Quantile Regression: A Gentle Introduction

Roger Koenker

University of Illinois, Urbana-Champaign

5th RMetrics Workshop, Meielisalp: 28 June 2011

Overview of the Course

- Some Basics: What, Why and How?
- Inference and Quantile Treatment Effects
- Nonparametric Quantile Regression
- Quantile Autoregression
- Risk Assessment and Choquet Portfolios

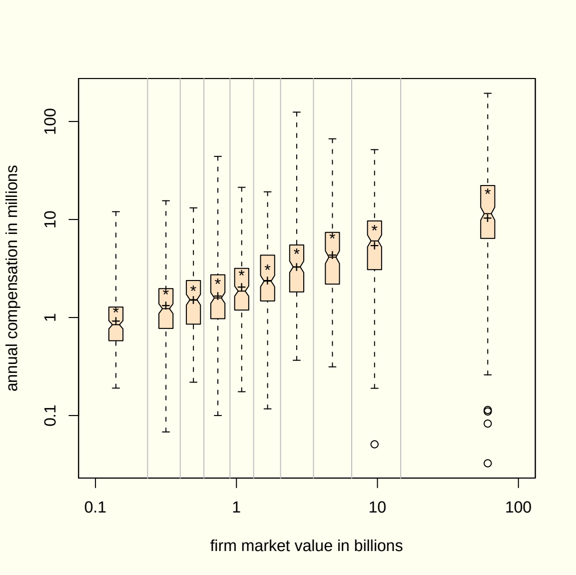
Course outline, lecture slides, an R FAQ, and even some proposed exercises can all be found at:

`http://www.econ.uiuc.edu/~roger/courses/RMetrics`.

A somewhat more extensive set of lecture slides can be found at:

`http://www.econ.uiuc.edu/~roger/courses/LSE`.

Boxplots of CEO Pay



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)

Motivation

Francis Galton in a famous passage defending the "charms of statistics" against its many detractors, chided his statistical colleagues

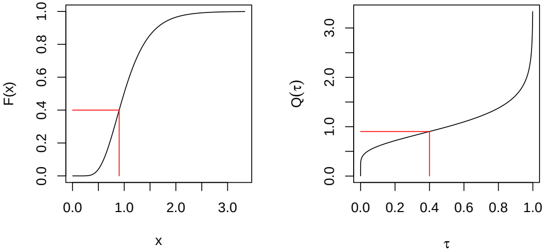
> *[who] limited their inquiries to Averages, and do not seem to revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of a native of one of our flat English counties, whose retrospect of Switzerland was that, if the mountains could be thrown into its lakes, two nuisances would be got rid of at once.* *Natural Inheritance, 1889*

Univariate Quantiles

Given a real-valued random variable, X, with distribution function F, we 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.

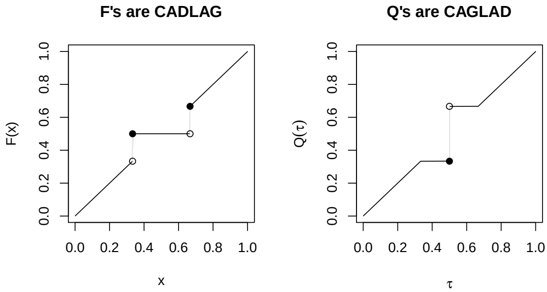


Univariate Quantiles

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

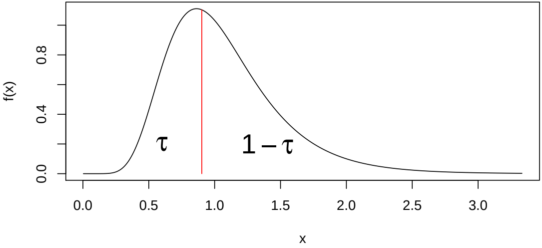
$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\{x \mid F(x) \geqslant \tau\}.$$

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



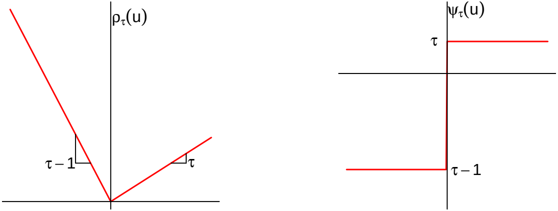
Univariate Quantiles

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:

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

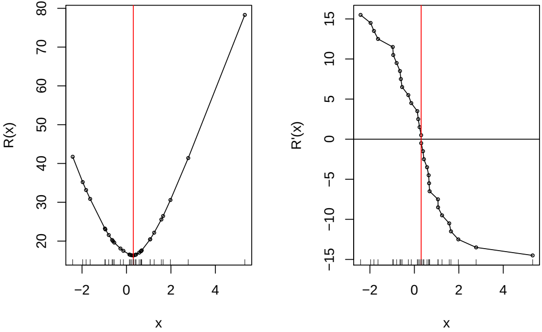
Proof: Let $\psi_\tau(u) = \rho'_\tau(u)$, so differentiating wrt to α:

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

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 the gradient simply counts observations above and below and weights 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.
Let $h \in \mathcal{H}$ index p-element subsets of $\{1, 2, ..., n\}$ then primal solutions take the form:

$$\hat{\beta} = \hat{\beta}(h) = X(h)^{-1}y(h)$$

Least Squares from the Quantile Regression Perspective

Exact fits to p observations:

$$\hat{\beta} = \hat{\beta}(h) = X(h)^{-1} y(h)$$

OLS is a weighted average of these $\hat{\beta}(h)$'s:

$$\hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y = \sum_{h \in \mathcal{H}} w(h) \hat{\beta}(h),$$

$$w(h) = |X(h)|^2 / \sum_{h \in \mathcal{H}} |X(h)|^2$$

Least Squares from the Quantile Regression Perspective

Exact fits to p observations:

$$\hat{\beta} = \hat{\beta}(h) = X(h)^{-1} y(h)$$

OLS is a weighted average of these $\hat{\beta}(h)$'s:

$$\hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y = \sum_{h \in \mathcal{H}} w(h) \hat{\beta}(h),$$

$$w(h) = |X(h)|^2 / \sum_{h \in \mathcal{H}} |X(h)|^2$$

The determinants $|X(h)|$ are the (signed) volumes of the parallelipipeds formed by the columns of the the matrices $X(h)$. In the simplest bivariate case, we have,

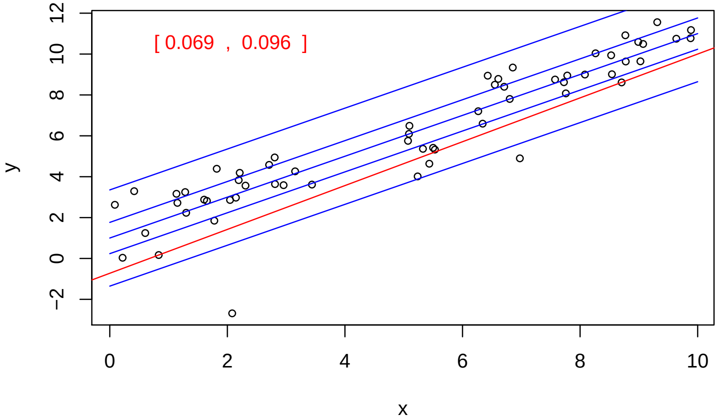
$$|X(h)|^2 = \begin{vmatrix} 1 & x_i \\ 1 & x_j \end{vmatrix}^2 = (x_j - x_i)^2$$

so pairs of observations that are far apart are given more weight.

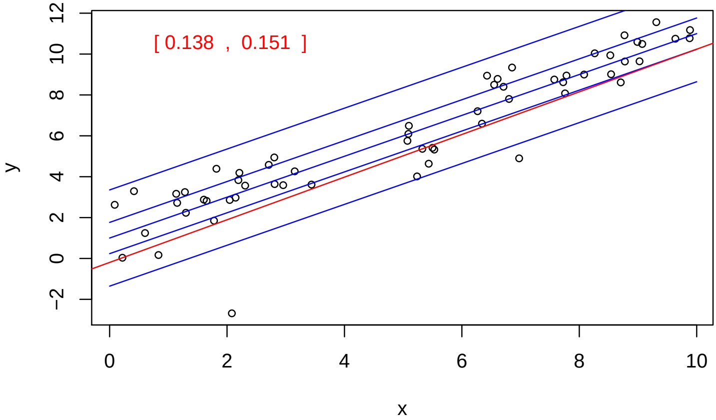
Quantile Regression: The Movie

- 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.

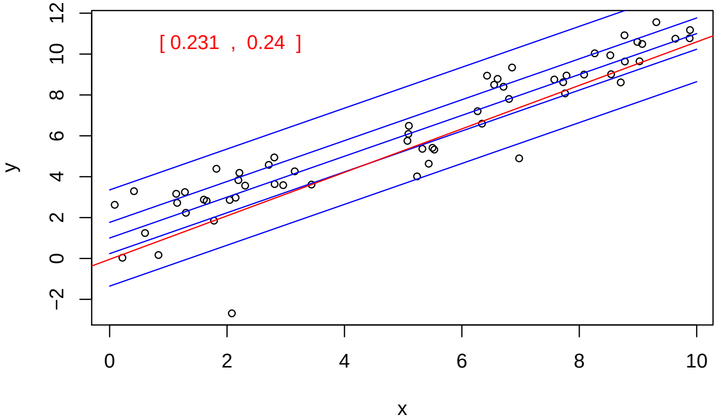
Quantile Regression in the iid Error Model



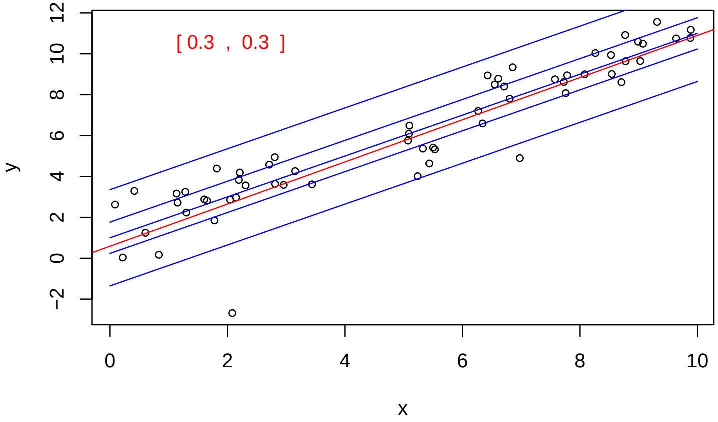
Quantile Regression in the iid Error Model



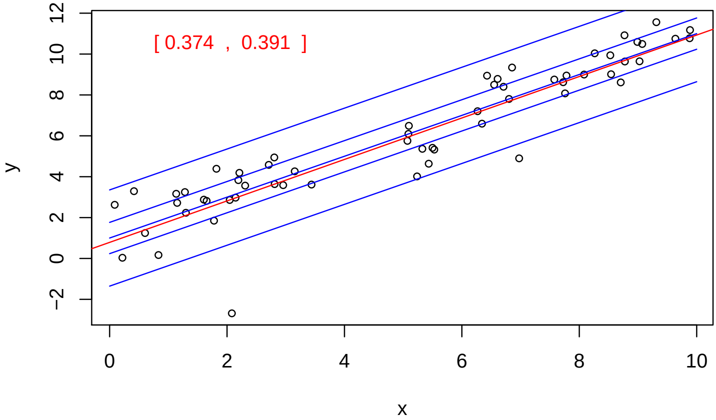
Quantile Regression in the iid Error Model



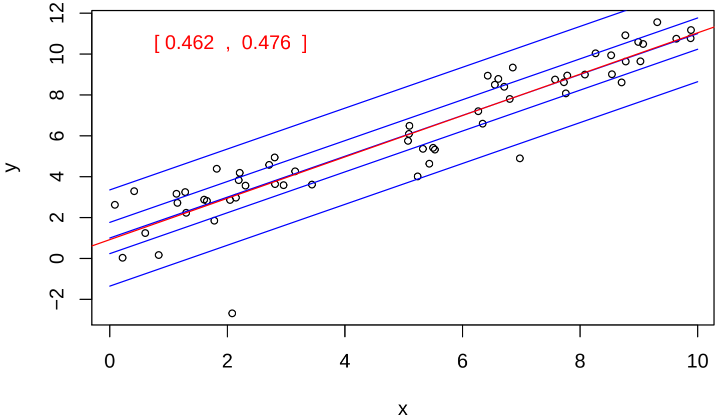
Quantile Regression in the iid Error Model



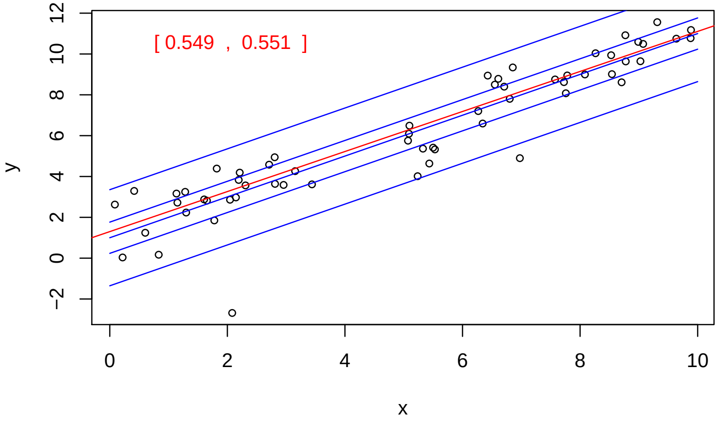
Quantile Regression in the iid Error Model



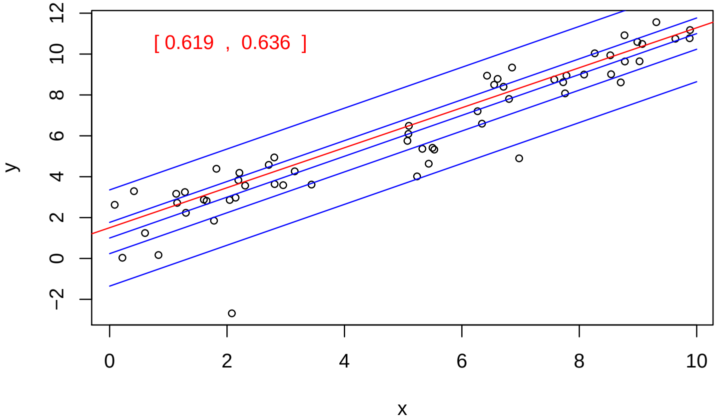
Quantile Regression in the iid Error Model



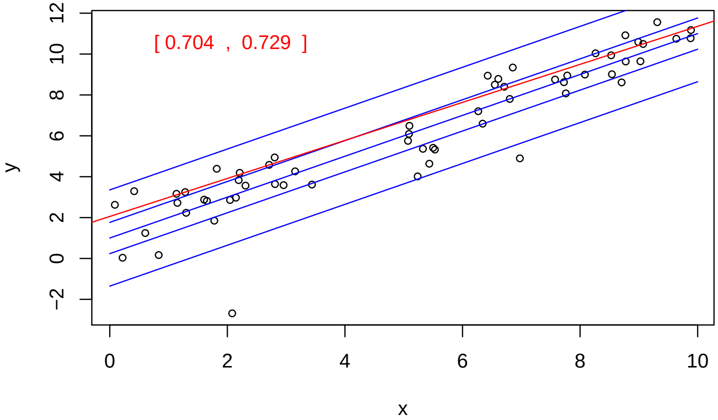
Quantile Regression in the iid Error Model



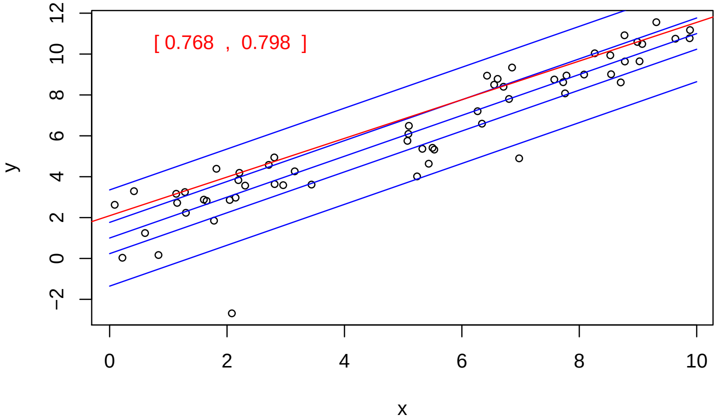
Quantile Regression in the iid Error Model



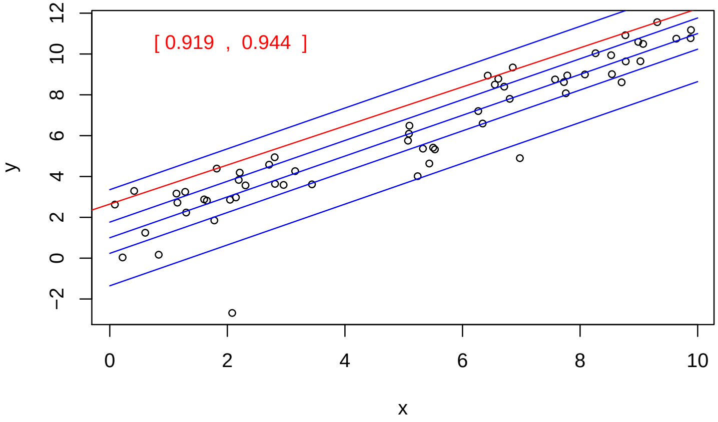
Quantile Regression in the iid Error Model



Quantile Regression in the iid Error Model



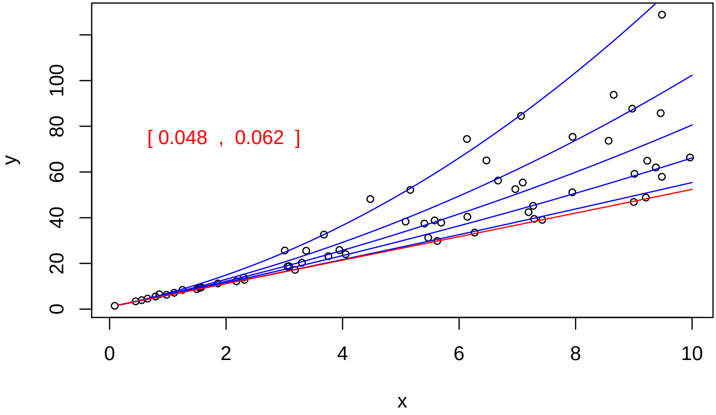
Quantile Regression in the iid Error Model



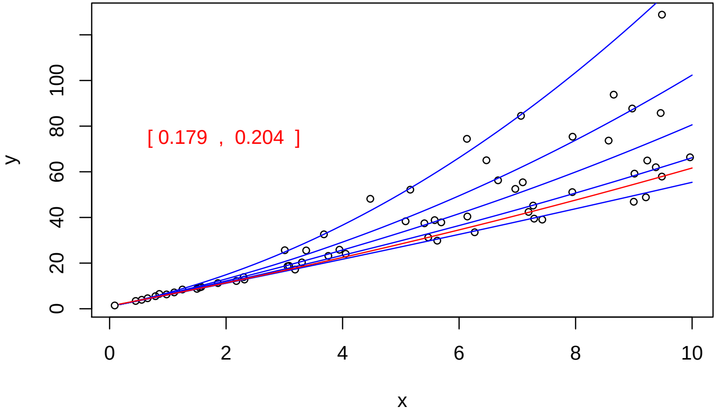
Virtual Quantile Regression II

- 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 $\tau \in (0, 1)$.

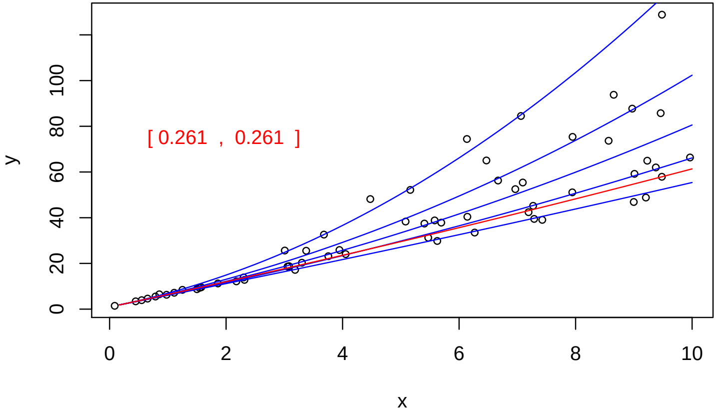
Quantile Regression in the Heteroscedastic Error Model



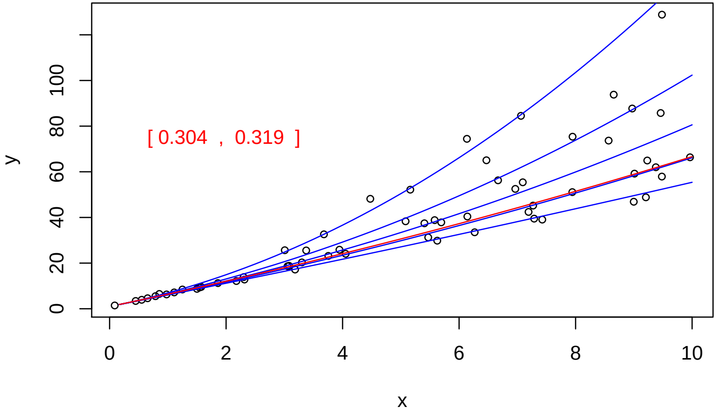
Quantile Regression in the Heteroscedastic Error Model



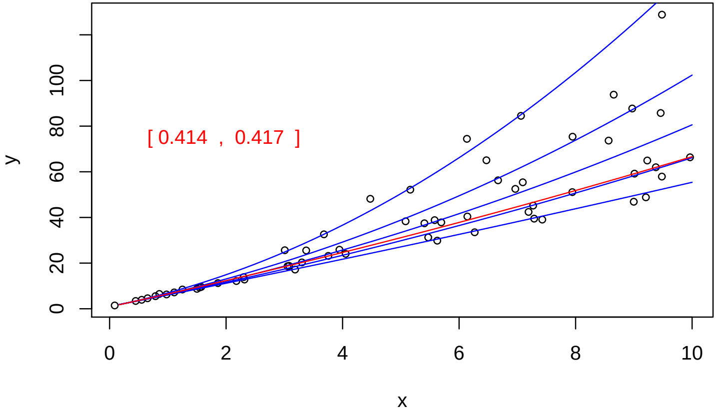
Quantile Regression in the Heteroscedastic Error Model



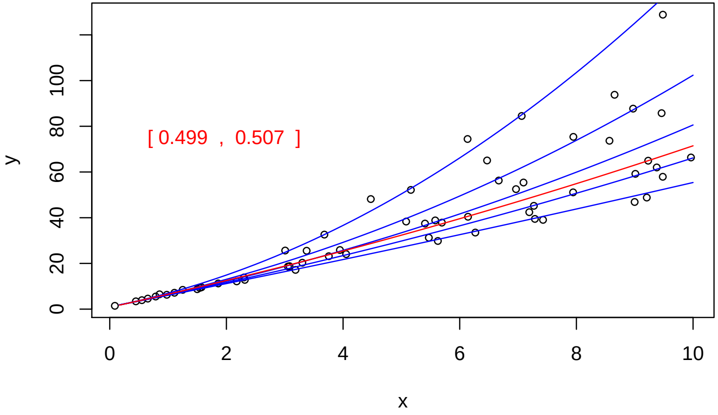
Quantile Regression in the Heteroscedastic Error Model



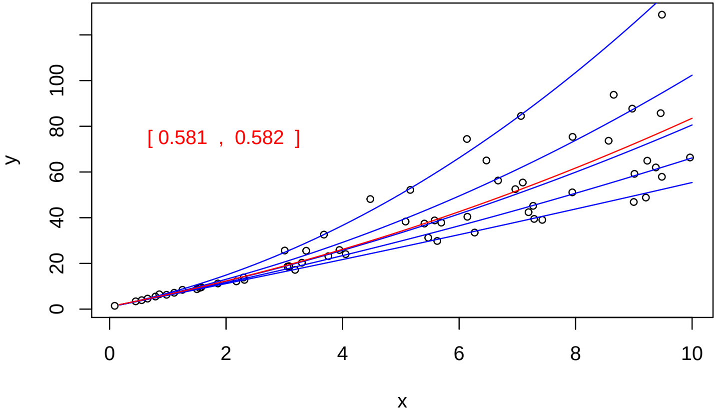
Quantile Regression in the Heteroscedastic Error Model



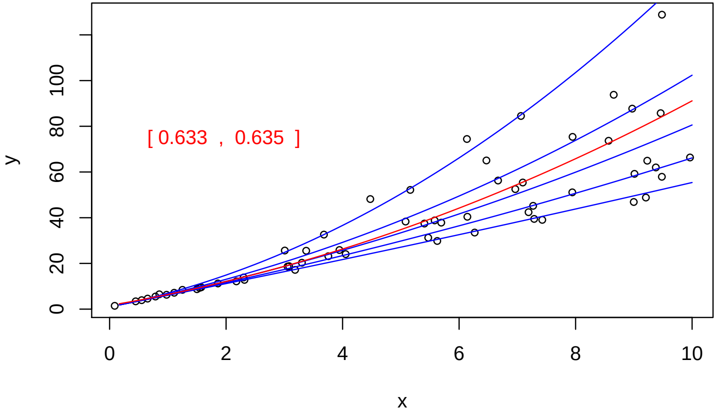
Quantile Regression in the Heteroscedastic Error Model



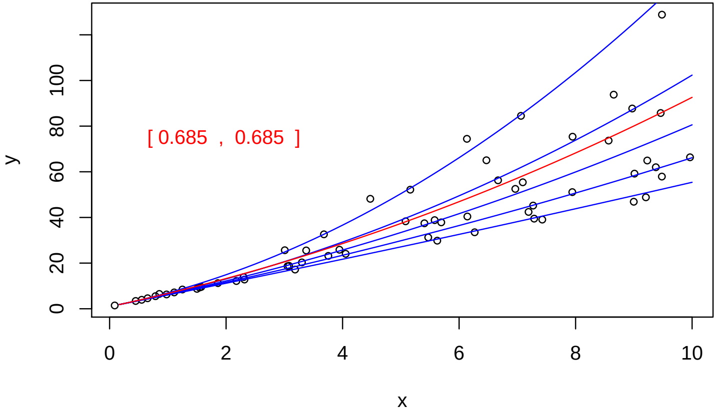
Quantile Regression in the Heteroscedastic Error Model



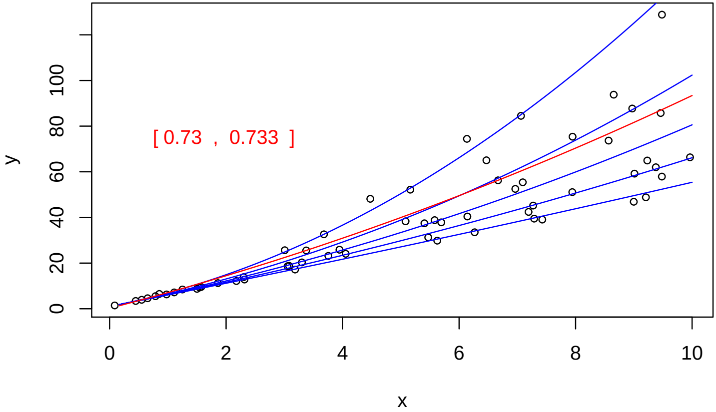
Quantile Regression in the Heteroscedastic Error Model



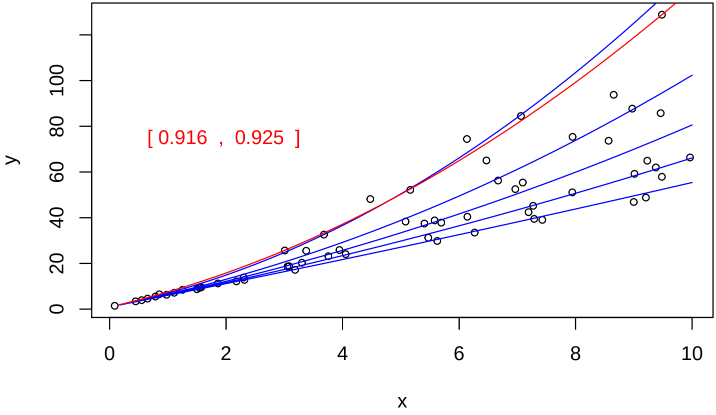
Quantile Regression in the Heteroscedastic Error Model



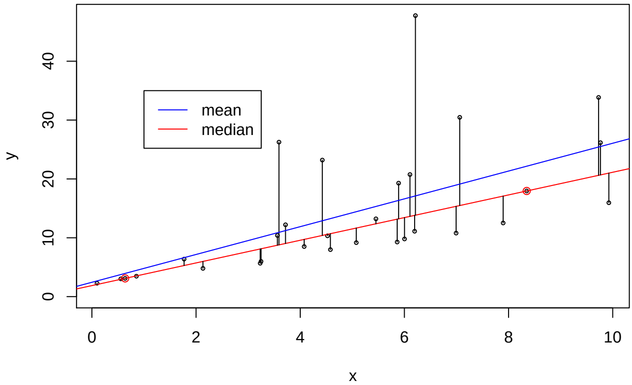
Quantile Regression in the Heteroscedastic Error Model



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.
$h(y) = \text{sgn}\{y\}$ 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.*

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

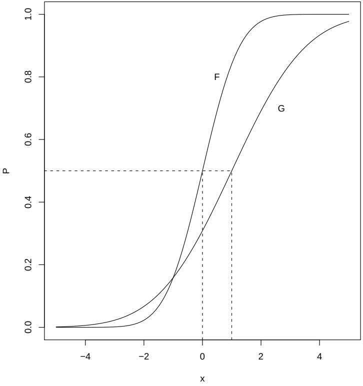
Then $\Delta(x)$ is uniquely defined as

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

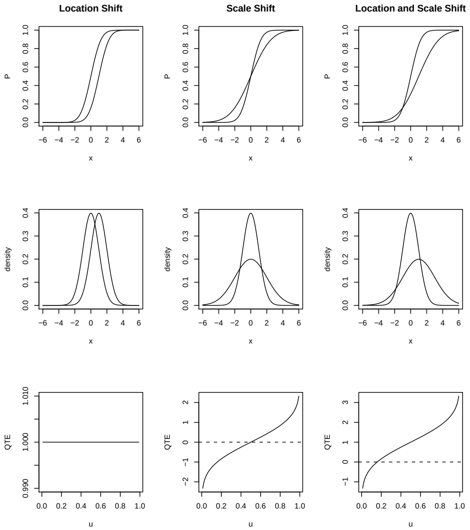
This is the essence of the conventional QQ-plot. Changing variables so $\tau = F(x)$ we have the quantile treatment effect (QTE):

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

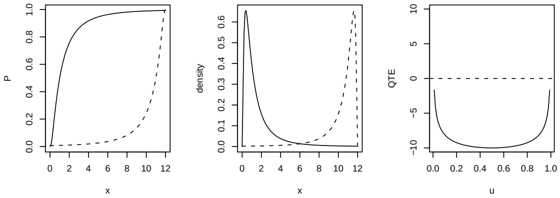
Lehmann-Doksum QTE



Lehmann-Doksum QTE



An Asymmetric Example



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

QTE via Quantile Regression

The Lehmann QTE is naturally estimable by

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

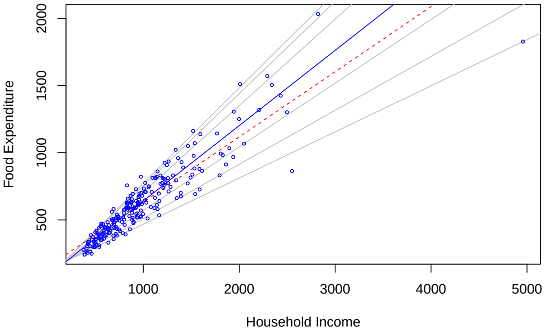
where $\hat{G}_n$ and $\hat{F}_m$ denote the empirical distribution functions of the treatment and control observations, Consider the quantile regression model

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

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,

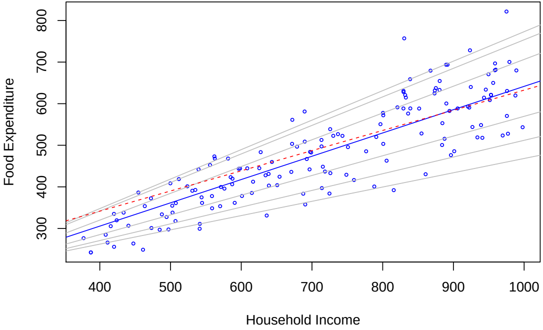
$$\min \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \delta D_i)$$

Engel's Food Expenditure Data



Engel Curves for Food: This figure plots data taken from Engel's (1857) study of the dependence of households' food expenditure on household income. Seven estimated quantile regression lines for $\tau \in \{.05, .1, .25, .5, .75, .9, .95\}$ are superimposed on the scatterplot. The median $\tau = .5$ fit is indicated by the blue solid line; the least squares estimate of the conditional mean function is indicated by the red dashed line.

Engel's Food Expenditure Data

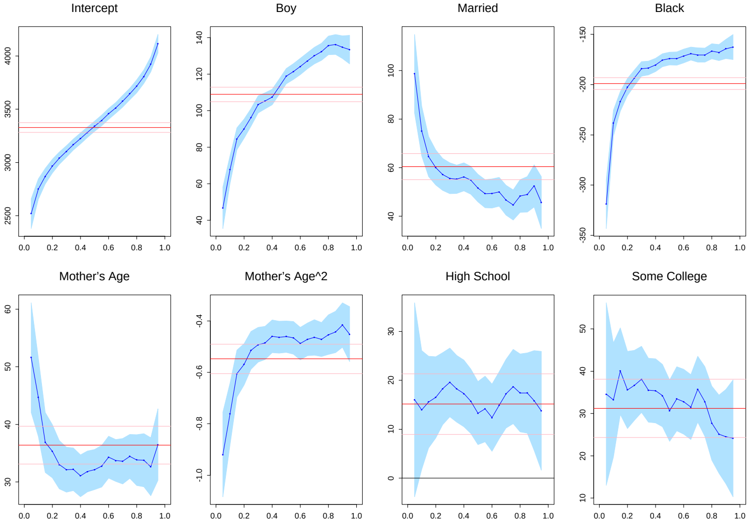


Engel Curves for Food: This figure plots data taken from Engel's (1857) study of the dependence of households' food expenditure on household income. Seven estimated quantile regression lines for $\tau \in \{.05, .1, .25, .5, .75, .9, .95\}$ are superimposed on the scatterplot. The median $\tau = .5$ fit is indicated by the blue solid line; the least squares estimate of the conditional mean function is indicated by the red dashed line.

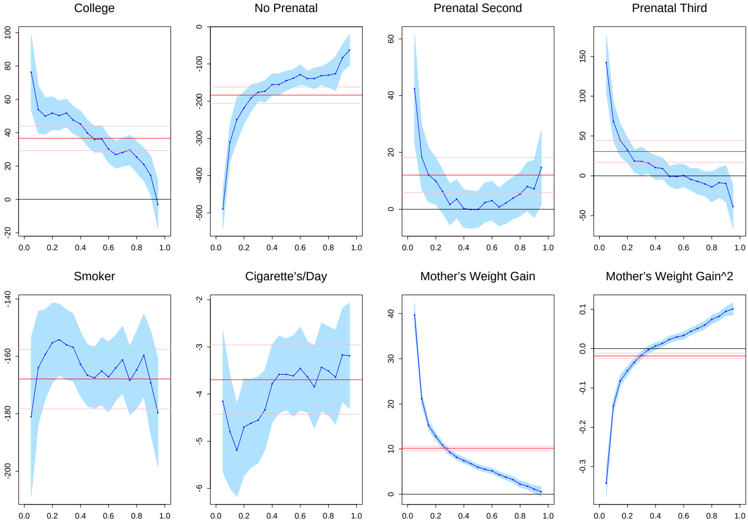
A Model of Infant Birthweight

- 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

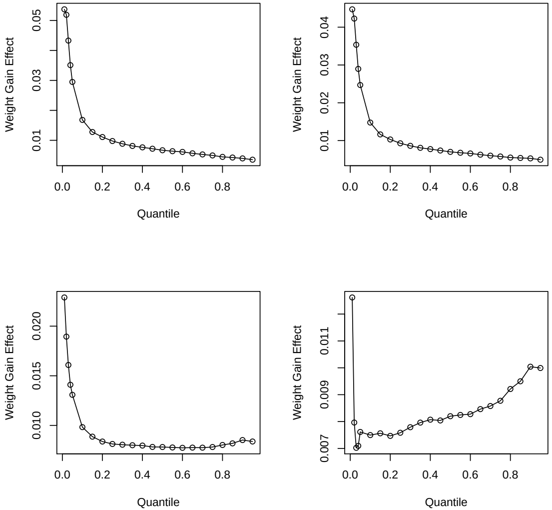
Quantile Regression Birthweight Model I



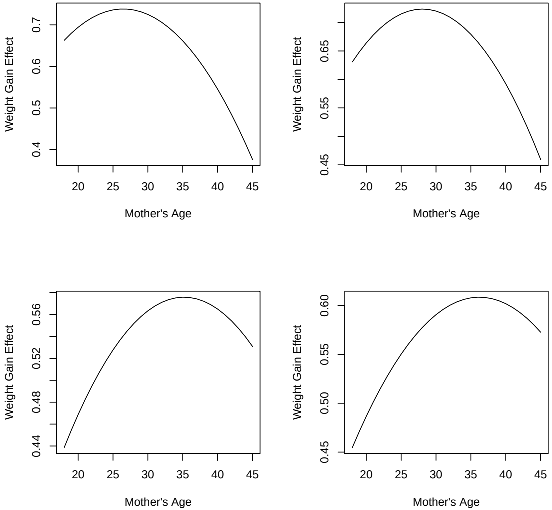
Quantile Regression Birthweight Model II



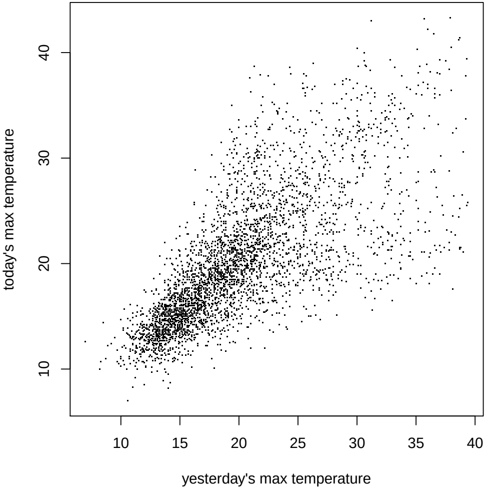
Marginal Effect of Mother's Age



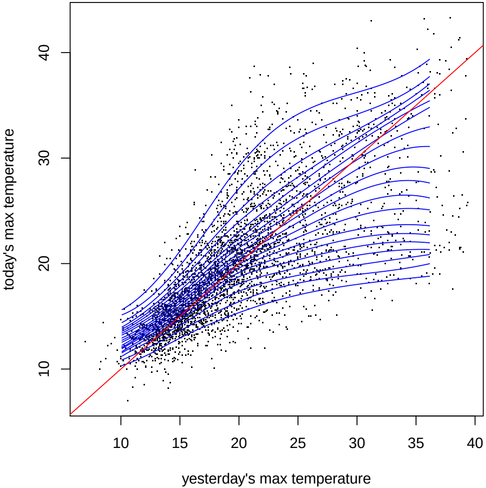
Marginal Effect of Mother's Weight Gain



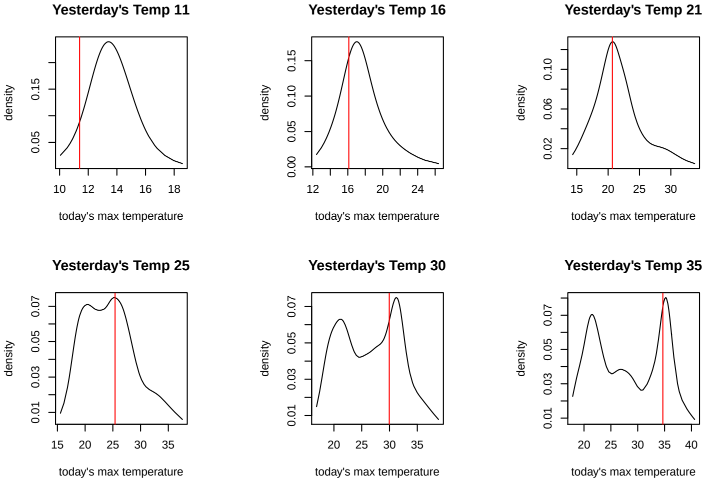
Daily Temperature in Melbourne: AR(1) Scatterplot



Daily Temperature in Melbourne: Nonlinear QAR(1) Fit



Conditional Densities of Melbourne Daily Temperature



Review

Least squares methods of estimating conditional mean functions

- were developed for, and
- promote the view that,

$$\text{Response} = \text{Signal} + \text{iid (Gaussian Measurement Error}$$

In fact the world is rarely this simple. Quantile regression is intended to expand the regression window allowing us to see a wider vista.