with common distribution function F_ε.

- ⊡ DGP1 (location shift): $y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta}_o + \varepsilon_t = \beta_0 + \tilde{\boldsymbol{x}}_t^\top \boldsymbol{\beta}_1 + \varepsilon_t$
  - $q_{Y|\boldsymbol{X}}(\tau) = \boldsymbol{x}_t^\top \boldsymbol{\beta}_o + F_\varepsilon^{-1}(\tau) = \beta_0 + F_\varepsilon^{-1}(\tau) + \tilde{\boldsymbol{x}}_t^\top \boldsymbol{\beta}_1$
  - Quantile functions differ only by the "intercept"
- ⊡ DGP2 (scale): $y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta}_o + (\boldsymbol{x}_t^\top \boldsymbol{\gamma}_o)\varepsilon_t$
  - $q_{Y|\boldsymbol{X}}(\tau) = \boldsymbol{x}_t^\top \boldsymbol{\beta}_o + (\boldsymbol{x}_t^\top \boldsymbol{\gamma}_o) F_\varepsilon^{-1}(\tau) = \boldsymbol{x}_t^\top \left[\boldsymbol{\beta}_o + \boldsymbol{\gamma}_o F_\varepsilon^{-1}(\tau)\right]$
  - Quantile functions differ not only by the "intercept" but also the "slope" term
  - The model can be expressed as

$$y_t = \boldsymbol{x}_t^\top \underbrace{\left[\boldsymbol{\beta}_o + \boldsymbol{\gamma}_o F_\varepsilon^{-1}(\tau)\right]}_{\boldsymbol{\beta}(\tau)} + \tilde{\varepsilon}_t,$$

where $q_{\tilde{\varepsilon}_t|\boldsymbol{X}} = 0$. The quantile estimator $\hat{\boldsymbol{\beta}}(\tau)$ converges to $\boldsymbol{\beta}(\tau)$

DGP and its quantile function

DGP3: $y_t = \sigma_t \varepsilon_t$, $\sigma_t = \alpha_0 + \alpha_1 |y_{t-1}| + \beta_1 \sigma_{t-1}$. (GARCH(1,1) on standard deviation).
The quantile function is

$$\begin{aligned} q_{Y|\mathcal{F}_{t-1}}(\tau) &= \left[\alpha_0 + \alpha_1 |y_{t-1}| + \beta_1 \sigma_{t-1}\right] q_\varepsilon(\tau) \\ &= \underbrace{\alpha_0 q_\varepsilon(\tau)}_{\tilde{\alpha}_0} + \underbrace{\alpha_1 q_\varepsilon(\tau)}_{\tilde{\alpha}_1} |y_{t-1}| + \beta_1 q_{Y|\mathcal{F}_{t-2}}(\tau). \end{aligned}$$

- The quantile function has a GARCH(1,1) form too
- This is one of the variation of the famous Conditional Autoregressive Value at Risk (CAViaR) model

Algebraic Properties: Equivalence

Let $\hat{\beta}(\tau)$ be the QR estimator of the quantile regression of y_t on $\boldsymbol{x}_t$. Let y_t^* be a translation of y_t and $\hat{\boldsymbol{\beta}}^*(\tau)$ be the QR estimator of y_t^* on $\boldsymbol{x}_t$.

- Scale equivariance: For scaled $y_t^* = cy_t$:
  - For $c > 0$, $\hat{\boldsymbol{\beta}}^*(\tau) = c\hat{\boldsymbol{\beta}}(\tau)$.
  - For $c < 0$, $\hat{\boldsymbol{\beta}}^*(1-\tau) = c\hat{\boldsymbol{\beta}}(\tau)$.
  - $\hat{\boldsymbol{\beta}}^*(0.5) = c\hat{\boldsymbol{\beta}}(0.5)$, regardless of the sign of c.
  - Example: y=yearly salary, x =age. Divide y by 1000 to balance the scale.
- Shift equivariance: For $y_t^* = y_t + \boldsymbol{x}_t^\top \boldsymbol{\gamma}$. Then $\hat{\boldsymbol{\beta}}^*(\tau) = \hat{\boldsymbol{\beta}}(\tau) + \boldsymbol{\gamma}$.

Algebraic Properties: Equivalence

- Equivariance to reparameterization of design: Let $\boldsymbol{X}^* = \boldsymbol{X}\boldsymbol{A}$, for nonsingular $\boldsymbol{A}$, then $\hat{\boldsymbol{\beta}}^*(\tau) = \boldsymbol{A}^{-1}\hat{\boldsymbol{\beta}}(\tau)$.
- For a nondecreasing function h,

$$\mathsf{P}\left\{y \le a\right\} = \mathsf{P}\left\{h(y) \le h(a)\right\} = \tau,$$

so that

$$q_{h(y)|\boldsymbol{x}}(\tau) = h\left(q_{y|\boldsymbol{x}}(\tau)\right),$$

note that the expectation does not have this property

- Example: if $\boldsymbol{x}_t^\top \boldsymbol{\beta}$ is the τth conditional quantile of log return, then $\exp(\boldsymbol{x}_t^\top \boldsymbol{\beta})$ is the τth conditional quantile of price ratio

Illustration: Engel's curve

- Engel's (1857) study of households' expenditure on food versus annual income. 235 obs.
- Hypothesis: Food expenditure constitutes a declining share of household income

Table 1: The slopes of Engel's curve.

τ	0.15	0.25	0.5	0.75	0.95	0.99
$\hat{\beta}(\tau)$	0.832	0.849	0.877	0.916	0.922	0.893

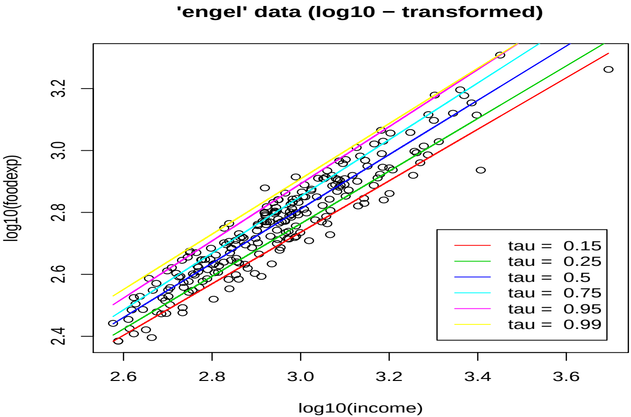


Figure 4: Engel curves for food. Households' expenditure on food versus annual income. 235 obs.

Asymptotic Properties: Heuristics

- ⊡ Consider $y_t = q + \varepsilon_t$, define

$$g_T(q) \overset{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{1}(y_t < q) - \tau\right)$$

which is the "FOC" of minimizing $T^{-1}\sum_{t=1}^{T}\rho_\tau(y_t - q)$.

- ⊡ Obviously $g_T(q)$ is non-decreasing in q, so that $\hat{q}(\tau) > q$ iff $g_T(q) < 0$. Thus,

$$\mathsf{P}\left[\sqrt{T}(\hat{q}(\tau) - q(\tau)) > c\right] = \mathsf{P}\left[g_T\left(q(\tau) + c/\sqrt{T}\right) < 0\right] \tag{3}$$

Moreover,

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

$$\mathsf{Var}\left[g_T\left(q(\tau) + \frac{c}{\sqrt{T}}\right)\right] = \frac{1}{T}F(1-F) \approx \frac{1}{T}\tau(1-\tau).$$

Asymptotic Properties: Heuristics

$$\begin{aligned}
\mathsf{P}\left[\sqrt{T}\{\hat{q}(\tau)-q(\tau)\}>c\right] &= \mathsf{P}\left[\frac{g_T\left(q(\tau)+c/\sqrt{T}\right)}{\sqrt{\tau(1-\tau)/T}}<0\right] \text{ (by (3))} \\
&= \mathsf{P}\left[\frac{g_T\left(q(\tau)+c/\sqrt{T}\right)}{\sqrt{\tau(1-\tau)/T}}-\frac{c}{\lambda}<-\frac{c}{\lambda}\right] \\
&= \mathsf{P}\left[\frac{g_T\left(q(\tau)+c/\sqrt{T}\right)-cf\left(q(\tau)\right)/\sqrt{T}}{\sqrt{\tau(1-\tau)/T}}<-\frac{c}{\lambda}\right] \\
&\overset{\mathcal{L}}{\to} 1-\Phi(c/\lambda),
\end{aligned}$$

where $\lambda^2=\tau(1-\tau)/f^2(q(\tau))$. By CLT. This implies

$$\sqrt{T}\{\hat{q}(\tau)-q(\tau)\}\overset{\mathcal{L}}{\to}N(0,\lambda^2).$$

QR as Extremal Estimator

Theorem (Newey and McFadden (1994), Theorem 2.1.)

Suppose that $\hat{\beta}$ *minimizes the objective function* $V_T(\boldsymbol{\beta})$ *in the parameter space* Θ *and* $\boldsymbol{\beta}_o$ *is the unique solution of* $F_{Y|\boldsymbol{X}}(\boldsymbol{x}_t^\top \boldsymbol{\beta}) = \tau$*.* $V_0(\boldsymbol{\beta}) = \mathrm{E}[V_T(\boldsymbol{\beta})]$*. If*

1. Θ *is compact*
2. $V_T(\boldsymbol{\beta})$ *converges uniformly to* $V_0(\boldsymbol{\beta})$ *in probability*
3. $V_0(\boldsymbol{\beta})$ *is continuous*
4. $V_0(\boldsymbol{\beta})$ *is uniquely minimized at* $\boldsymbol{\beta}_o$*.*

Then $\hat{\beta} \xrightarrow{\mathrm{P}} \boldsymbol{\beta}_o$*.*

Proof.

(2) follows from LLN and (1). Via FOC of $\mathrm{E}[\rho_\tau(y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta})]$, the fourth condition follows from $F_{Y|\boldsymbol{X}}(\boldsymbol{x}_t^\top \boldsymbol{\beta}_o) = \tau$. □

Asymptotic Normality

- There is no "first order condition" for QR (i.e. $\partial_{\boldsymbol{\beta}} V_T(\boldsymbol{\beta}) = 0$)
- The QR estimator satisfies "asymptotic FOC":

$$\sqrt{T}\Psi_n(\hat{\boldsymbol{\beta}}(\tau)) \overset{\text{def}}{=} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{x}_t \psi_\tau(y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\beta}}) = \mathcal{O}_P(1)\,.$$

where $\psi_\tau(u) = \mathbf{1}(u < 0) - \tau$.

- Stochastic equicontinuous condition (SEC):

$$\sqrt{T}\left[\Psi(\hat{\boldsymbol{\beta}}) - \Psi(\boldsymbol{\beta}_o) - \{\Psi_n(\hat{\boldsymbol{\beta}}) - \Psi_n(\boldsymbol{\beta}_o)\}\right] = \mathcal{O}_p(1).$$

where $\Psi(\boldsymbol{\beta}) = \mathrm{E}[\Psi_n(\hat{\boldsymbol{\beta}})]$.

- In particular, $\Psi(\boldsymbol{\beta}_o) = 0$ and $\Psi(\boldsymbol{\beta})$ is differentiable

Digression: some conditions guarantee SEC

Let ψ be the (sub)differential of the likelihood function, $q_{t,T}(\boldsymbol{\beta}) = \nabla g_{t,T}(\boldsymbol{\beta})\psi(y_t - g(\boldsymbol{x}_t, \boldsymbol{\beta}))$, and g is the specified function,

$$\Psi_n(\boldsymbol{\beta}) = T^{-1} \sum_{t=1}^{T} q_{t,T}(\boldsymbol{\beta}),$$

$\Phi_n(\boldsymbol{\beta}) = \mathrm{E}[\Psi_n(\boldsymbol{\beta})]$

$$\mu_{t,T}(\boldsymbol{\beta}, d) = \sup_{\gamma} \left\{ \| q_{t,T}(\gamma) - q_{t,T}(\boldsymbol{\beta}) \| : \| \boldsymbol{\beta} - \gamma \| \leq d \right\}.$$

Huber (1967) gives conditions required for SEC:

N1 For each t, $\psi_{t,T}(\boldsymbol{\beta})$ is measurable for each $\boldsymbol{\beta} \in \Theta$ and is separable in the sense of Doob(1953)

N2 For each T, there is some $\boldsymbol{\beta}_o$ such that $\Phi_n(\boldsymbol{\beta}_o) = 0$

Digression: some conditions guarantee SEC

N3 There are strictly positive numbers a, b, c, d_0 and T_0 such that for all t and $T \geq T_0$,

1. $\|\Phi_n(\boldsymbol{\beta})\| \geq a\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|$ for $\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| \leq d_0$
2. $\mathrm{E}\,\|\mu_{t,T}(\boldsymbol{\beta}, d)\| \leq bd$ for $\|\boldsymbol{\beta} - \boldsymbol{\beta}_n\| + d \leq d_0$, $d \geq 0$
3. $\mathrm{E}\,\|\mu_{t,T}(\boldsymbol{\beta}, d)\|^r \leq cd$ for $\|\boldsymbol{\beta} - \boldsymbol{\beta}_n\| + d \leq d_0$ for some $r > 2$

N4 There exists $K < 0$ such that $\mathrm{E}\left[\|q_{t,T}(\hat{\boldsymbol{\beta}})\|^2\right] < K$ is finite for all t, T

For time dependent model (e.g. ARMA, ARCH), Weiss (1991) adds one more condition:

N5 $\{\boldsymbol{x}_t, \varepsilon_t\}$ is α-mixing with $\alpha(L) \leq \Delta L^{-\lambda}$ for some $\lambda > 2r/(r-2)$, $r > 2$

Asymptotic Normality

Mean value theorem on $\Psi(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_o$ gives

$$\sqrt{T}\Psi(\hat{\boldsymbol{\beta}}) = \underbrace{\sqrt{T}\Psi(\boldsymbol{\beta}_o)}_{=0 \text{ correct dynamic specification}} + \sqrt{T}\nabla_{\boldsymbol{\beta}}\Psi(\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o),$$

for $\tilde{\boldsymbol{\beta}}$ lies between $\boldsymbol{\beta}_o$ and $\hat{\boldsymbol{\beta}}$. With $\hat{\boldsymbol{\beta}} \xrightarrow{\mathsf{P}} \boldsymbol{\beta}_o$ and continuity of $\nabla_{\boldsymbol{\beta}}\Psi$,

$$\nabla_{\boldsymbol{\beta}}\Psi(\tilde{\boldsymbol{\beta}}) \xrightarrow{\mathsf{P}} \nabla_{\boldsymbol{\beta}}\Psi(\boldsymbol{\beta}_o) \overset{\text{def}}{=} G_o.$$

Suppose G_o is nonsingular,

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \xrightarrow{\mathsf{P}} G_o^{-1}\sqrt{T}\Psi(\hat{\boldsymbol{\beta}}).$$

Asymptotic Normality

Note that by SEC and suppose that G_o^{-1} has finite norm, by Slutzky theorem

▸ Slutzky Thm.

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \xrightarrow{\mathrm{P}} G_o^{-1}\sqrt{T}\left\{\Psi_n(\boldsymbol{\beta}_o) - \Psi_n(\hat{\boldsymbol{\beta}})\right\}.$$

Via a CLT,

$$\sqrt{T}G_o^{-1}\Psi_n(\boldsymbol{\beta}_o) = G_o^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{x}_t\psi(y_t - \boldsymbol{x}_t^\top\boldsymbol{\beta}_o) \xrightarrow{\mathcal{L}} N(0, G_o^{-1}\Sigma_o G_o^{-1})$$

where

$$\Sigma_o = \mathrm{E}\left[\boldsymbol{x}_t\boldsymbol{x}_t^\top \,\mathrm{E}\left[\psi(y_t - \boldsymbol{x}_t^\top\boldsymbol{\beta}_o)^2|\boldsymbol{x}_t\right]\right],$$

and $\sqrt{T}\Psi_n(\hat{\boldsymbol{\beta}}) = \mathcal{O}_p(1)$ by "asymptotic FOC". Hence, via Slutzky theorem, we get

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

Slutzky Theorem

Theorem (Slutzky)

If $X_n \xrightarrow{\text{d}} X$ *and* $Y_n \xrightarrow{\text{P}} c$, *where c is a constant, then:*

1. $X_n + Y_n \xrightarrow{\text{d}} X + c$
2. $X_n Y_n \xrightarrow{\text{d}} cX$, *if* $c \neq 0$
   $X_n Y_n \xrightarrow{\text{P}} 0$, *if* $c = 0$
3. $X_n / Y_n \xrightarrow{\text{d}} X / c$, *if* $c \neq 0$

▸ Asymp. Normality

Asymptotic Normality

The form of G_o and Σ_o can be explicitly computed under the i.i.d. error ε_t:

⊡ Under the interchangeability of differentiation and integration,

$$\begin{aligned} G_o = \nabla_{\boldsymbol\beta} \Psi(\boldsymbol\beta_o) &= \nabla_{\boldsymbol\beta} \operatorname{E}\left[\operatorname{E}[\boldsymbol{x}_t \psi(y_t - \boldsymbol{x}_t^\top \boldsymbol\beta_o)] | \boldsymbol{x}_t \right] \\ &= \nabla_{\boldsymbol\beta} \operatorname{E}\left[\boldsymbol{x}_t (F_{Y|\boldsymbol{X}}(\boldsymbol{x}_t^\top \boldsymbol\beta_o) - \tau)\right] \\ &= \operatorname{E}\left[\boldsymbol{x}_t \boldsymbol{x}_t^\top f_{Y|\boldsymbol{X}}(\boldsymbol{x}_t^\top \boldsymbol\beta_o)\right] \\ &= \operatorname{E}\left[\boldsymbol{x}_t \boldsymbol{x}_t^\top f_{\varepsilon|\boldsymbol{X}}(0)\right]. \end{aligned}$$

⊡ $\psi_\tau(\varepsilon_t)$ is i.i.d. Bernoulli with mean 0 and variance $\tau(1-\tau)$ given $\boldsymbol{x}_t$ (why?),

$$\begin{aligned} \Sigma_o &= \operatorname{E}\left[\boldsymbol{x}_t \boldsymbol{x}_t^\top \operatorname{E}\left[\{\mathbf{1}(y_t - \boldsymbol{x}_t^\top \boldsymbol\beta_o) - \tau\}^2 | \boldsymbol{x}_t\right]\right] \\ &= \tau(1-\tau) \operatorname{E}\left[\boldsymbol{x}_t \boldsymbol{x}_t^\top\right] \overset{\text{def}}{=} \tau(1-\tau) \boldsymbol{M}_{\boldsymbol{XX}} \end{aligned}$$

Asymptotic Normality

Theorem

$$\boldsymbol{\sqrt{T}(\hat{\beta} - \beta_o) \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau) G_o^{-1} M_{XX} G_o^{-1}),}$$

where $G_o = \mathrm{E}\left[\boldsymbol{x}_t \boldsymbol{x}_t^\top f_{\varepsilon|\boldsymbol{X},t}(0)\right]$ *and* $M_{\boldsymbol{XX}} = \mathrm{E}\left[\boldsymbol{x}_t \boldsymbol{x}_t^\top\right]$.

Remarks:

⊡ Conditional heterogeneityis characterized by the conditional density $f_{\varepsilon|\boldsymbol{X}}(0)$ in G_o, which is not limited to heteroskedasticity. This theorem applies to independently but not identically distributed data, e.g. DGP2, but not DGP3.

⊡ If $f_{\varepsilon|\boldsymbol{X}}(0) = f_\varepsilon(0)$, i.e. conditional homoskedasticity, then

$$\sqrt{T}(\hat{\beta} - \beta_o) \xrightarrow{\mathcal{L}} N\left(0, \frac{\tau(1-\tau)}{f_\varepsilon(0)^2} \boldsymbol{M}_{\boldsymbol{XX}}^{-1}\right)$$

Estimation of Asymptotic Covariance Matrix

Goal: Consistently estimate $\hat{D}(\boldsymbol{\beta}_o) = G_o^{-1} \boldsymbol{M}_{\boldsymbol{XX}} G_o^{-1}$,

- ⊡ $\boldsymbol{M}_{\boldsymbol{XX}} : \boldsymbol{M}_T = T^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^\top$
- ⊡ It is suggested by Powell in his lecture notes that

$$G_T = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h} K \left\{ \frac{y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\beta}}}{h} \right\} \boldsymbol{x}_t \boldsymbol{x}_t^\top,$$

where K is a kernel function satisfies $\int K(u) du = 1$, the bandwidth h sastisfies $h \to 0$ and $nh \to \infty$. In practice h can be chosen by standard methods like cross-validation, plug-in method.

- ⊡ Both estimators are robust to non i.i.d. data

Wald Test

$H_0 : \boldsymbol{R}\beta(\tau) = \boldsymbol{r}$, where $\boldsymbol{R}$ is $q \times p$ and $\boldsymbol{r}$ is $q \times 1$. Let $\boldsymbol{D}(\beta) = G_o^{-1} \boldsymbol{M_{XX}} G_o^{-1}$,

- $\boldsymbol{\sqrt{T}(\hat{\beta} - \beta_o) \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)D(\beta_o))}$,
- Under the null,

$$\boldsymbol{\sqrt{T}R(\hat{\beta} - \beta_o) = \sqrt{T}(R\hat{\beta} - r) \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)\Gamma(\beta_o))},$$

where $\boldsymbol{\Gamma}(\beta_o) = \boldsymbol{R}\boldsymbol{D}(\beta_o)\boldsymbol{R}^\top$.

Theorem (The Null Distribution of the Wald Test)

$$\mathcal{W}_T = T\big[\boldsymbol{R}\hat{\beta} - \boldsymbol{r}\big]^\top \hat{\boldsymbol{\Gamma}}^{-1}\big[\boldsymbol{R}\hat{\beta} - \boldsymbol{r}\big] / \big[\tau(1-\tau)\big] \xrightarrow{\mathcal{L}} \chi^2(q),$$

where $\hat{\boldsymbol{\Gamma}} = \boldsymbol{R}\hat{\boldsymbol{D}}(\beta_o)\boldsymbol{R}^\top$.

Sup Wald Test

- ⊡ $H_0 : \boldsymbol{R}\beta(\tau) = \boldsymbol{r}$ for all $\tau \in \mathcal{S} \subset (0,1)$ a compact set, $\boldsymbol{R} : q \times p$
- ⊡ Define the Brownian bridge: $\boldsymbol{B}_q \overset{\mathrm{d}}{=} [\tau(1-\tau)]^{1/2} N(0, I_q)$, for $0 < \tau < 1$, and hence

$$\hat{\boldsymbol{\Gamma}}^{-1/2}\sqrt{T}[\boldsymbol{R}\hat{\beta}(\tau) - \boldsymbol{r}] \xrightarrow{\mathcal{L}} \boldsymbol{B}_q(\tau).$$

Thus, $\mathcal{W}_T(\tau) \xrightarrow{\mathcal{L}} \|\boldsymbol{B}_q(\tau)/\sqrt{\tau(1-\tau)}\|^2$, for all τ

Theorem

$$\sup_{\tau\in\mathcal{S}} \mathcal{W}_T(\tau) \xrightarrow{\mathcal{L}} \sup_{\tau\in\mathcal{S}} \left\| \frac{\boldsymbol{B}_q(\tau)}{\sqrt{\tau(1-\tau)}} \right\|^2,$$

where $S \subset (0,1)$ is a compact set.

Sup Wald Test

- If $\mathcal{S} = [a, b]$, select $a = \tau_1 < ... < \tau_n = b$, and compute

$$\sup_{\tau \in \mathcal{S}} \mathcal{W}_T(\tau) \approx \sup_{i=1,..,n} \mathcal{W}_T(\tau_i)$$

- For $s = \tau/(1-\tau)$, $B(\tau)/\sqrt{\tau(1-\tau)} \overset{\mathrm{d}}{=} W(s)/\sqrt{s}$, where W: Wiener process, so that:

$$\mathrm{P}\left\{ \sup_{\tau \in [a,b]} \left\| \frac{\boldsymbol{B}_q(\tau)}{\sqrt{\tau(1-\tau)}} \right\|^2 < c \right\} = \mathrm{P}\left\{ \sup_{s \in [1, s_2/s_1]} \left\| \frac{\boldsymbol{W}_q(s)}{\sqrt{s}} \right\|^2 < c \right\}, \tag{4}$$

for all $c > 0$ with $s_1 = a/(1-a)$, $s_2 = b/(1-b)$

- Critical values of (4) can be obtained via simulations; some special cases were tabulated in in DeLong (1981) and Andrews (1993)

Likelihood Ratio Test

- Let $\hat{\beta}(\tau)$ and $\tilde{\beta}(\tau)$ be the constrained and unconstrained estimators and $\hat{V}_T(\tau) = V_T(\hat{\beta}(\tau);\tau)$ and $\tilde{V}_T(\tau) = V_T(\tilde{\beta}(\tau);\tau)$ be the corresponding objective functions
- Given the asymmetric Laplace density: $f_\tau(u) = \tau(1-\tau)\exp\big[-\rho_\tau(u)\big]$, the log-likelihood is

$$L_T(\beta;\tau) = T\log(\tau(1-\tau)) - \sum_{t=1}^{T}\rho_\tau(y_t - \boldsymbol{x}_t^\top\beta).$$

- -2 times the log-likelihood ratio is

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

Likelihood Ratio Test

⊡ Koenker and Machado (1999):

$$\mathcal{LR}_T(\tau) = \frac{2\left[\tilde{V}_T(\tau) - \hat{V}_T(\tau)\right]}{\tau(1-\tau)\left[f_\varepsilon(0)\right]^{-1}} \overset{\mathcal{L}}{\to} \chi^2(q).$$

where q is the number of restrictions. The test is also known as the quantile ρ test

Nonlinear Quantile Regression

- The quantile specification can be nonlinear, i.e.

$$q_{Y|\boldsymbol{x}}(\tau|\boldsymbol{x}) = g(\boldsymbol{x}, \boldsymbol{\beta}_o(\tau)),$$

  where g is nonlinear in $\boldsymbol{\beta}$.

- Define the nonlinear quantile regression estimator

$$\hat{\boldsymbol{\beta}}(\tau) = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \rho_\tau\big(y_t - g(\boldsymbol{x}_t, \boldsymbol{\beta})\big).$$

- Replace the "asymptotic FOC" in the linear case by

$$\sqrt{T}\Psi_n(\hat{\boldsymbol{\beta}}(\tau)) \overset{\text{def}}{=} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}}) \psi_\tau\big(y_t - g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}})\big) = \mathcal{O}_P(1).$$

Nonlinear Quantile Regression

The following theorem can be proved by imitating that of linear QR:

Theorem

Let $\Psi_n(\boldsymbol{\beta}) \overset{\text{def}}{=} T^{-1} \sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}}) \psi_\tau(y_t - g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}}))$ *and* $\Psi(\boldsymbol{\beta}) = \mathrm{E}[\Psi_n(\boldsymbol{\beta})]$. *Suppose that*

1. $\nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}})$ *is continuous in* $\boldsymbol{\beta}$
2. *The stochastic equicontinuous condition holds for* Ψ_n *and* Ψ
3. $\sqrt{T}\Psi_n(\boldsymbol{\beta}_o) \xrightarrow{\mathcal{L}} N(0, \Sigma_o)$ *where* $\Sigma_o = \tau(1-\tau)\,\mathrm{E}\left[\nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o) \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o)^\top\right]$

Then,

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

where $G_o = -\,\mathrm{E}\left[\nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o) \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o)^\top f_{\varepsilon|\boldsymbol{x}}(0)\right]$.

Nonlinear Quantile Regression

- Koenker (2005) Sec. 4.4 states the same limiting theorem under somewhat weaker condition
- The estimator

$$\boldsymbol{M}_T = \frac{1}{T}\sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}}) \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}})^\top$$

$$\xrightarrow{\mathsf{P}} \mathsf{E}\left[\nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o) \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o)^\top\right] \overset{\text{def}}{=} \boldsymbol{M}_{\boldsymbol{x}\boldsymbol{x}}.$$

- The estimator

$$G_T = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{h} K\left\{\frac{y_t - g(\boldsymbol{x}_t, \hat{\boldsymbol{\beta}})}{h}\right\} \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o) \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o)^\top$$

is again consistent for G_o

QR as *M*-estimator

Definition (Serfling (1980))

For any function $\psi(x, \boldsymbol{\beta})$, the *M*-functional is defined by the solution of

$$\int \psi(x, \boldsymbol{\beta}) dF(x) = 0,$$

and the associate *M*-estimator is defined by the solution of

$$\int \psi(x, \boldsymbol{\beta}) dF_T(x) = \frac{1}{T} \sum_{t=1}^{T} \psi(X_t, \boldsymbol{\beta}) = 0,$$

where $F_T(x) = T^{-1} \sum_{t=1}^{T} \mathbf{1}(X_t \leq x)$.

- Example: $\psi(u) = u - \beta$, the *M*-functional is the mean of X; the *M*-estimator is the sample mean

QR as *M*-estimator

- Fix $0 < \tau < 1$, let $\psi_\tau(u, q) = \mathbf{1}(u < q) - \tau$, the *M*-functional is the τth quantile of X, as we have shown before; the so-called "asymtotic *M*-estimator" satisfies instead

$$\sum_{t=1}^{T} \psi(X_t, q) = \mathcal{O}(\delta_T),$$

  where $\delta_T \to \infty$.

- This corresponds to the "asymptotic FOC" in the previous section on asymptotic normality of QR

QR as *M*-estimator

- Fix $0 < \tau < 1$, nonlinear τth QR M-functional can be defined by the solution of

$$\int \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta})\psi_\tau\big(y - g(\boldsymbol{x}, \boldsymbol{\beta})\big) dF_{Y|\boldsymbol{X}}(y) = 0,$$

  which exactly corresponds to $\Psi(\boldsymbol{\beta}_o) = 0$ in the previous discussion for asymp. normality, and $\hat{\boldsymbol{\beta}}(\tau)$ is the "asymptotic M-estimator"
- QR satisfies $\delta_T = \sqrt{T}$, for $\dim(\boldsymbol{\beta}) = p$ fixed and bounded $\nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}, \boldsymbol{\beta})$
- Huber (1967) shows consistency with $\delta_T = T$ and asymptotic normality with $\delta_T = \sqrt{T}$

Bahadur Representation

Theorem

Under the same conditions for proving the asymptotic normality of nonlinear quantile estimator, we have

$$\sqrt{T}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_o(\tau)) = G_o^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o) \psi_\tau(u_t(\tau)) + R_T,$$

where $u_t(\tau) = y_t - g(\boldsymbol{x}_t, \boldsymbol{\beta}_o(\tau))$ *and* $R_T = o_P(1)$.

- $u_t(\tau) = y_t - g(\boldsymbol{x}_t, \boldsymbol{\beta}_o(\tau)) = \varepsilon_t$ in additive error model
- The asymptotic normality easily follows from the Bahadur representation
- For $\dim(\boldsymbol{x}) = 1$, i.i.d. error, the rate of R_n

  $$R_T = \mathcal{O}(T^{-1/4}(\log\log T)^{3/4})$$

  which is the sharpest possible rate achieved by Kiefer (1967).

Bahadur Representation

- Bahadur representation can be derived for "asymptotic M-estimator"
- He and Shao (1996) prove Bahadur representation under weaker conditions and non identically distributed errors

$$\sqrt{T}\big(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_o(\tau)\big)$$

$$= G_o^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}} g(\boldsymbol{x}_t, \boldsymbol{\beta}_o) \psi_\tau\big(u_t(\tau)\big) + R_{T,1} + R_{T,2},$$

where $R_{T,1}$ relates to the accumulated error of non identical distribution and $R_{T,2}$ relates to Taylor approximation error of Ψ and δ_T

- Wu (2007) proves Bahadur representation under a version of dependent errors

Inversion of Empirical Distribution

- Suppose the data $\{(Y_t, \boldsymbol{X}_t); t = 1, ..., T\}$ are i.i.d. and we would like to estimate the τth conditional quantile function of the response Y, given $\boldsymbol{X} = \boldsymbol{x}$:

$$g(\boldsymbol{x}) = q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x})$$

- The first idea: find

$$q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) = \inf\left\{y : \hat{F}_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) \geq \tau\right\},$$

where $\hat{F}_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ is an estimate for $F_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$. Such approach is proposed by Li and Racine (2007)

Conditional CDF estimator

Li and Racine (2007) propose two estimators:

$$\tilde{F}(y|\boldsymbol{x}) = \frac{T^{-1}\sum_{t=1}^{T} \mathbf{1}(Y_i \leq y) K_H(\boldsymbol{X}_t - \boldsymbol{x})}{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})}$$

and

$$\hat{F}(y|\boldsymbol{x}) = \frac{T^{-1}\sum_{t=1}^{T} G\left\{(y - Y_i)/h_0\right\} K_H(\boldsymbol{X}_t - \boldsymbol{x})}{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})}$$

where $G(\cdot)$ is a kernel CDF (e.g. standard normal CDF), $K_H(\cdot) = K_{h_1}(\cdot)...K_{h_p}(\cdot)$ is a product kernel function with$K_{h_j}(x) = K\left(x/h_j\right)$, $H = (h_1, ..., h_p)$ bandwidthes and

$$\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{T}\sum_{t=1}^{T} K_H(\boldsymbol{x} - \boldsymbol{X}_t).$$

Asymptotic Normality

Assumptions:

(A1) Both $f_{\boldsymbol{X}}(\boldsymbol{x})$ and $F(y|\boldsymbol{x})$ have continuous second order derivative with respect to $\boldsymbol{x}$.

(A2) $nh_1 ... h_p \to \infty$ and $h_j \to 0$ for $j = 1, ..., p$

(A3) $K(\cdot)$ is symmetric, bounded, compactly supported and integrated to 1.

Let

$$\tilde{M}(y, \boldsymbol{x}) = \hat{f}_{\boldsymbol{X}}(\boldsymbol{x})\left[\tilde{F}(y|\boldsymbol{x}) - F(y|\boldsymbol{x})\right]$$

$$\hat{M}(y, \boldsymbol{x}) = \hat{f}_{\boldsymbol{X}}(\boldsymbol{x})\left[\hat{F}(y|\boldsymbol{x}) - F(y|\boldsymbol{x})\right]$$

Asymptotic Normality of $\tilde{F}(y|\mathbf{x})$

Theorem (A)

Under (A1)-(A3) and $f_{\mathbf{X}}(\mathbf{x}) > 0$ and $F(y|\mathbf{x}) > 0$, we have

1. $\mathsf{E}[\tilde{M}(y,\mathbf{x})] = f_{\mathbf{X}}(\mathbf{x})\big[\sum_{j=1}^{p} h_j^2 B_j(y,\mathbf{x})\big] + \mathcal{O}(\sum_{j=1}^{p} h_j^2)$, *where* $B_j(y,\mathbf{x}) = (1/2)\kappa_2\big[F_{jj}(y|\mathbf{x}) + 2f_{\mathbf{X},j}(\mathbf{x})F_j(y|\mathbf{x})/f_{\mathbf{X}}(\mathbf{x})\big]$.
2. $Var[\tilde{M}(y,\mathbf{x})] = (nh_1...h_p)^{-1}f_{\mathbf{X}}(\mathbf{x})^2\Sigma_{y|\mathbf{x}} + \mathcal{O}((nh_1...h_p)^{-1})$, *where* $\Sigma_{y|\mathbf{x}} = \|K\|_2^{2p}F(y|\mathbf{x})\big[1 - F(y|\mathbf{x})\big]/f_{\mathbf{X}}(\mathbf{x})$
3. *If* $(nh_1...h_p)^{-1/2}\sum_{j=1}^{p} h_j^3 = \mathcal{O}(1)$, *then*

$$(nh_1...h_p)^{-1/2}\left[\tilde{F}(y|\mathbf{x}) - F(y|\mathbf{x}) - \sum_{j=1}^{p} h_j^2 B_j(y,\mathbf{x})\right] \xrightarrow{\mathcal{L}} N(0,\Sigma_{y|\mathbf{x}})$$

Asymptotic Normality of $\hat{F}(y|\mathbf{x})$

Theorem (B)

Define $B_0(y,\mathbf{x}) = (1/2)\kappa_2 F_{yy}(y|\mathbf{x})$ *and let* $\Omega(y,\mathbf{x}) = \|K\|_2^{2p} F_y(y|\mathbf{x})/f_{\mathbf{X}}(\mathbf{x})$. *Letting* $|\bar{h}|^2 = \sum_{j=0}^{p} h_j^2$, *then under conditions similar to the last theorem,*

1. $\mathrm{E}[\hat{M}(y,\mathbf{x})] = f_{\mathbf{X}}(\mathbf{x})\left[\sum_{j=0}^{p} h_j^2 B_j(y,\mathbf{x})\right] + \mathcal{O}(|h|^2)$, *where* $B_j(y,\mathbf{x}) = (1/2)\kappa_2\left[F_{jj}(y|\mathbf{x}) + 2 f_{\mathbf{X},j}(\mathbf{x}) F_j(y|\mathbf{x})/f_{\mathbf{X}}(\mathbf{x})\right]$.
2. $Var[\hat{M}(y,\mathbf{x})] = (nh_1 \ldots h_p)^{-1} f_{\mathbf{X}}(\mathbf{x})^2\left[\Sigma_{y|\mathbf{x}} - h_0 C_K \Omega(y,\mathbf{x})\right] + \mathcal{O}((nh_1 \ldots h_p)^{-1})$, *where* $C_K = 2\int G(v)K(v)vdv$
3. *If* $(nh_1 \ldots h_p)^{-1/2}\sum_{j=1}^{p} h_j^3 = \mathcal{O}(1)$, *then*

$$(nh_1 \ldots h_p)^{-1/2}\left[\hat{F}(y|\mathbf{x}) - F(y|\mathbf{x}) - \sum_{j=1}^{p} h_j^2 B_j(y,\mathbf{x})\right] \xrightarrow{\mathcal{L}} N(0,\Sigma_{y|\mathbf{x}})$$

Asymptotic normality of conditional quantile estimator

Define

$$\tilde{q}_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) \overset{\mathrm{def}}{=} \arg\min_{q} |\tau - \hat{F}(q|\boldsymbol{x})|$$

$$\hat{q}_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) \overset{\mathrm{def}}{=} \arg\min_{q} |\tau - \hat{F}(q|\boldsymbol{x})|$$

and

$$B_{\tau,j}(y,\boldsymbol{x}) \overset{\mathrm{def}}{=} \frac{B_j(y,\boldsymbol{x})}{f_{Y|\boldsymbol{X}}(q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}))}$$

Asymptotic normality of conditional quantile estimator

Theorem

Assume that the density of $F(y|\boldsymbol{x})$ exists. Under similar condition as in Theorem (A),

$$(nh_1 ... h_p)^{-1/2} \left[\tilde{q}_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) - q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) - \sum_{j=1}^{p} h_j^2 B_{\tau,j}(y, \boldsymbol{x}) \right]$$

$$\xrightarrow{\mathcal{L}} N(0, V_\tau(\boldsymbol{x})),$$

where $V_\tau(\boldsymbol{x}) = \tau(1-\tau)\|K\|_2^{2p} / \left[f^2_{Y|\boldsymbol{X}}(q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x})|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \right]$

Asymptotic normality of conditional quantile estimator

Theorem

Assume that the density of $F(y|\boldsymbol{x})$ exists. Under similar condition as in Theorem (B),

$$(nh_1 \dots h_p)^{-1/2}\left[\tilde{q}_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) - q_{Y|\boldsymbol{X}}(\tau|\boldsymbol{x}) - \sum_{j=0}^{p} h_j^2 B_{\tau,j}(y,\boldsymbol{x})\right] \\ \stackrel{\mathcal{L}}{\to} N(0, V_\tau(\boldsymbol{x})),$$

where $V_\tau(\boldsymbol{x})$ is similar to the last theorem

Kernel Functions

Kernel	$K(u)$	$\|K\|_2^2$	$\kappa_2(K)$
Uniform	$\frac{1}{2}\mathbf{1}(\|u\| \leq 1)$	$1/2$	$1/3$
Epanechnikov	$\frac{3}{4}(1-u^2)\mathbf{1}(\|u\| \leq 1)$	$3/5$	$1/5$
Quartic	$\frac{15}{16}(1-u^2)^2\mathbf{1}(\|u\| \leq 1)$	$5/7$	$1/7$
Triweight	$\frac{35}{32}(1-u^2)^3\mathbf{1}(\|u\| \leq 1)$	$350/429$	$1/9$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$	$1/2\sqrt{\pi}$	1

Table 2: Common Second-Order kernels (symmetric)

Choice of Bandwidth

- The MSE of $\tilde{F}(y|\boldsymbol{x})$ is

$$\mathrm{MSE}(h_1, \ldots, h_p) \sim \Big[\sum_{j=1}^{p} h_j^2 B_j(y, \boldsymbol{x})\Big]^2 + \frac{\Sigma_{y|\boldsymbol{x}}}{n h_1 \ldots h_p}$$

  minimizing the MSE suggests that the bandwidth $h_j \sim n^{1/(4+p)}$

- The MSE of $\hat{F}(y|\boldsymbol{x})$ is

$$\mathrm{MSE}(h_1, \ldots, h_p) \sim \Big[\sum_{j=0}^{p} h_j^2 B_j(y, \boldsymbol{x})\Big]^2 + \frac{\Sigma_{y|\boldsymbol{x}} - h_0 C_K \Omega(y, \boldsymbol{x})}{n h_1 \ldots h_p}$$

  minimizing the MSE suggests $h_j \sim n^{1/(4+p)}$ while $h_0 = \mathcal{O}(h_j)$.

Cross Validation

- Li, Lin and Racine (2013) propose the following cross-validation function:

  $$\mathrm{CV}(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\int \left\{\mathbf{1}(Y_t \leq y) - \hat{F}_{-t}(y|\boldsymbol{X}_t)\right\}^2 \mathcal{M}(\boldsymbol{X}_t)M(y)dy,$$

  where $\hat{F}_{-t}(y|\boldsymbol{X}_t)$ is the leave-one-out estimator of $F(y|\boldsymbol{X}_t)$, defined by

  $$\hat{F}_{-t}(y|\boldsymbol{X}_t) = \left\{\frac{1}{T}\sum_{s\neq t}\mathbf{1}(Y_s \leq y)K_H(\boldsymbol{x} - \boldsymbol{X}_s)\right\}/\hat{f}_{-t}(\boldsymbol{X}_t).$$

- One can apply R package "np"

Local Polynomial Quantile Regression

Definition

Consider $\dim(\boldsymbol{x}) = 1$, for fix x, the local polynomial quantile regression problem with order p is

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \sum_{t=1}^{T} K_h(x_i - x)\rho_\tau\left\{y_i - \beta_0 - \beta_1(x_i - x) - \ldots - \beta_p(x_i - x)^p\right\}$$

where $K_h(x) = h^{-1}K\left(x/h\right)$ and K is a kernel function, $h \rightarrow 0$ and $nh \rightarrow \infty$

- ⊡ The most popular type is $p = 1$, which avoids boundary effect
- ⊡ $\beta_j/j!$ is the estimator for jth derivative of the true function g

Asymptotic Normality

Theorem (Theorem 3 of Fan, Hu and Truong (1994))

1. *(Interior property) Under regularity conditions and* $nh \to \infty$ *and* $h \to 0$. *The local linear estimator* $\hat{l}_\tau$

$$\hat{l}_\tau(x) - l_\tau(x) \xrightarrow{\mathcal{L}} N\left(\beta(x)h^2, \frac{r(x)}{nh}\right),$$

where

$$\beta(x) = \frac{l''_\tau(x)}{2} \int u^2 K(u)du,$$

$$r^2(x) = \frac{\int K^2(u)du}{f_X(x)} \frac{\tau(1-\tau)}{f_{Y|X}(l_\tau(x)|x)^2}.$$

Asymptotic Normality

2 (Boundary behavior) Under regularity conditions,

$$\hat{l}_\tau(x) - l_\tau(x) \xrightarrow{\mathcal{L}} N\left(\beta h^2, \frac{r^2}{nh}\right),$$

where $x_n = ch$, $\beta = 0.5\alpha(c) l''_\tau(0)$,

$$r^2 = \frac{\beta(c)\tau(1-\tau)}{f_{Y|X}(l_\tau(0)|0)^2 f_X(0)}$$

.

Remarks:

⊡ Boundary effect: the convergence rate is of low order on the end points ← local linear estimator resolves this problem

⊡ The bias of local linear QR does not depend on the design ($f_X(x)$)

Choice of bandwidth

- The MSE follows from the theorem

$$\text{MSE}(h) \simeq \frac{1}{4} h^4 \mu_2(K)^2 l''_\tau(x)^2 + \frac{\|K\|_2^2 \tau(1-\tau)}{nhf_X(x) f_{Y|X}(l_\tau(x)|x)^2},$$

  where $\mu_2(K) = \int u^2 K(u) du$.

- The h minimizes the MSE satisfies

$$h_\tau = \frac{\|K\|_2^2 \tau(1-\tau)}{n\mu_2(K)^2 l''_\tau(x)^2 f_X(x) f_{Y|X}(l_\tau(x)|x)^2}$$

- Suppose $f_{Y|X}$ is normal with $f_{Y|X}(l_\tau(x)|x) = \sigma_X^{-1}\phi(\Phi^{-1}(\tau))$

$$\left(\frac{h_{\tau_1}}{h_{\tau_2}}\right)^5 = \frac{\tau_1(1-\tau_1)\phi(\Phi^{-1}(\tau_1))^2}{\tau_2(1-\tau_2)\phi(\Phi^{-1}(\tau_2))^2},$$

Choice of bandwidth

- Hence, Yu and Jones (1998) suggest the rule-of-thumb:

  $$h_\tau = \left[4\tau(1-\tau)\phi(\Phi^{-1}(\tau))^{-2}\right]^{1/5} h_{1/2},$$

  where $h_{1/2}$ is chosen by some standard method for mean
- Cross-validation: quantile estimator can be viewed as a by-product of CDF estimation, it seems reasonable to se- lect bandwidths by a method optimal for CDF estimation

Illustration: motorcycle data

- Experimental measurements of the acceleration of the head in simulated motorcycle accident
- Two variables:
  - times: in milliseconds after impact
  - accel: in "g"
- Note the crossing of the estimated quantile curves beyond about 55 milliseconds
- For the first few milliseconds, variability is almost negligible and gradually increases thereafter

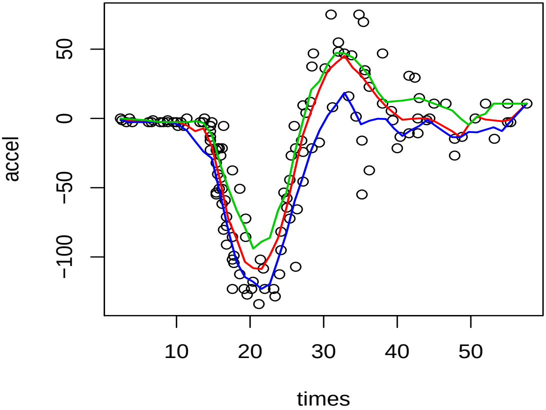


Figure 5: Motorcycle data. Red line: $\tau = 0.5$. Blue line: $\tau = 0.3$. Green line: $\tau = 0.7$. $h = 2$.

Penalty Method

- To penalize the roughness of the fitted function
- Bosch, Ye and Woodworth (1995) consider the problem:

$$\min_{g \in \mathcal{G}} \sum_{t=1}^{T} \rho_\tau(y_t - g(x)) + \lambda \int (g''(x))^2 dx$$

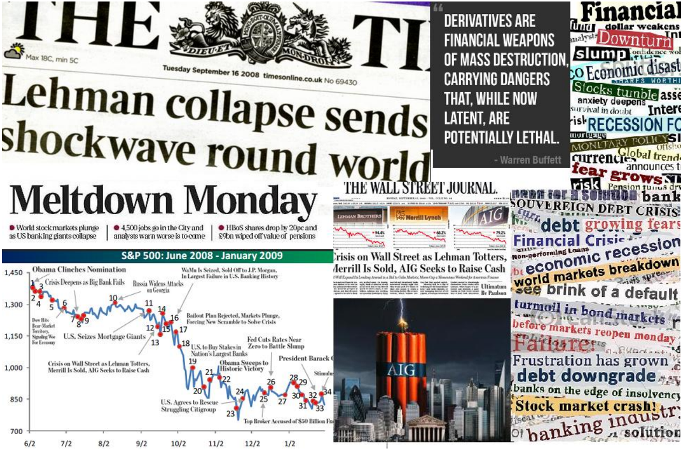
where $\mathcal{G}$ is the Sobolev space of C^2 function with square integrable second derivatives

- Koenker, Ng and Portnoy (1994) consider the L_p penalties:

$$J(g) = \|g''\|_p = \left[\int |g''(x)|^p dx \right]^{1/p},$$

and focus on $p = 1$ and $p = \infty$

The impact of the subprime crisis


THE TI
Max 18C, min 5C
Tuesday September 16 2008 timesonline.co.uk No 69430
Lehman collapse sends shockwave round world
Meltdown Monday
World stock markets plunge as US banking giants collapse
4,500 jobs go in the City and analysts warn worse is to come
HBoS shares drop by 20pc and £9bn wiped off value of pensions
DERIVATIVES ARE FINANCIAL WEAPONS OF MASS DESTRUCTION, CARRYING DANGERS THAT, WHILE NOW LATENT, ARE POTENTIALLY LETHAL.
- Warren Buffett
THE WALL STREET JOURNAL.
LEHMAN BROTHERS
Merrill Lynch
AIG
Crisis on Wall Street as Lehman Totters, Merrill Is Sold, AIG Seeks to Raise Cash
S&P 500: June 2008 - January 2009
Obama Clinches Nomination
Crisis Deepens as Big Bank Fails
Russia Widens Attacks on Georgia
WaMu Is Seized, Sold Off to J.P. Morgan, In Largest Failure in U.S. Banking History
Dow Hits Bear-Market Territory, Signaling Woe For Economy
U.S. Seizes Mortgage Giants
Bailout Plan Rejected, Markets Plunge, Forcing New Scramble to Solve Crisis
U.S. to Buy Stakes in Nation's Largest Banks
Fed Cuts Rates Near Zero to Battle Slump
Crisis on Wall Street as Lehman Totters, Merrill Is Sold, AIG Seeks to Raise Cash
Obama Sweeps to Historic Victory
President Barack
Stimulus
U.S. Agrees to Rescue Struggling Citigroup
Top Broker Accused of $50 Billion Fr
1,450
1,300
1,150
1,000
850
700
6/2
7/2
8/2
9/2
10/2
11/2
12/2
1/2
Financial
dollar weakens
Downturn
slump
Economic disast
Stocks tumble
anxiety deepens
survival in doubt
Interest
RECESSION
MONETARY POLICY
Global trend
currencies
announces
fear grows
risk
Pension funds
bank
SOVEREIGN DEBT CRISIS
debt growing fears
Financial Crisis
Non-performing Loans
economic recession
world markets breakdown
see brink of a default
turmoil in bond markets
before markets reopen monday
Failure
Frustration has grown
debt downgrade
banks on the edge of insolvency
Stock market crash!
banking industry
solution

Value at Risk (VaR)

Value at risk (VaR) has become the standard measure of market risk used by financial institutions and their regulators.

- R. Engle and S. Manganelli (2004)

- VaR is a measure of how much a certain portfolio can lose within a given time period, for a given confidence level

VaR and Basel Accords

- Basel accords: a set of recommendations on banking law and regulation that applies to all banks
- Basel committee on banking supervision: formed by central bankers from G10 countries (now 27) in 1975, in response to the subsequent international financial turmoil followed by the liquidation of Herstatt Bank

VaR and Basel Accords

- Banks are allowed to use internal risk measures on their trading book, but certain rules should be fulfilled
- Backtesting is a technique used to compare the predicted losses from VaR with the actual losses realised at the end of the period of time.
- Key points on backtesting:
  1. Data sets should be updated at least once every 3 months
  2. VaR must be calculated on a daily basis 99th percentile one-tailed confidence interval is to be used
  3. A 10 day movement in prices should be used as the instant price shock
  4. 1 year is classified as a minimum period for "historical" observations

Basel III modifications

- Basel III: commencing on Jan. 1, 2011, with most changes becoming effective within the next six years
- Additional requirement: Banks will be subject to new "stressed" value-at-risk(SVaR) models, increased counterparty risk charges, more restricted netting of offsetting positions, increased charges for exposures to other financial institutions and increased charges for securitisation exposures
- On a daily basis, a bank must meet the capital requirement expressed as the higher of its latest SVaR number and an average of SVaR numbers calculated over the preceding 60 business days multiplied by the multiplication factor

Stressed VaR key points

1. The stressed VaR is computed on a 10-day 99% confidence basis, but with inputs taken from times of significant financial stress relevant to the firm¡s portfolio. Therefore, altogether, in addition to the current requirement of between three to four timesthe 10-day 99% VaR, three times the 10-day 99% SVaR will be required
2. Model inputs are calibrated to historical data from a continuous 12-month period of significant financial stress (equivalent to a VaR measure calculated over a dataset including 2008 and 2009)
3. Data sets update every month and reassess whenever a material change in market prices takes place
4. Risk factors incorporated in pricing models should also be included in VaR calculations and omissions must be justified

Definition (Value-at-Risk)

If y_t is the asset return and $\tau \in (0,1)$, the $\mathrm{VaR}_{t,\tau}$ is defined by

$$\mathrm{P}(y_t < -\mathrm{VaR}_{t,\tau} | \mathcal{F}_{t-1}) = \tau.$$

Definition (Coherent risk measure, Artzner et al. (1999))

A risk measure R is said to be coherent if for portfolios P, P_1, P_2, the followings are satisfied:

1. Translation invariance: for any constant c, $R(P+c) = R(P) - c$
2. Linear homogeneity: for any constant $\lambda > 0$, $R(\lambda P) = \lambda R(P)$
3. Monotonicity: If P_1 stochastically dominates P_2, i.e. $F_{P_1}(y) \leq F_{P_2}(y)$, then $R(P_1) \leq R(P_2)$
4. Subadditivity: $R(P_1 + P_2) \leq R(P_1) + R(P_2)$

VaR violates subadditivity

Consider ε_i and η_i are independent,

$$X_i = \varepsilon_i + \eta_i, \varepsilon_i \overset{\text{iid}}{\sim} N(0,1), \eta_i \overset{\text{iid}}{\sim} \begin{cases} 0, & p = 0.991; \\ -10, & p = 0.009. \end{cases} \qquad i = 1, 2.$$

The 1% VaR for X_1 is 3.1(why?), which is only slightly higher than the VaR if the shocks $\eta = 0$ ($z_{0.01} = -2.3$). X_2 follows the same distribution as asset X_1. Compare

$$P_1 = X_1 + X_2, \quad P_2 = 2X_1.$$

In the former case, the 1% portfolio VaR is 9.8, because for $(X1 + X2)$ the probability of getting the -10 draw for either X_1 or X_2 is higher than 1% $(0.991 * 0.009 * 2 \approx 0.018)$

$$\text{VaR}(P_1) = \text{VaR}(X_1 + X_2) = 9.8 > \text{VaR}(P_2) = 2\text{VaR}(X_1) = 6.2.$$

VaR violates subadditivity

- Artzner et al. (1999): "a merger does not create extra risk", usually it should reduce the risk
- Special case: VaR is globally ($\forall\tau$) subadditive when asset returns are normally distributed, or more generally, log-concave distributed
- Daníelsson et al.(2013) suggests that
  - if the asset returns distribution has jointly fat tail (e.g. student-t with df>1), with very small τ, VaR$_\tau$ is subadditive
  - the occurrence of subadditivity depends on the estimation method, sample size
- Despite of non-subadditivity, VaR still prevails because:
  - smaller data requirement
  - ease for backtesting

Stylized facts of asset returns

- What is a stylized fact? Nontrivial statistical properties commonly shared by random variations of asset prices in different markets and instruments
- stylized facts are usually formulated in terms of qualitative properties of asset returns and may not be precise enough to distinguish among different parametric models
- It is not easy to exhibit even an (ad hoc) stochastic process which possesses the same set of properties and one has to go to great lengths to reproduce them with a model

Stylized facts of asset returns

Some common stylized facts of asset returns (Cont (2001)):

1. **Absence of autocorrelations**: except for intraday data
2. **Heavy tails**: the (unconditional) distribution of returns seems to display a power-law
3. **Gain/loss asymmetry**: large losses are observed but no equally large gain
4. **Aggregational Gaussianity**: as the time scale Δt increases, the calculated returns look more normally distributed
5. **Intermittency**: at any time scale one observes high degree of variability
6. **Volatility clustering**: volatility shows a positive autocorrelation over several days, high volatility events tend to cluster in time

Stylized facts of asset returns

7. **Conditional heavy tails**: even correcting returns for volatility clustering, the residual time series still exhibit heavy tails
8. **Slow decay of autocorrelation in absolute returns**: the autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent $[0.2, 0.4]$
9. **Leverage effect**: volatility negatively correlated with the (previous) returns
10. **Volume/volatility correlation**: trading volume is correlated with all measures of volatility
11. **Asymmetry in time scales**: volatility measure with Δt large predicts Δt small volatility better than the other way around

ARCH review

The AutoRegressive Conditional Heteroskedasticity (ARCH) model for asset return modeling of Engle (1982):

- $y_t = \sqrt{h_t}\varepsilon_t$, where ε_t are i.i.d. with mean 0 and variance 1,

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2, \quad \alpha_0 > 1, \alpha_1 \geq 0.$$

- $\mathrm{E}[y_t|\mathcal{F}_{t-1}] = 0$ and $\mathrm{E}[y_t^2|\mathcal{F}_{t-1}] = h_t$
- $\mathrm{E}[y_t y_s] = 0$ for $t \neq s$ (why?)
- y_t^2 are serially correlated with AR(1):

$$y_t^2 = h_t + (y_t^2 - h_t) = \alpha_0 + \alpha_1 y_{t-1}^2 + \tilde{\varepsilon}_t,$$

where $\tilde{\varepsilon}_t = h_t \varepsilon_t^2 - 1$ are innovations with mean and covariance zero

ARCH review

- Assume that $\varepsilon_t \sim N(0,1)$, $\mathrm{E}[y_t^4|\mathcal{F}_{t-1}] = 3h_t^2$, and

$$m_4 \overset{\text{def}}{=} \mathrm{E}[y_t^4] = 3\big[\alpha_0^2 + 2\alpha_0\alpha_1\,\mathrm{E}[h_t] + \alpha_1^2\,\mathrm{E}(y_{t-1}^4)\big]$$
$$= 3\alpha_0^2\left(1 + \frac{2\alpha_0}{1-\alpha_1}\right) + 3\alpha_1 m_4,$$

  hence,

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}.$$

  In order to make m_4 well-defined, it is required that $0 \le \alpha_1^2 < 1/3$

- y_t are leptokurtic because the kurtosis of y_t is

$$\frac{m_4}{\mathrm{Var}(y_t)^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3, \text{ if } \alpha_1 \neq 0.$$

GARCH review

The Generalized ARCH (GARCH) model of Bollerslev (1986):

- GARCH(1,1): $y_t = \sqrt{h_t}\varepsilon_t$, with

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}, \quad \alpha_0 > 0,\ \alpha_1, \beta_1 \geq 0.$$

- y_t^2 have an ARMA(1,1) representation:

$$\begin{aligned} y_t^2 &= h_t + (y_t^2 - h_t) \\ &= \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 + h_t(\varepsilon_t^2 - 1) - \beta_1 h_{t-1}(\varepsilon_{t-1}^2 - 1), \end{aligned}$$

where $h_t(\varepsilon_t^2 - 1)$ can be viewed as serially uncorrelated innovations

GARCH review

- y_t has mean zero, and since ε_t is independent of h_t,

  $$\mathrm{Var}(y_t) = \mathrm{E}[y_t^2] = \mathrm{E}[h_t \varepsilon_t^2] = \mathrm{E}[h_t]\,\mathrm{E}[\varepsilon_t^2] = \mathrm{E}[h_t].$$

  Suppose h_t is stationary,

  $$\mathrm{E}[h_t] = \mathrm{Var}(y_t) = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}.$$

- y_t and y_s are uncorrelated for $t \neq s$
- If $\varepsilon_t \sim N(0,1)$,

  $$\frac{m_4}{\mathrm{Var}(y_t)^2} = 3\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} > 3,$$

  provided that $1-(\alpha_1+\beta_1)^2-2\alpha_1^2 > 0$

EGARCH

Exponential GARCH (EGARCH) model of Nelson (1992)

- $y_t = \sqrt{h_t}\varepsilon_t$, with

$$h_t = \exp\left[\alpha_0 + \beta_1 \log(h_{t-1}) + \left(\theta_1 \frac{y_{t-1}}{\sqrt{h_{t-1}}} + \gamma_1 \left|\frac{y_{t-1}}{\sqrt{h_{t-1}}}\right|\right)\right].$$

- θ_1 is interpreted as a measure of "leverage" effect, while γ_1 is interpreted as the "magnitude" effect. θ_1 tends to be negative empirically, while γ_1 tends to be positive. $\Rightarrow$ Negative shock has more impact on valitility than positive shock
- Due to the exponential transform, there is no constraint on the coefficients in h_t

News Impact Curve

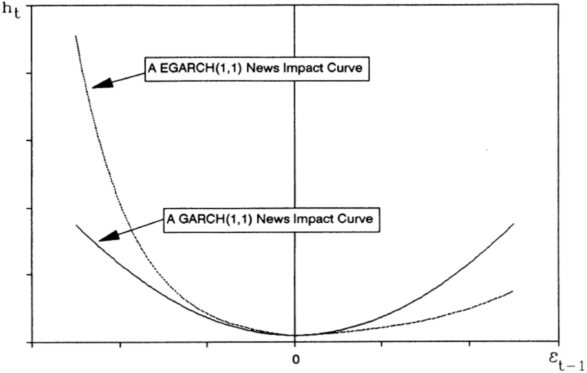


Figure 6: News impact curve. Source: Engle and Ng (1993) Figure 1, p.1754

CAViaR model

The Conditional Autoregressive Value at Risk (CAViaR) of Engle and Manganelli (2004): a family of time series models for quantile

(a) Symmetric absolute value:

$$q_{\tau,t}(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{\tau,t-1}(\boldsymbol{\beta}) + \beta_3 |y_{t-1}|.$$

(b) Asymmetric slope:

$$q_{\tau,t}(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{\tau,t-1}(\boldsymbol{\beta}) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^-.$$

(c) Indirect GARCH(1,1):

$$q_{\tau,t}(\boldsymbol{\beta}) = \left[\beta_1 + \beta_2 q_{\tau,t-1}(\boldsymbol{\beta})^2 + \beta_3 (y_{t-1})^2\right]^{1/2}.$$

(d) Adaptive: for $0 < G < \infty$,

$$q_{\tau,t}(\beta_1) = q_{\tau,t-1}(\beta_1) + \beta_1 \left\{ \frac{1}{1 + \exp\left(G[y_{t-1} - q_{\tau,t-1}(\beta_1)]\right)} - \tau \right\}$$

CAViaR model

(a) Symmetric absolute value:

$$q_{\tau,t}(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{\tau,t-1}(\boldsymbol{\beta}) + \beta_3 |y_{t-1}|.$$

(b) Asymmetric slope:

$$q_{\tau,t}(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{\tau,t-1}(\boldsymbol{\beta}) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^-.$$

Symmetric absolute value and asymmetric slope are induced from GARCH with standard deviation is modeled symmetrically or asymmetrically.

CAViaR model

(c) Indirect GARCH$(1,1)$:

$$q_{\tau,t}(\boldsymbol{\beta}) = \left[\beta_1 + \beta_2 q_{\tau,t-1}(\boldsymbol{\beta})^2 + \beta_3 (y_{t-1})^2\right]^{1/2}.$$

Indirect GARCH is correctly specified if the DGP is GARCH(1,1) with an i.i.d. error distribution. To see this, note that $y_t = \sqrt{h_t}\varepsilon_t$, so

$$\begin{aligned} q_{\tau,t} &= \sqrt{h_t} q_{\varepsilon,\tau} = \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}} q_{\varepsilon,\tau} \\ &= \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 q_{\tau,t-1}^2} \end{aligned}$$

CAViaR model

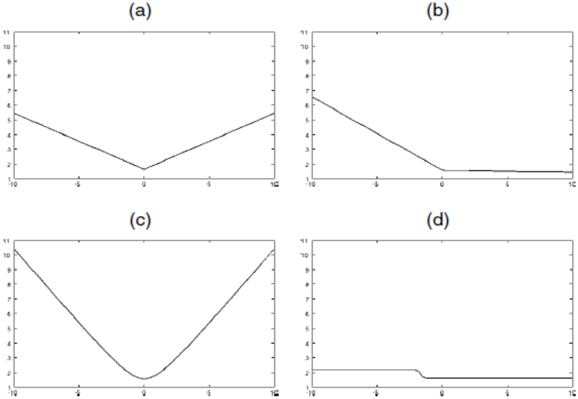
(d) Adaptive: for $0 < G < \infty$,

$$q_{\tau,t}(\beta_1) = q_{\tau,t-1}(\beta_1) + \beta_1 \left\{ \underbrace{\frac{1}{1 + \exp\left(G[y_{t-1} - q_{\tau,t-1}(\beta_1)]\right)}}_{(*)} - \tau \right\}$$

When $G \to \infty$, $(*) \to \mathbf{1}\{y_{t-1} \le q_{\tau,t-1}(\beta_1)\}$, given that τ is small:

- ⊡ When $y_{t-1} \le q_{\tau,t-1}(\beta_1)$, $q_{\tau,t}(\beta_1) = q_{\tau,t-1}(\beta_1) + \beta_1(1 - \tau)$
- ⊡ When $y_{t-1} > q_{\tau,t-1}(\beta_1)$, $q_{\tau,t}(\beta_1) = q_{\tau,t-1}(\beta_1) - \beta_1\tau$

CAViaR news impact



source: Engle and Manganelli (2004) Figure 2

Estimation

- The objective function is:

$$Q_T(\boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\{y_t - q_t(\boldsymbol{\beta})\},$$

$$\hat{\boldsymbol{\beta}} \overset{\mathrm{def}}{=} \operatorname*{arg\,min}_{\boldsymbol{\beta}} Q_T(\boldsymbol{\beta}).$$

- Difficulty: this is a dynamic model. $q_t(\boldsymbol{\beta})$ depends on $q_{t-1}(\boldsymbol{\beta})$.
- Manganelli provides the code on his website

Asymptotic Theory

Consider the model

$$\begin{aligned} y_t &= f(y_{t-1}, \boldsymbol{x}_{t-1}, ..., y_1, \boldsymbol{x}_1; \boldsymbol{\beta}^0) + \varepsilon_t \\ &\stackrel{\mathrm{def}}{=} f_t(\boldsymbol{\beta}^0) + \varepsilon_t, \quad t = 1, ..., T, \end{aligned} \tag{5}$$

where $q_\tau(\varepsilon_t | \mathcal{F}_t) = 0$.

Theorem (C)

In model (5)*, under assumption C0-C7,* $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}^0$*, where* $\hat{\boldsymbol{\beta}}$ *is the solution* $\min_{\boldsymbol{\beta}} Q_T(\boldsymbol{\beta})$*.*

▸ C0-C7

Asymtptic Theory

Theorem (AN)

In model (5)*, under assumptions AN1-AN4 and the conditions of Theorem C,*

▸ AN1-AN4

$$\sqrt{T}\boldsymbol{A}_T^{1/2}\boldsymbol{D}_T(\hat{\beta}-\beta^0) \xrightarrow{d} N(0,1),$$

where

$$\boldsymbol{A}_T = \mathsf{E}\left[T^{-1}\tau(1-\tau)\sum_{t=1}^{T}\nabla f_t(\beta^0)^\top \nabla f_t(\beta^0)\right],$$

$$\boldsymbol{D}_T = \mathsf{E}\left[T^{-1}\tau(1-\tau)\sum_{t=1}^{T}\nabla g_t(0|\mathcal{F}_t) f_t(\beta^0)^\top \nabla f_t(\beta^0)\right],$$

and $\hat{\beta}$ is computed as in Theorem C.

Need to show two things:

Asymptotc FOC: Let $\mathrm{Hit}_t(\boldsymbol{\beta}^0) \overset{\mathrm{def}}{=} \mathbf{1}\{y_t < f_t(\boldsymbol{\beta}^0)\} - \tau$, $Q_j(\delta) \overset{\mathrm{def}}{=} -T^{-1/2} \sum_{t=1}^{T} \rho_\tau\{y_t - f_t(\boldsymbol{\beta} + \delta \boldsymbol{e}_j)\}$ and $G_j(\delta) \overset{\mathrm{def}}{=} -T^{-1/2} \sum_{t=1}^{T} \nabla_j f_t(\hat{\boldsymbol{\beta}} + \delta \boldsymbol{e}_j)\mathbf{Hit}_t(\hat{\boldsymbol{\beta}} + \delta \boldsymbol{e}_j)$, where $\boldsymbol{e}_j$ are standard basis of $\mathbb{R}^p$. Because $Q_j(\delta)$ is continuous in δ and achieves a maximum at 0,

$$|G_j(0)| \leq G_j(\delta) - G_j(\delta)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\nabla_j f_t(\hat{\boldsymbol{\beta}} + \delta \boldsymbol{e}_j) \mathrm{Hit}_t(\hat{\boldsymbol{\beta}} + \delta \boldsymbol{e}_j) + \nabla_j f_t(\hat{\boldsymbol{\beta}} - \delta \boldsymbol{e}_j) \mathrm{Hit}_t(\hat{\boldsymbol{\beta}} - \delta \boldsymbol{e}_j) \right]$$

Letting $\delta \to 0$,

$$
\begin{aligned}
|G_j(0)| &\leq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left|\nabla_j f_t(\hat{\boldsymbol{\beta}})\right| \mathbf{1}\left\{y_t = f_t(\hat{\boldsymbol{\beta}})\right\} \\
&\leq \frac{1}{\sqrt{T}} \Big[\max_{1 \leq t \leq T} H(\mathcal{F}_t)\Big] \sum_{t=1}^{T} \mathbf{1}\big(y_t = f_t(\hat{\boldsymbol{\beta}})\big).
\end{aligned}
$$

Assumption AN1 implies $T^{-1/2}[\max_{1\leq t\leq T} H(\mathcal{F}_t)] =_P (1)$ and C2 implies $\sum_{t=1}^{T} \mathbf{1}(y_t = f_t(\hat{\boldsymbol{\beta}})) = \mathcal{O}(1)$ a.s. because the $\{\varepsilon_t = 0\}$ has probability 0.

Stochastic equicontinuous condition: ▸ SEC

N2 : AN1 implies $\boldsymbol{\beta}^0$ lies in the interior of B.
$\mathrm{E}[q_t(\boldsymbol{\beta}^0)] = \mathrm{E}\left[\mathrm{E}[\mathrm{Hit}_t(\boldsymbol{\beta}^0)|\mathcal{F}_t]\nabla f_t(\boldsymbol{\beta}^0)^\top\right] = 0.$

N3 : too technical

Asymtptic Theory

Theorem (VC)

Under assumptions VC1-VC3 and the conditions of Theorems C and AN, $\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T \xrightarrow{P} 0$ *and* $\hat{\boldsymbol{D}}_T - \boldsymbol{D}_T \xrightarrow{P} 0$, *where*

▸ VC1-VC3

$$\hat{\boldsymbol{A}}_T = T^{-1}\tau(1-\tau)\nabla f_t(\hat{\beta})^\top \nabla f_t(\hat{\beta})$$

$$\hat{\boldsymbol{D}}_T = (2T\hat{c}_T)^{-1}\sum_{t=1}^{T} \mathbf{1}(|y_t - f_t(\hat{\beta})| < \hat{c}_T)\nabla f_t(\hat{\beta})^\top \nabla f_t(\hat{\beta}),$$

$\boldsymbol{A}_T$ *and* $\boldsymbol{D}_T$ *are defined in Theorem AN, and* $\hat{c}_T$ *is a bandwidth defined in assumption VC1.*

Backtesting

- Define the hit sequence

$$\mathrm{Hit}_t(\beta^0) \overset{\text{def}}{=} \mathbf{1}\big(y_t < f_t(\beta^0)\big) - \tau$$

- A natural way to test the validity of the forecast model is to check whether the sequence Hit_t is i.i.d.
- This concept has a drawback: define a i.i.d. sequence (like flipping a coin)

$$z_t = \begin{cases} 1, & \text{with probability } \tau; \\ -1, & \text{with probability } 1-\tau. \end{cases}$$

Setting $f_t(\beta^0) = K z_{t-1}$ for K large. Once z_{t-1} is observed, the probability of y_t exceeding $f_t(\beta)$ is almost 0 or 1

Testing in-sample fit

Define

$$\boldsymbol{X}_t(\hat{\beta}) \in \mathbb{R}^q : \text{ the typical row of } \boldsymbol{X}_t(\hat{\beta}) \in F_t,$$

$$\mathbf{Hit}_t(\hat{\beta}) \stackrel{\text{def}}{=} (\mathrm{Hit}_1(\hat{\beta}), ..., \mathrm{Hit}_T(\hat{\beta}))^\top$$

$$\boldsymbol{M}_T \stackrel{\text{def}}{=} \boldsymbol{X}^\top(\beta^0) - \mathsf{E}\left[T^{-1}\boldsymbol{X}^\top(\beta^0)\boldsymbol{H}\nabla f(\beta^0)\right]\boldsymbol{D}_T^{-1}\nabla^\top f(\beta^0),$$

$$\boldsymbol{H} = \mathrm{diag}\big(g_1(0|\mathcal{F}_1), ..., g_T(0|\mathcal{F}_T)\big)$$

Theorem (In-sample dynamic quantile test)

Under the asuumption of Theorem C and AN and assumptions DQ1-DQ6,

▸ DQ1-DQ7

$$\left[\tau(1-\tau)\,\mathsf{E}[T^{-1}\boldsymbol{M}_T\boldsymbol{M}_T^\top]\right]^{-1/2} T^{-1/2}\boldsymbol{X}^\top(\hat{\beta})\boldsymbol{X}(\hat{\beta}) \overset{\mathsf{d}}{\sim} N(0,I).$$

If assumption DQ7 and the conditions of Theorem VC also hold, then

$$DQ_{IS} \overset{\text{def}}{=} \frac{\boldsymbol{Hit}^\top(\hat{\beta})\boldsymbol{X}(\hat{\beta})(\hat{\boldsymbol{M}}_T\hat{\boldsymbol{M}}_T^\top)^{-1}\boldsymbol{X}^\top(\hat{\beta})\boldsymbol{Hit}^\top(\hat{\beta})}{\tau(1-\tau)} \overset{\mathsf{d}}{\sim} \chi^2_q, \quad T\to\infty,$$

where

$$\hat{\boldsymbol{M}}_T \overset{\text{def}}{=}$$

$$\boldsymbol{X}^\top(\hat{\beta}) - \left\{(2T\hat{c}_T)^{-1}\sum_{t=1}^{T}\boldsymbol{1}\left(|y_t - f_t(\hat{\beta})| < \hat{c}_T\right)\boldsymbol{X}_t^\top(\hat{\beta})\nabla f_t(\hat{\beta})\right\}\hat{\boldsymbol{D}}_T^{-1}\nabla f(\hat{\beta})^\top.$$

Out-of-sample test

Define

T_R : no. in-sample obs

N_R : no. out-of-sample obs.

$\boldsymbol{X}_n(\hat{\boldsymbol{\beta}}_{T_R}) \in \mathbb{R}^q$: the typical row of $\boldsymbol{X}_t(\hat{\boldsymbol{\beta}}_{T_R}) \in F_t$,

$$n = T_R + 1, ..., T_R + N_R$$

$$\textbf{Hit}_t(\hat{\boldsymbol{\beta}}_{T_R}) \overset{\text{def}}{=} \left(\text{Hit}_{T_R+1}(\hat{\boldsymbol{\beta}}_{T_R}), ..., \text{Hit}_{T_R+N_R}(\hat{\boldsymbol{\beta}}_{T_R})\right)^\top$$

Theorem (Out-of-sample dynamic quantile test)

Under the assumption of Theorem C and AN and assumptions DQ1-DQ3, DQ8 and DQ9,

▸ DQ8-DQ9

$$DQ_{OOS} \stackrel{\mathrm{def}}{=} N_R^{-1} \boldsymbol{Hit}^\top(\hat{\beta}_{T_R}) \boldsymbol{X}(\hat{\beta}_{T_R}) \left[\boldsymbol{X}(\hat{\beta}_{T_R})^\top \boldsymbol{X}(\hat{\beta}_{T_R})\right]^{-1} \boldsymbol{X}^\top(\hat{\beta}_{T_R}) \boldsymbol{Hit}^\top(\hat{\beta}_{T_R}) / [\tau(1-\tau)] \overset{\mathsf{d}}{\sim} \chi^2_q, \quad R \to \infty$$

- ⊡ The In-sample DQ test is useful for model selection
- ⊡ The out-of-sample DQ test is useful for regulators to check whether the VaR estimates satisfy basic requirements, such as unbiasedeness, independent hits, and independence of the quantile estimates

Application

- 3392 daily prices (April 7, 1986-April 7, 1999) of GM, IBM and S&P500
- Estimation: first 2892 obs. last 500 for backtesting
- $G = 10$ in adaptive model

Quantile Regression: Primary Techniques

Shih-Kang Chao

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

A1:Convex Analysis

Definition (Subdifferetial/subgradient)

g is called a subdifferetial/subgradient of a convex function f at $\boldsymbol{x} \in \mathrm{dom} f$ if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + g^\top(\boldsymbol{y} - \boldsymbol{x}).$$

Remark: if f is differentiable, then its subdifferential is unique.

Theorem

If f is convex. Its composition with an affine function is again convex; namely, $f(\alpha + \boldsymbol{\beta}^\top \boldsymbol{x})$ is convex.
Remark: composition with arbitrary function may lose convexity.

Theorem

The subgradient obeys chain rule.

A1:Convex Analysis

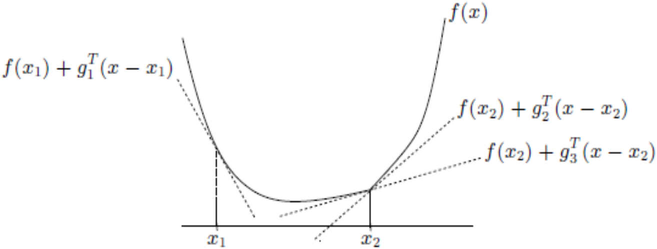

$f(x)$
$f(x_1) + g_1^T(x - x_1)$
$f(x_2) + g_2^T(x - x_2)$
$f(x_2) + g_3^T(x - x_2)$
x_1
x_2

Figure 7: g_1 is the unique subgradient at x_1. g_2 and g_3 are subgradients at x_2.

C0 (Ω, F, P) is acomplete probability space, and $\{\varepsilon_t, \boldsymbol{x}_t : t = 1, ...\}$ are random vectors on this space

C1 $f_t(\boldsymbol{\beta}) : \mathbb{R}^{k_t} \times B \rightarrow \mathbb{R}$ is such that for each $\boldsymbol{\beta} \in B$, a compact subset of $\mathbb{R}^p$, $f_t(\boldsymbol{\beta})$ is measurable with respect to the information set $\mathcal{F}_t$ and $f_t(\boldsymbol{\beta})$ is continuous in B, $t = 1, 2, ...$, for a given choice of explanatory variables $\{y_{t-1}, \boldsymbol{x}_{t-1}, ..., y_1, \boldsymbol{x}_1\}$

C2 Conditional on all the past information $\mathcal{F}_t$, the error terms ε_t form a stationary process, with continuous conditional density $g_t(\varepsilon|\mathcal{F}_t)$

C3 There exists $h > 0$ such that for all t, $g_t(0|\mathcal{F}_t) \geq h$

C4 $|f_t(\boldsymbol{\beta})| < K(\mathcal{F}_t)$ for each $\boldsymbol{\beta} \in B$ and for all t, where $K(\mathcal{F}_t)$ is some (possibly) stochastic function of variables that belong to the information set, such that $\mathsf{E}[|K(\mathcal{F}_t)|] \leq K_0 < \infty$

C5 $\mathsf{E}[|K(\mathcal{F}_t)|] < \infty$ for all t

C6 $\rho_\tau\big(y_t - f_t(\boldsymbol{\beta})\big)$ obeys the unifrom law of large numbers

C7 $\forall \xi > 0$, there exists a $\delta > 0$, such that if $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \geq \xi$, then

$$\liminf_{T\to\infty} T^{-1} \sum \mathsf{P}\left[|f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0)| > \delta\right] > 0$$

▸ Consistency

AN1 $f_t(\boldsymbol{\beta})$ is differentiable in B and $\forall \boldsymbol{\beta}$ and γ in a neighborhood ν_0 of $\boldsymbol{\beta}^0$, such $\|\boldsymbol{\beta} - \gamma\| \leq d$ for d sufficiently small and for all t:

- a $\|\nabla f_t(\boldsymbol{\beta})\| \leq H(\mathcal{F}_t)$, where $H(\mathcal{F}_t)$ is some (possibly) stochastic function of variables that belong to the information set and $\mathsf{E}\left[H(\mathcal{F}_t)^3\right] \leq H_0 < \infty$, for some constant H_0.
- b $\|\nabla f_t(\boldsymbol{\beta}) - \nabla f_t(\gamma)\| \leq M(\mathcal{F}_t, \boldsymbol{\beta}, \gamma) = \mathcal{O}(\|\boldsymbol{\beta} - \gamma\|)$, where $M(\mathcal{F}_t, \boldsymbol{\beta}, \gamma)$ is some function such that $\mathsf{E}[M(\mathcal{F}_t, \boldsymbol{\beta}, \gamma)]^2 \leq M_0\|\boldsymbol{\beta} - \gamma\| < \infty$ and $\mathsf{E}[M(\mathcal{F}_t, \boldsymbol{\beta}, \gamma)H(\mathcal{F}_t)] \leq M_1\|\boldsymbol{\beta} - \gamma\| < \infty$ for some constants M_0 and M_1

AN2

- a $g_t(\varepsilon|\mathcal{F}_t) \leq N < \infty \; \forall t$, for some constant N
- b $g_t(\varepsilon|\mathcal{F}_t)$ satisfies the Lipschitz condition $|g_t(\lambda_1|\mathcal{F}_t) - g_t(\lambda_2|\mathcal{F}_t)| \leq L|\lambda_1 - \lambda_2|$ for some constant $L < \infty \; \forall t$

AN3 The matrices $\boldsymbol{A}_T$ and $\boldsymbol{D}_T$ have the smallest eigenvalues bounded below by a positive constant for T sufficiently large

AN4 The sequence

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\tau - \mathbf{1}(y_t < f_t(\beta^0))\right] \nabla f_t(\beta^0)^\top$$

obeys the central limit theorem.

▸ Asym. Normality

VC1 $\hat{c}_T / c_T \xrightarrow{\text{P}} 1$, where the nonstochastic positive sequence c_T satisfies $c_T = (1)$ and $c_T^{-1} = (T^{1/2})$.

VC2 $\mathsf{E}\left[|H(\mathcal{F}_t)|^4\right] \leq H_1 < \infty$ for all t and for some constant H_1, where $H(\mathcal{F}_t)$ has been defined in assumption AN1(a)

VC3

$$T^{-1}\tau(1-\tau)\nabla f_t(\boldsymbol{\beta}^0)^\top \nabla f_t(\boldsymbol{\beta}^0) - \boldsymbol{A}_T \xrightarrow{\text{P}} 0$$

$$T^{-1}\sum_{t=1}^{T} g_t(0|\mathcal{F}_t)\nabla f_t(\boldsymbol{\beta}^0)^\top \nabla f_t(\boldsymbol{\beta}^0) - \boldsymbol{D}_T \xrightarrow{\text{P}} 0.$$

▸ VC

DQ1 $\boldsymbol{X}_t(\boldsymbol{\beta})$ is different element wise from $\nabla f_t(\boldsymbol{\beta})$, is measurable $\mathcal{F}_t$, $\|\boldsymbol{X}_t(\boldsymbol{\beta})\| \leq W(\mathcal{F}_t)$, where $W(\mathcal{F}_t)$ is some (possibly) stochastic function of variables that belong to the information set, such that $\mathsf{E}\left[W(\mathcal{F}_t)M(\mathcal{F}_t,\boldsymbol{\beta},\boldsymbol{\gamma})\right] \leq W_0\|\boldsymbol{\beta}-\boldsymbol{\gamma}\| < \infty$ and $\mathsf{E}\left[\{W(\mathcal{F}_t)H(\mathcal{F}_t)\}^2\right] < W_1 < \infty$ for some finite constant W_1, and $H(\mathcal{F}_t)$ and $M(\mathcal{F}_t,\boldsymbol{\beta},\boldsymbol{\gamma})$ are defined in AN1.

DQ2 $\|\boldsymbol{X}_t(\boldsymbol{\beta}) - \boldsymbol{X}_t(\boldsymbol{\gamma})\| \leq S(\mathcal{F}_t,\boldsymbol{\beta},\boldsymbol{\gamma})$, where
$\mathsf{E}[S(\mathcal{F}_t,\boldsymbol{\beta},\boldsymbol{\gamma})] \leq S_0\|\boldsymbol{\beta}-\boldsymbol{\gamma}\| < \infty$,
$\mathsf{E}\left[W(\mathcal{F}_t)S(\mathcal{F}_t,\boldsymbol{\beta},\boldsymbol{\gamma})\right] \leq S_1\|\boldsymbol{\beta}-\boldsymbol{\gamma}\| < \infty$, and for some constant S_0, S_1.

DQ3 Let $\{\varepsilon_t^1, ..., \varepsilon_t^{J_i}\}$ be the set of values for which $\boldsymbol{X}_t(\beta)$ is not differentiable. Then $\mathsf{P}(\varepsilon_t = \varepsilon_t^j) = 0$ for $j = 1, ..., J_i$. Whenever the derivative exists, $\|\nabla \boldsymbol{X}_t(\beta)\| \leq Z(\mathcal{F}_t)$, where $Z(\mathcal{F}_t)$ is some (possibly) stochastic function of variables that belong to the information set, such that $\mathsf{E}[Z(\mathcal{F}_t)^r] < Z_0 < \infty$, $r = 1, 2$, for some constant Z_0

DQ4 $T^{-1}\boldsymbol{X}_t^\top(\beta^0)\boldsymbol{H}\nabla f(\beta^0) - \mathsf{E}\left[T^{-1}\boldsymbol{X}^\top(\beta^0)\boldsymbol{H}\nabla f(\beta^0)\right] \xrightarrow{\mathsf{P}} 0.$

DQ5 $T^{-1}\boldsymbol{M}_T\boldsymbol{M}^\top - T^{-1}\,\mathsf{E}[\boldsymbol{M}_T\boldsymbol{M}_T^\top] \xrightarrow{\mathsf{P}} 0$

DQ6 The sequence $\{T^{-1/2}\boldsymbol{M}_T\mathbf{Hit}(\beta^0)\}$ obeys the central limit theorem

DQ7 $T^{-1}\,\mathsf{E}[\boldsymbol{M}_T\boldsymbol{M}_T^\top]$ is a nonsingular matrix

DQ8 $\lim_{R\to\infty} T_R = \infty$, $\lim_{R\to\infty} N_R = \infty$, and $\lim_{R\to\infty} N_R/T_R = 0$

DQ9 The sequence $\{N_R^{-1/2}\boldsymbol{X}^\top(\beta^0)\mathbf{Hit}(\beta^0)\}$ obeys the central limit theorem.

▸ In-sample test ▸ Out-of-sample test