Quantile Regression: An Introduction

Roger Koenker

University of Illinois, Urbana-Champaign

Aarhus: 21 June 2010

Danish Graduate Programme in Economics: Short Course

Overview of the Course

- The Basics: What, Why and How?
- Inference: Wald and Rank Tests
- Lab Session
- Computation: From the Inside and Outside
- Nonparametric QR
- QR Survival Analysis
- Lab Session
- Quantile Autoregression
- Risk Assessment and Choquet Portfolios
- QR for Longitudinal Data
- Endogoneity and IV Methods

The Basics: What, Why and How?

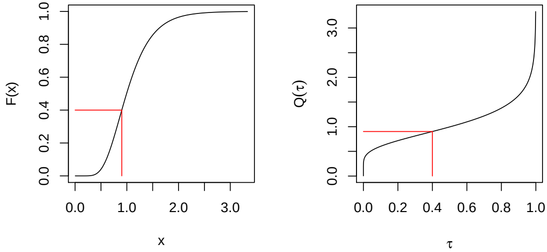
1. Univariate Quantiles
2. Scatterplot Smoothing
3. Equivariance Properties
4. Quantile Treatment Effects

Univariate Quantiles

Given a real-valued random variable, X, with distribution function F, we will define the τth quantile of X as

$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\{x \mid F(x) \geqslant \tau\}.$$

This definition follows the usual convention that F is CADLAG, and Q is CAGLAD as illustrated in the following pair of pictures.

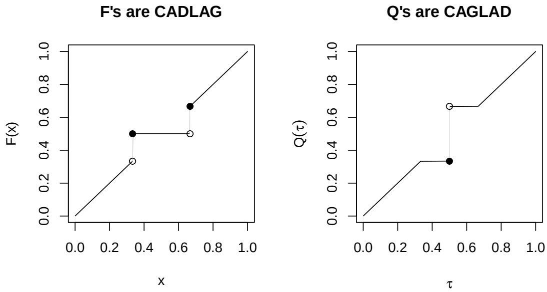


Univariate Quantiles

Given a real-valued random variable, X, with distribution function F, we will define the τth quantile of X as

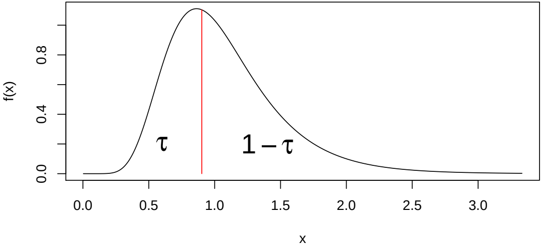
$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\{x \mid F(x) \geqslant \tau\}.$$

This definition follows the usual convention that F is CADLAG, and Q is CAGLAD as illustrated in the following pair of pictures.



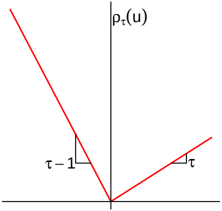
Univariate Quantiles

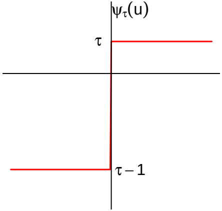
Viewed from the perspective of densities, the τth quantile splits the area under the density into two parts: one with area τ below the τth quantile and the other with area $1 - \tau$ above it:



Two Bits Worth of Convex Analysis

A convex function ρ and its subgradient ψ:





The subgradient of a convex function $f(u)$ at a point u consists of all the possible "tangents." Sums of convex functions are convex.

Population Quantiles as Optimizers

Quantiles solve a simple optimization problem:

$$\hat{\alpha}(\tau) = \text{argmin}\ \mathbb{E}\ \rho_\tau(Y - \alpha)$$

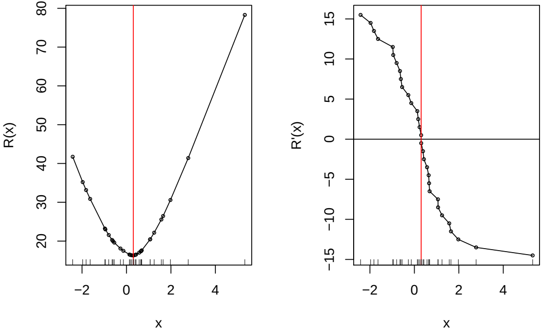
Proof: Let $\psi_\tau(u) = \rho'_\tau(u)$, so differentiating wrt to α:

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} \psi_\tau(y - \alpha) dF(y) \\ &= (\tau - 1) \int_{-\infty}^{\alpha} dF(y) + \tau \int_{\alpha}^{\infty} dF(y) \\ &= (\tau - 1) F(\alpha) + \tau(1 - F(\alpha)) \end{aligned}$$

implying $\tau = F(\alpha)$ and thus $\hat{\alpha} = F^{-1}(\tau)$.

Sample Quantiles as Optimizers

For sample quantiles replace F by $\hat{F}$, the empirical distribution function. The objective function becomes a polyhedral convex function whose derivative is monotone decreasing, in effect we are simply counting observations above and below and weighting the sums by τ and $1 - \tau$.



Conditional Quantiles: The Least Squares Meta-Model

The unconditional mean solves

$$\mu = \text{argmin}_m \mathbb{E}(Y - m)^2$$

Conditional Quantiles: The Least Squares Meta-Model

The unconditional mean solves

$$\mu = \text{argmin}_m \mathbb{E}(Y - m)^2$$

The conditional mean $\mu(x) = E(Y|X = x)$ solves

$$\mu(x) = \text{argmin}_m \mathbb{E}_{Y|X=x}(Y - m(X))^2.$$

Conditional Quantiles: The Least Squares Meta-Model

The unconditional mean solves

$$\mu = \text{argmin}_m \mathbb{E}(Y - m)^2$$

The conditional mean $\mu(x) = E(Y|X = x)$ solves

$$\mu(x) = \text{argmin}_m \mathbb{E}_{Y|X=x}(Y - m(X))^2.$$

Similarly, the unconditional τth quantile solves

$$\alpha_\tau = \text{argmin}_a \mathbb{E}\rho_\tau(Y - a)$$

Conditional Quantiles: The Least Squares Meta-Model

The unconditional mean solves

$$\mu = \text{argmin}_m \mathbb{E}(Y - m)^2$$

The conditional mean $\mu(x) = E(Y|X = x)$ solves

$$\mu(x) = \text{argmin}_m \mathbb{E}_{Y|X=x}(Y - m(X))^2.$$

Similarly, the unconditional τth quantile solves

$$\alpha_\tau = \text{argmin}_a \mathbb{E}\rho_\tau(Y - a)$$

and the conditional τth quantile solves

$$\alpha_\tau(x) = \text{argmin}_a \mathbb{E}_{Y|X=x}\rho_\tau(Y - a(X))$$

Computation of Linear Regression Quantiles

Primal Formulation as a linear program, split the residual vector into positive and negative parts and sum with appropriate weights:

$$\min\{\tau 1^\top u + (1-\tau)1^\top v | y = Xb + u - v, (b, u, v) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}\}$$

Dual Formulation as a Linear Program

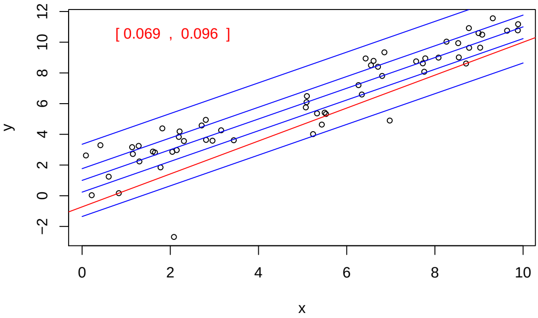
$$\max\{y'd | X^\top d = (1-\tau)X^\top 1, d \in [0,1]^n\}$$

Solutions are characterized by an exact fit to p observations.

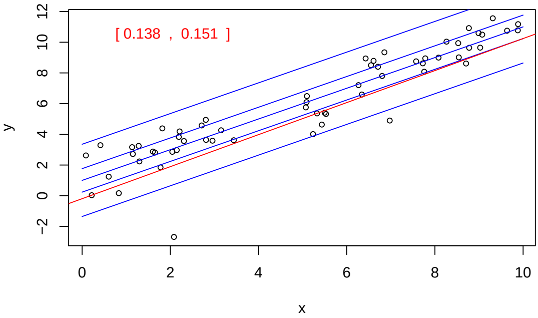
Quantile Regression: The Movie

- Bivariate linear model with iid Student t errors
- Conditional quantile functions are parallel in blue
- 100 observations indicated in blue
- Fitted quantile regression lines in red.
- Intervals for $\tau \in (0, 1)$ for which the solution is optimal.

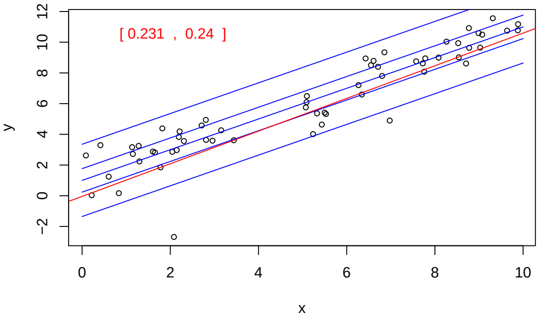
Quantile Regression in the iid Error Model



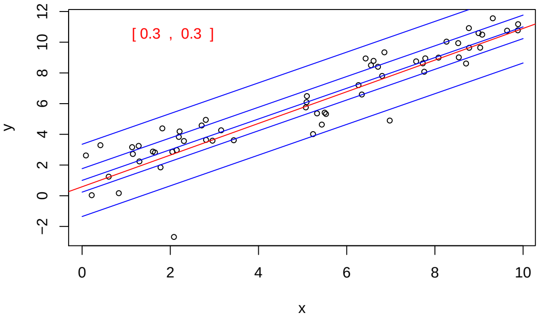
Quantile Regression in the iid Error Model



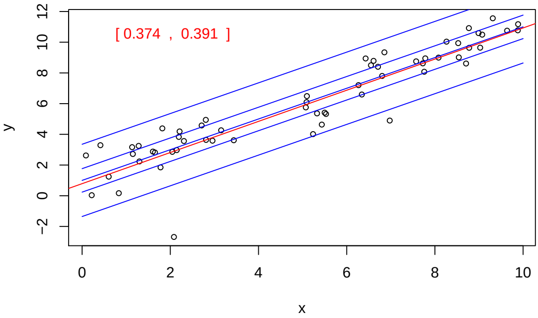
Quantile Regression in the iid Error Model



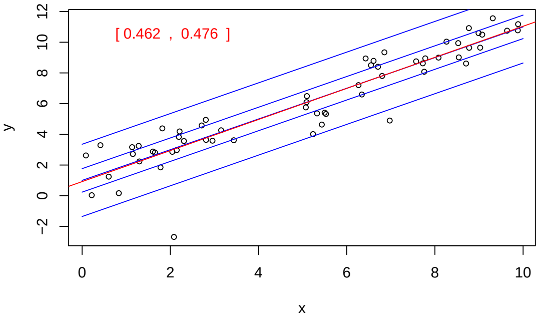
Quantile Regression in the iid Error Model



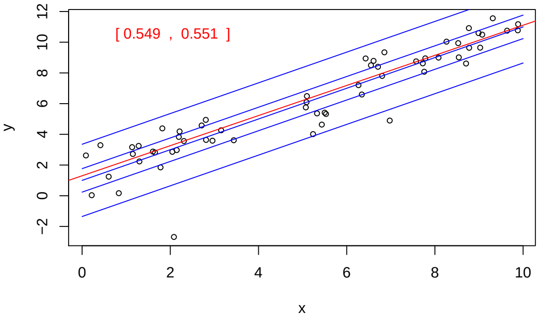
Quantile Regression in the iid Error Model



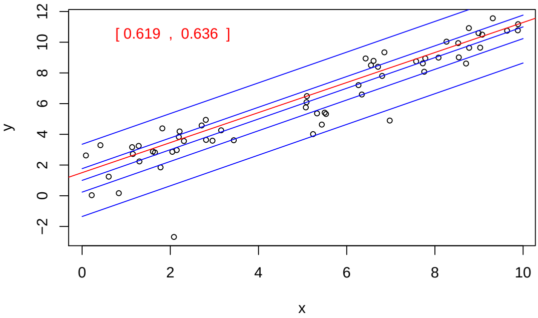
Quantile Regression in the iid Error Model



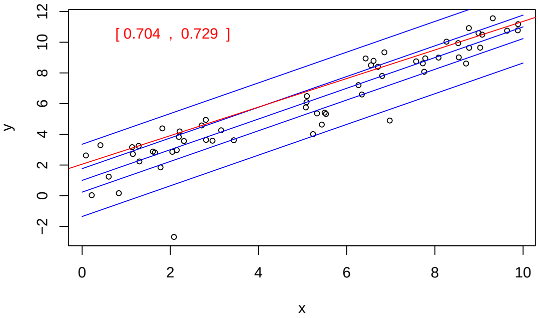
Quantile Regression in the iid Error Model



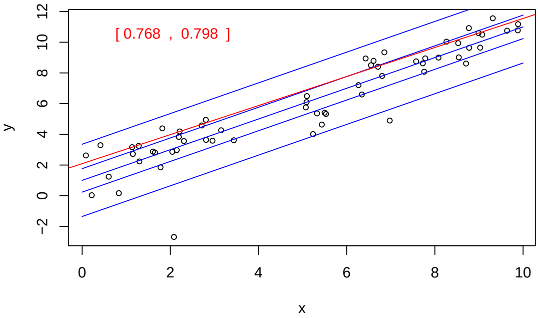
Quantile Regression in the iid Error Model



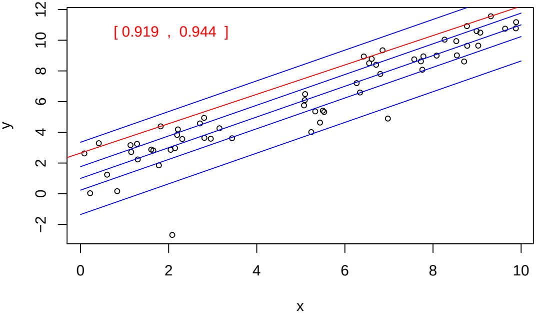
Quantile Regression in the iid Error Model



Quantile Regression in the iid Error Model



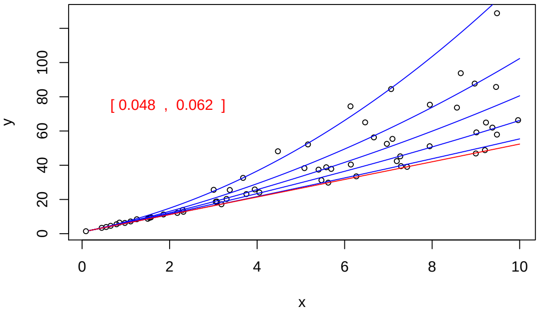
Quantile Regression in the iid Error Model



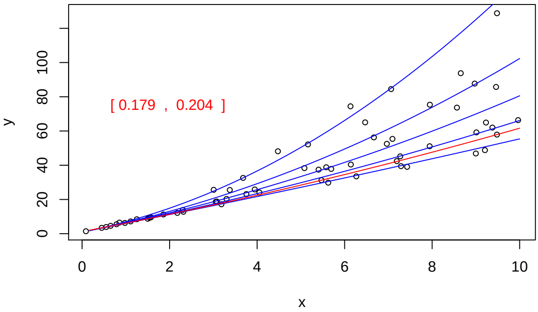
Virtual Quantile Regression II

- Bivariate quadratic model with Heteroscedastic χ^2 errors
- Conditional quantile functions drawn in blue
- 100 observations indicated in blue
- Fitted quadratic quantile regression lines in red
- Intervals of optimality for $\tau \in (0, 1)$.

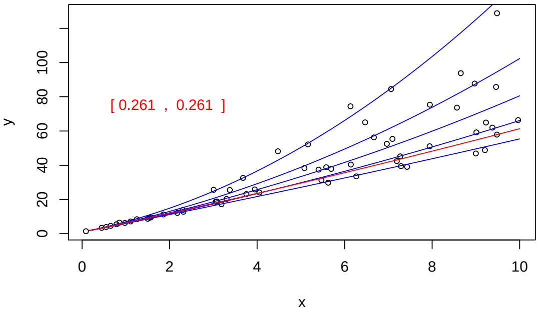
Quantile Regression in the Heteroscedastic Error Model



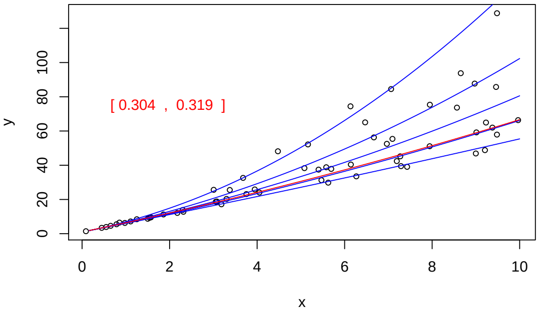
Quantile Regression in the Heteroscedastic Error Model



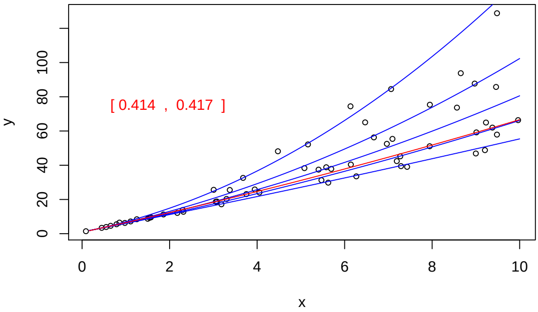
Quantile Regression in the Heteroscedastic Error Model



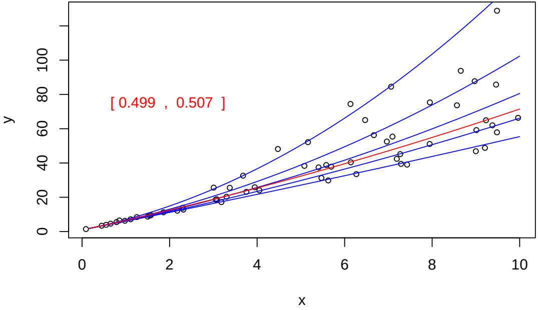
Quantile Regression in the Heteroscedastic Error Model



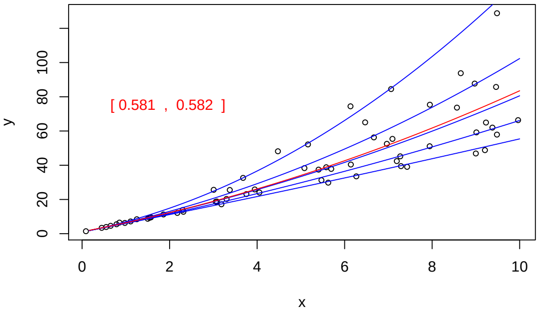
Quantile Regression in the Heteroscedastic Error Model



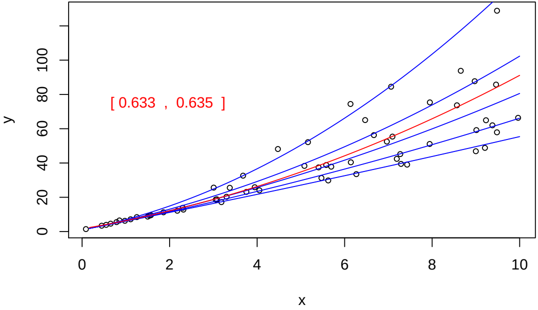
Quantile Regression in the Heteroscedastic Error Model



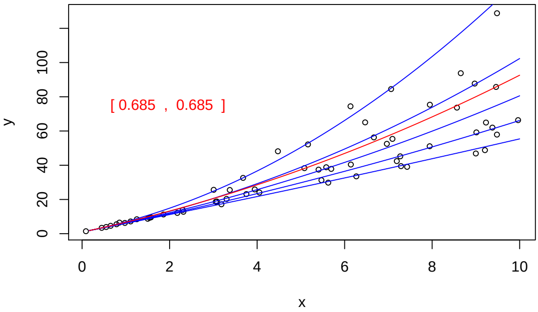
Quantile Regression in the Heteroscedastic Error Model



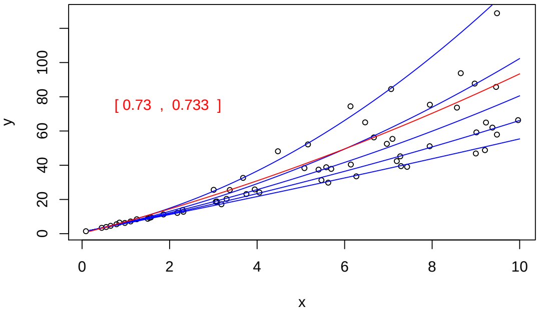
Quantile Regression in the Heteroscedastic Error Model



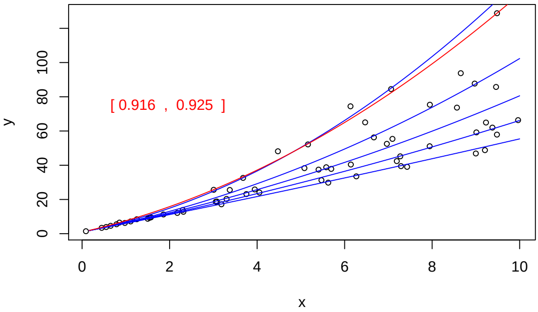
Quantile Regression in the Heteroscedastic Error Model



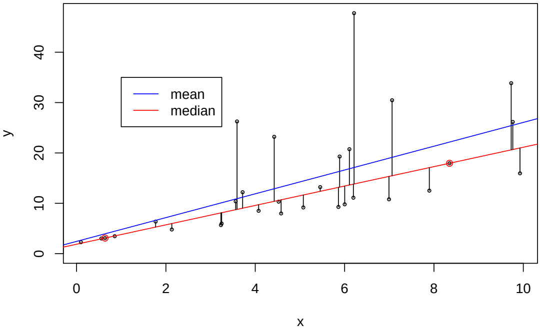
Quantile Regression in the Heteroscedastic Error Model



Quantile Regression in the Heteroscedastic Error Model



Conditional Means vs. Medians



Minimizing absolute errors for median regression can yield something quite different from the least squares fit for mean regression.

Equivariance of Regression Quantiles

- Scale Equivariance: For any $a > 0$, $\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; y, X)$ and $\hat{\beta}(\tau; -ay, X) = a\hat{\beta}(1-\tau; y, X)$

Equivariance of Regression Quantiles

- Scale Equivariance: For any $a > 0$, $\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; y, X)$ and $\hat{\beta}(\tau; -ay, X) = a\hat{\beta}(1-\tau; y, X)$
- Regression Shift: For any $\gamma \in \mathbb{R}^p$ $\hat{\beta}(\tau; y + X\gamma, X) = \hat{\beta}(\tau; y, X) + \gamma$

Equivariance of Regression Quantiles

- Scale Equivariance: For any $a > 0$, $\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; y, X)$ and $\hat{\beta}(\tau; -ay, X) = a\hat{\beta}(1-\tau; y, X)$
- Regression Shift: For any $\gamma \in \mathbb{R}^p$ $\hat{\beta}(\tau; y + X\gamma, X) = \hat{\beta}(\tau; y, X) + \gamma$
- Reparameterization of Design: For any $|A| \neq 0$, $\hat{\beta}(\tau; y, AX) = A^{-1}\hat{\beta}(\tau; yX)$

Equivariance of Regression Quantiles

- Scale Equivariance: For any $a > 0$, $\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; y, X)$ and $\hat{\beta}(\tau; -ay, X) = a\hat{\beta}(1-\tau; y, X)$
- Regression Shift: For any $\gamma \in \mathbb{R}^p$ $\hat{\beta}(\tau; y + X\gamma, X) = \hat{\beta}(\tau; y, X) + \gamma$
- Reparameterization of Design: For any $|A| \neq 0$, $\hat{\beta}(\tau; y, AX) = A^{-1}\hat{\beta}(\tau; yX)$
- Robustness: For any diagonal matrix D with nonnegative elements. $\hat{\beta}(\tau; y, X) = \hat{\beta}(\tau, y + D\hat{u}, X)$

Equivariance to Monotone Transformations

For any monotone function h, conditional quantile functions $Q_Y(\tau|x)$ are equivariant in the sense that

$$Q_{h(Y)|X}(\tau|x) = h(Q_{Y|X}(\tau|x))$$

In contrast to conditional mean functions for which, generally,

$$E(h(Y)|X) \neq h(EY|X)$$

Examples:
$h(y) = \min\{0, y\}$, Powell's (1985) censored regression estimator.
$h(y) = \mathrm{sgn}\{y\}$ Rosenblatt's (1957) perceptron, Manski's (1975) maximum score estimator. estimator.

Beyond Average Treatment Effects

Lehmann (1974) proposed the following general model of treatment response:

> *"Suppose the treatment adds the amount $\Delta(x)$ when the response of the untreated subject would be x. Then the distribution G of the treatment responses is that of the random variable $X + \Delta(X)$ where X is distributed according to F."*

Lehmann QTE as a QQ-Plot

Doksum (1974) defines $\Delta(x)$ as the "horizontal distance" between F and G at x, *i.e.*

$$F(x) = G(x + \Delta(x)).$$

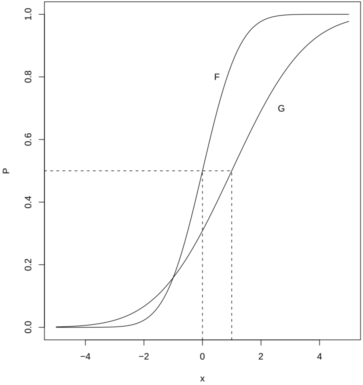
Then $\Delta(x)$ is uniquely defined as

$$\Delta(x) = G^{-1}(F(x)) - x.$$

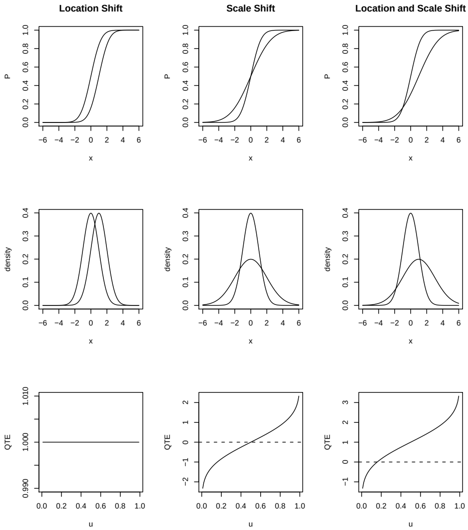
This is the essence of the conventional QQ-plot. Changing variables so $\tau = F(x)$ we have the quantile treatment effect (QTE):

$$\delta(\tau) = \Delta(F^{-1}(\tau)) = G^{-1}(\tau) - F^{-1}(\tau).$$

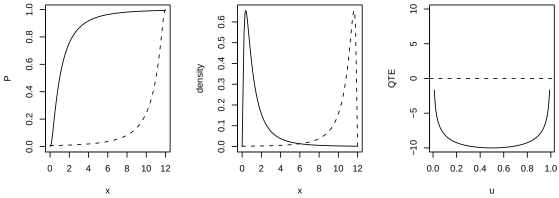
Lehmann-Doksum QTE



Lehmann-Doksum QTE



An Asymmetric Example



Treatment shifts the distribution from right skewed to left skewed making the QTE U-shaped.

The Erotic is Unidentified

The Lehmann QTE characterizes the difference in the marginal distributions, F and G, but it cannot reveal anything about the joint distribution, H. The copula function, Schweizer and Wolf (1981), Genest and McKay, (1986),

$$\varphi(u, v) = H(F^{-1}(u), G^{-1}(v)),$$

is *not* identified. Lehmann's formulation *assumes* that the treatment leaves the ranks of subjects invariant. If a subject was going to be the median control subject, then he will also be the median treatment subject. This is an inherent limitation of the Neymann-Rubin potential outcomes framework.

QTE via Quantile Regression

The Lehmann QTE is naturally estimable by

$$\hat{\delta}(\tau) = \hat{G}_n^{-1}(\tau) - \hat{F}_m^{-1}(\tau)$$

where $\hat{G}_n$ and $\hat{F}_m$ denote the empirical distribution functions of the treatment and control observations, Consider the quantile regression model

$$Q_{Y_i}(\tau | D_i) = \alpha(\tau) + \delta(\tau) D_i$$

where D_i denotes the treatment indicator, and $Y_i = h(T_i)$, *e.g.* $Y_i = \log T_i$, which can be estimated by solving,

$$\min \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \delta D_i)$$

Francis Galton's (1885) Anthropometric Quantiles

ANTHROPOMETRIC PER-CENTILES

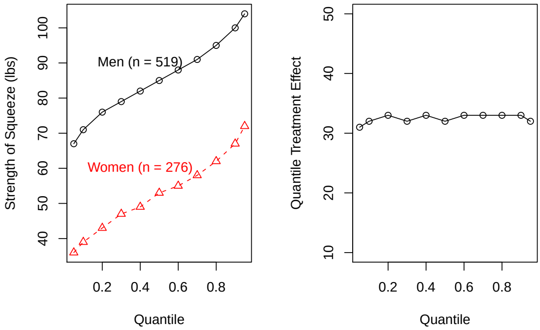
Values surpassed, and Values unreached, by various percentages of the persons measured at the Anthropometric Laboratory in the late International Health Exhibition

(*The value that is unreached by n per cent. of any large group of measurements, and surpassed by 100–n of them, is called its nth percentile*)

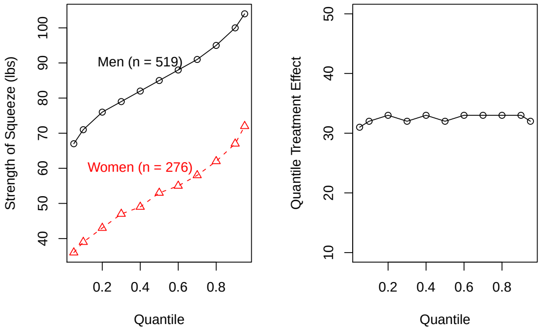
Subject of measurement	Age	Unit of measurement	Sex	No. of persons in the group	95	90	80	70	60	50	40	30	20	10	5
				Values unreached by per-cents. as below	5	10	20	30	40	50	60	70	80	90	95
Height, standing, without shoes ...	23–51	Inches	M.	811	63·2	64·5	65·8	66·5	67·3	67·9	68·5	69·2	70·0	71·3	72·4
			F.	770	58·8	59·9	61·3	62·1	62·7	63·3	63·9	64·6	65·3	66·4	67·3
Height, sitting, from seat of chair ...	23–51	Inches	M.	1013	33·6	34·2	34·9	35·3	35·4	36·0	36·3	36·7	37·1	37·7	38·2
			F.	775	31·8	32·3	32·9	33·3	33·6	33·9	34·2	34·6	34·9	35·6	36·0
Span of arms ...	23–51	Inches	M.	811	65·0	66·1	67·2	68·2	69·0	69·9	70·6	71·4	72·3	73·6	74·8
			F.	770	58·6	59·5	60·7	61·7	62·4	63·0	63·7	64·5	65·4	66·7	68·0
Weight in ordinary indoor clothes ...	23–26	Pounds	M.	520	121	125	131	135	139	143	147	150	156	165	172
			F.	276	102	105	110	114	118	122	129	132	136	142	149
Breathing capacity	23–26	Cubic inches	M.	212	161	177	187	199	211	219	226	236	248	277	290
			F.	277	92	102	115	124	131	138	144	151	164	177	186
Strength of pull as archer with bow	23–26	Pounds	M.	519	56	60	64	68	71	74	77	88	82	89	96
			F.	276	30	32	34	36	38	40	42	44	47	51	54
Strength of squeeze with strongest hand	23–26	Pounds	M.	519	67	71	76	79	82	85	88	91	95	100	104
			F.	276	36	39	43	47	49	52	55	58	62	67	72
Swiftness of blow.	23–26	Feet per second	M.	516	13·2	14·1	15·2	16·2	17·3	18·1	19·1	20·0	20·9	22·3	23·6
			F.	271	9·2	10·1	11·3	12·1	12·8	13·4	14·0	14·5	15·1	16·3	16·9
Sight, keenness of—by distance of reading diamond test-type	23–26	Inches	M.	398	13	17	20	22	23	25	26	28	30	32	34
			F.	433	10	12	16	19	22	24	26	27	29	31	32

(Values surpassed by per-cents as below: 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5)

Quantile Treatment Effects: Strength of Squeeze



Quantile Treatment Effects: Strength of Squeeze

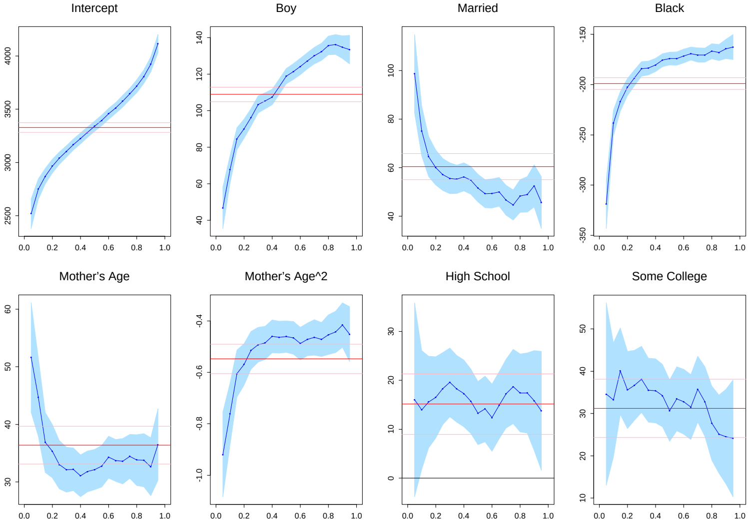


"Very powerful women exist, but happily perhaps for the repose of the other sex, such gifted women are rare."

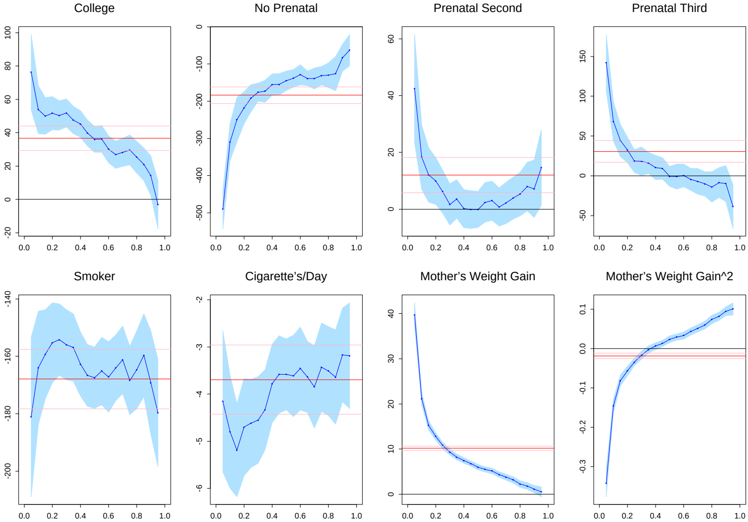
A Model of Infant Birthweight

- Reference: Abrevaya (2001), Koenker and Hallock (2001)
- Data: June, 1997, Detailed Natality Data of the US. Live, singleton births, with mothers recorded as either black or white, between 18-45, and residing in the U.S. Sample size: 198,377.
- Response: Infant Birthweight (in grams)
- Covariates:
  - Mother's Education
  - Mother's Prenatal Care
  - Mother's Smoking
  - Mother's Age
  - Mother's Weight Gain

Quantile Regression Birthweight Model I



Quantile Regression Birthweight Model II



Motivation

What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of of x*'s. 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.*

Motivation

What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of of x's. 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.

Motivation

What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of of x's. 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.

Mosteller and Tukey (1977)