Introduction to Quantile Regression

CHUNG-MING KUAN

Department of Finance & CRETA

National Taiwan University

June 13, 2011

Lecture Outline

Lecture Outline (cont'd)

5. Estimation of Asymptotic Covariance Matrix

6. Hypothesis Testing
   - Wald Tests
   - Likelihood Ratio Tests

7. Treatment Effect
   - Average Treatment Effect (ATE)
   - Quantile Treatment Effect (QTE)

8. Empirical Applications
   - Return-Volume Relations
   - Effects of NHI on Saving

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

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

  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.

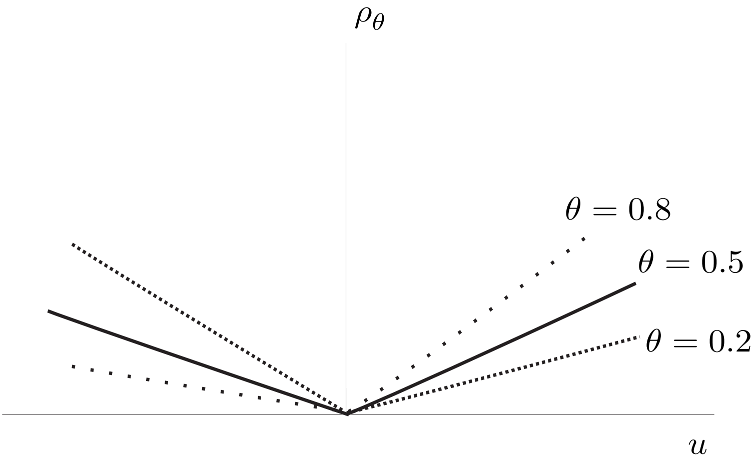


Figure: Check function $\rho_\theta(u) = (\theta - 1_{\{u<0\}})u$.

Quantile Regression (QR) Method

Koenker and Basset (1978)

Given $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$, the θth 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 θth conditional quantile function $Q_{y|\mathbf{x}}(\theta)$, with $\hat{\beta}_i(\theta)$ the estimated marginal effect of the ith regressor on $Q_{y|\mathbf{x}}(\theta)$.

Finding the Solution to V_T

- Difficulties in estimation:
  - The QR estimator $\hat{\boldsymbol{\beta}}(\theta)$ 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

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

  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

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

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

$$\min_{\mathbf{z}} \frac{1}{T}\mathbf{c}^\prime\mathbf{z}, \qquad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{z}, \;\; \mathbf{z} \geq 0.$$

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

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

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

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

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

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 $\boldsymbol{\beta}_o$,

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

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 $\hat{\beta}(\theta)$ satisfies the "asymptotic FOC":

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

- The (approximate) estimating function is thus

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

- The expectation of the estimating function is

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

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

Asymptotic Distribution

- When integration and differentiation can be interchanged,

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

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

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

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

  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

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

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,

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

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

The Null Distribution of the Wald Test

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

where $\widehat{\mathbf{\Gamma}}(\theta) = \mathbf{R}\widehat{\mathbf{D}}(\theta)\mathbf{R}'$, 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

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

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 $f_{e_{0.5}}(0) = 1/4$.

Average Treatment Effect (ATE)

- 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 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

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

- 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 (QTE)

- 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 θth 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

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

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

Empirical Study: Return-Volume Relations

- Granger non-causality is defined in terms of distribution.
- Existing tests focus on its implications:
  - Non-causality in mean (linear model): Granger (1969, 1980).
  - Non-causality in variance: Granger et al. (1986), Cheung and Ng (1996).
  - Nonlinear causality: Hiemstra and Jones (1994).
  - Non-causality in some quantiles: Lee and Yang (2006), Hong et al. (2006).
- Empirical studies on return-volume relations are typically based on least-squares regressions and find that volume does not Granger cause return.

Notions of Granger Non-Causality

- x does not Granger cause y in distribution if

$$F_{y_t}(\eta|(\mathcal{Y},\mathcal{X})_{t-1}) = F_{y_t}(\eta|\mathcal{Y}_{t-1}), \quad \forall \eta \in \mathbb{R}.$$

- x does not Granger cause y in mean if

$$\mathbb{E}[y_t|(\mathcal{Y},\mathcal{X})_{t-1}] = \mathbb{E}(y_t|\mathcal{Y}_{t-1}).$$

- x does not Granger cause y in all quantiles if

$$Q_{y_t}(\tau \big| (\mathcal{Y},\mathcal{X})_{t-1}) = Q_{y_t}(\tau \big| \mathcal{Y}_{t-1}), \quad \forall \tau \in [0,1];$$

cf. Lee and Yang (2006), Hong et al. (2006).

Notions of Granger Non-Causality

- x does not Granger cause y in distribution if

$$F_{y_t}(\eta|(\mathcal{Y},\mathcal{X})_{t-1}) = F_{y_t}(\eta|\mathcal{Y}_{t-1}), \quad \forall \eta \in \mathbb{R}.$$

- x does not Granger cause y in mean if

$$\mathbb{E}[y_t|(\mathcal{Y},\mathcal{X})_{t-1}] = \mathbb{E}(y_t|\mathcal{Y}_{t-1}).$$

- x does not Granger cause y in all quantiles if

$$Q_{y_t}(\tau\big|(\mathcal{Y},\mathcal{X})_{t-1}) = Q_{y_t}(\tau\big|\mathcal{Y}_{t-1}), \quad \forall \tau \in [0,1];$$

cf. Lee and Yang (2006), Hong et al. (2006).

Notions of Granger Non-Causality

- x does not Granger cause y in distribution if

$$F_{y_t}(\eta|(\mathcal{Y},\mathcal{X})_{t-1}) = F_{y_t}(\eta|\mathcal{Y}_{t-1}), \quad \forall \eta \in \mathbb{R}.$$

- x does not Granger cause y in mean if

$$\mathbb{E}[y_t|(\mathcal{Y},\mathcal{X})_{t-1}] = \mathbb{E}(y_t|\mathcal{Y}_{t-1}).$$

- x does not Granger cause y in all quantiles if

$$Q_{y_t}(\tau\big|(\mathcal{Y},\mathcal{X})_{t-1}) = Q_{y_t}(\tau\big|\mathcal{Y}_{t-1}), \quad \forall \tau \in [0,1];$$

cf. Lee and Yang (2006), Hong et al. (2006).

Causal Relations between Return and Volume

Chuang, Kuan, and Lin (2009, JBF):

- 2 stock market indices: NYSE and S&P 500, from Jan 1990 to June 30, 2006 with 4135 and 4161 obs.
- Model:

$$r_t = a(\tau) + b(\tau)\,\frac{t}{T} + c(\tau)\left(\frac{t}{T}\right)^2 + \sum_{j=1}^{q} \alpha_j(\tau) r_{t-j} + \sum_{j=1}^{q} \beta_j(\tau) \ln v_{t-j} + e_t,$$

  where $\beta_j(\tau)$ represents the quantile causal effect of $\ln v_{t-j}$ on r_t, cf. Gallant, Rossi, and Tauchen (1992).
- Null hypothesis: $\beta_j(\tau) = 0$ for all τ in $(0, 1)$, i.e., Granger non-causality in quantiles. A sup-Wald test will do.

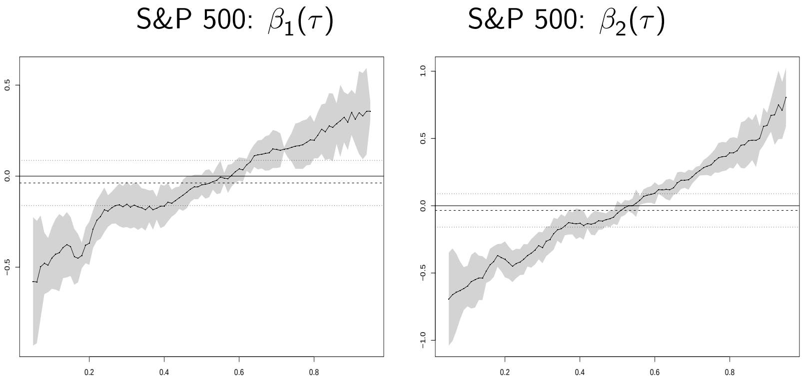
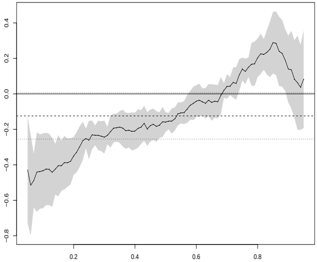
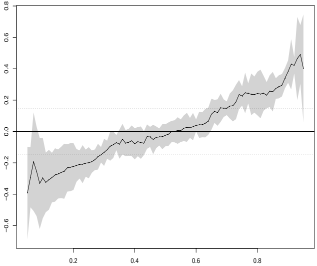


Figure: QR and LS estimates of the causal effects of log volume on return.

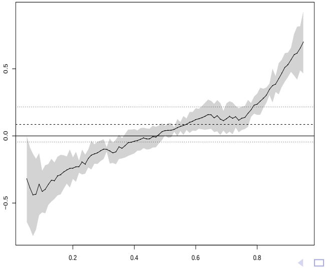
NYSE: $\beta_1(\tau)$



NYSE: $\beta_2(\tau)$



NYSE: $\beta_3(\tau)$



A Summary

- There is no causality in mean, but there are significant and heterogeneous causal effects of volume on return quantiles.
  - Causal effects have opposite signs at the two sides of dist.
  - Causal effects are stronger at more extreme quantiles.
  - The pairwise causal effects are symmetric about the median.
- With log volume (return) on the vertical (horizontal) axis, causal relations exhibit symmetric V shapes across quantiles, cf. Karpoff (1987), Gallant et al. (1992), and Blume et al. (1994). This implies that return dispersion (volatility) increases with volume.
- These results remain valid when the model includes r_{t-j}^2, and they are robust in different sample periods.

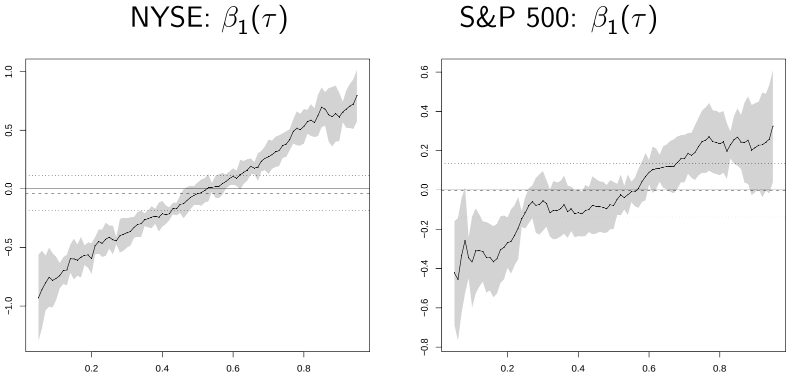


Figure: QR and LS estimates of the causal effects of log volume on return: 1995–2006.

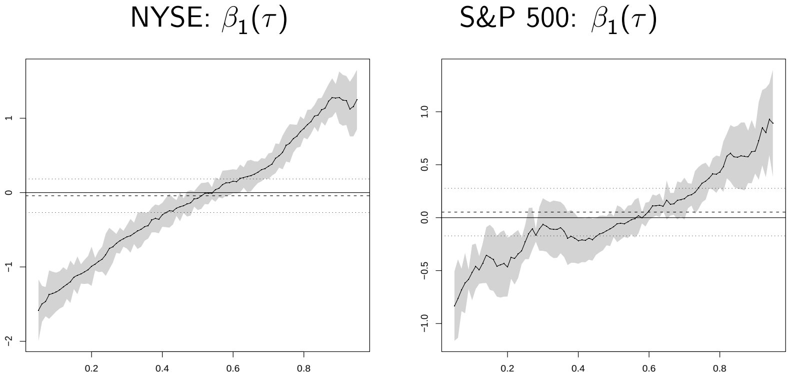


Figure: QR and LS estimates of the causal effects of log volume on return: 2000–2006.

Empirical Study: Effects of NHI on Saving

- There were 3 major health insurance programs in Taiwan: Labor Insurance (1950), Government Employees' Insurance (1958), and Farmers' Health Insurance (1985).
  - Government Employees' Insurance (GEI): Covers the employees (including retirees) in the public sector; the coverage was extended to spouses in 1982, to parents in 1989, and to children in 1992.
  - Labor Insurance: Covers only the employees in the private sector whose ages are 15–60, but not their spouse, parents, and children.
- The NHI launched in 1995 covers every citizen in Taiwan without discrimination, and its coverage is the same as that of GEI.
- We want to examine the effect of the enforcement of NHI on the precautionary saving in Taiwan.

Data from the SFIE

- The data are taken from the Survey of Family Income and Expenditure (SFIE). This nation-wide survey is not panel and uses new sampling every year.
- The sample size is about 16,000 each year in the early 90's and drops to about 14,000 from the mid-90's.
- The modules in the SFIE include individual socio-demographic and socio-economic characteristics along with income and outlays of each income earner within the household.
- The SFIE provides details on each household member's income sources and job attributes that are crucial for estimation of permanent income and identification of treatment/control groups.

- We employ 10 waves of SFIE from 1990 through 2000 (excluding 1995), with 1990–1994 and 1996–2000 as the periods before and after the enforcement of NHI.
- Our sample includes only the households with non-farm heads aged 20-69 years.
- There are 68,738 and 58,355 observations for the periods 1990–1994 and 1996–2000, respectively.
- While the average household real income increases moderately from NTD 0.97 million during 1990–1994 to more than NTD 1.2 million during 1996–2000, the corresponding average saving rates decrease by 50%, from 0.23 during 1990–1994 to 0.12 during 1996–2000.

Treatment and Control Groups

Groups are partitioned according to the working status of household head and his/her spouse.

1. Case 1:
   - Control: At least one of head and spouse in the public sector;
   - Treatment: Neither head nor spouse in the public sector.
2. Case 2:
   - Strict control: Head and spouse in the public sector;
   - Quasi control: One of head and spouse in the public sector;
   - Treatment 1: Head and spouse in the private sector;
   - Treatment 2: One of head and spouse in the private sector.

Unconditional Group Means and ATE

	1990–1994			1996–2000			
Sample	Obs.	Mean	S.D.	Obs.	Mean	S.D.	Difference
Full Sample	68,738	0.229	0.282	58,355	0.114	0.244	−0.115 (0.001)
Control	13,589	0.281	0.220	9,520	0.195	0.215	−0.085 (0.003)
Treatment	55,149	0.216	0.294	48,835	0.098	0.246	−0.118 (0.002)
Strict Control	2,476	0.341	0.195	1,717	0.263	0.200	−0.078 (0.006)
Quasi Control	11,113	0.267	0.223	7,803	0.180	0.215	−0.087 (0.003)
Treatment 1	24,545	0.227	0.228	22,279	0.107	0.220	−0.120 (0.002)
Treatment 2	30,600	0.208	0.338	26,556	0.091	0.265	−0.117 (0.003)

Comparison with Chou, Liu, & Hammitt (2003)

1. Data:
   - CLH03 include 1995 but we do not.
   - CLH03 exclude the sample with negative saving but we do not. Note that such sample was 11.8% in 1990–1994 but jumped to 27.2% in 1996–2000, leading to an average of 18.9% of the entire sample.
2. Dependent variable: CLH03 uses $\ln(Y - C)$, but we approximate the saving rate by $\ln Y - \ln C$.
3. Covariates: We partition some covariates (income and head age) into groups and study the groupwise ATE and QTEs.
4. Model: CLH03 make pairwise comparisons, but we estimate a model with multiple treatment groups.

Empirical Models

1. Control vs. one treatment group:

$$S_i = \alpha + X_i'\beta + D_i\gamma + (\text{NHI}_i)\delta + [(\text{NHI}_i) \times D_i]\zeta + e_i.$$

2. Control vs. multiple treatment groups:

$$S_i = \alpha + X_i'\beta + \sum_{j=0}^{2} D_i(j)\gamma_j + (\text{NHI}_i)\delta + \sum_{j=0}^{2}[(\text{NHI}_i) \times D_i(j)]\zeta_j + e_i.$$

where $D(j)$ are the indicators of the j-th treatment group ($j = 0$ for quasi-control group, $j = 1$ and 2 for treatment 1 and 2 groups).

1. To explore possible heterogeneity of the treatment effects, we partition the data into 5 permanent income groups (based on income quintiles) and 5 age groups (20–29,..., 60–69). We estimate the following model:

$$S_i = \alpha + X_i'\beta + \sum_{j=1}^{5} D_i K_i(j)\gamma_j + \sum_{j=1}^{5} (\mathrm{NHI}_i \times K_i(j))\delta_j + \sum_{j=1}^{5} [(\mathrm{NHI}_i) \times D_i \times K_i(j)]\zeta_j + e_i,$$

where D is the binary indicator of the treatment group, and $K(j)$, $j = 1, \ldots, 5$, are the indicators of 5 income groups or 5 age groups.

Empirical Result 1

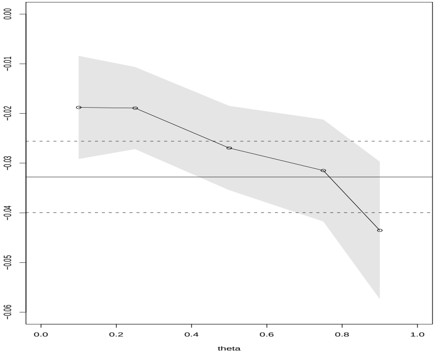


Figure: DID estimates of the ATE and QTE: 1 treatment group.

Empirical Result 2

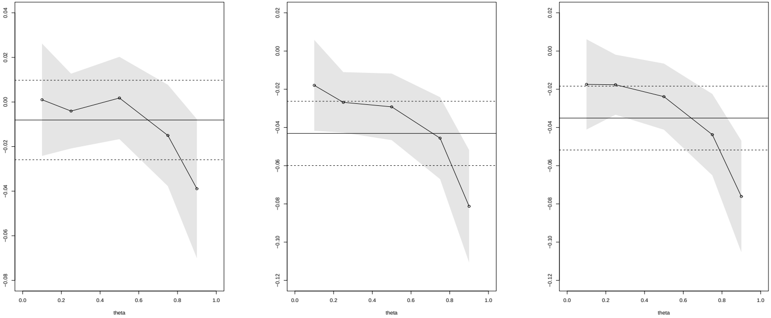


Figure: DID estimates of the ATE and QTE: Quasi control (left), treatment 1 (middle), and treatment 2 (right).

Empirical Result 3

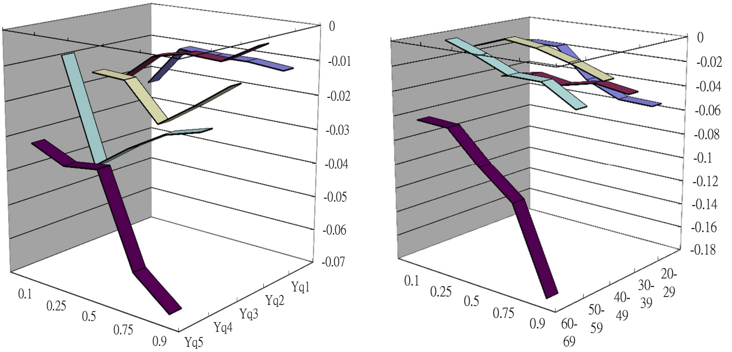


Figure: DID estimates of the ATE and QTE: 5 income groups (left) and 5 age groups (right).

A Summary

- The NHI has significantly negative (crowd-out) effect on precautionary saving, and such impact (QTE) is stronger for higher savers (magnitude increasing with saving quantile).
- The estimated conditional ATEs are quite close to the unconditional ATEs. For example, for the treatment 1 and treatment 2 groups, the estimated ATEs are -4.3% and -3.5% (vs. unconditional ATEs: -4.3% and -4.1%). For the quasi control group, the ATE is insignificant.
- The QTE patterns for treatment 1 and 2 groups are opposite to those of CLH03.
- The negative impact is the largest for the highest income group and the oldest (age 60–69) group, cf. CLH04.