Quantile Regression:
A Gentle Introduction for R-ophiles

Roger Koenker

University of Illinois, Urbana-Champaign

Social Science Research Using R: 19 June 2009

What is Quantile Regression?

- Quantiles Efficiently Describe Marginal Distributions
  - Proportion τ of students perform better than the τth quantile.

What is Quantile Regression?

- Quantiles Efficiently Describe Marginal Distributions
  - Proportion τ of students perform better than the τth quantile.
- Regression Quantiles Describe Conditional Distributions
  - Given characteristics X, proportion τ of students of type X perform better than τth conditional quantile.

What is Quantile Regression?

- Quantiles Efficiently Describe Marginal Distributions
  - Proportion τ of students perform better than the τth quantile.
- Regression Quantiles Describe Conditional Distributions
  - Given characteristics X, proportion τ of students of type X perform better than τth conditional quantile.
- Quantiles minimize asymmetric linear loss
  - Sorting can be replaced by convex optimization.

What is Quantile Regression?

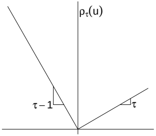
- Quantiles Efficiently Describe Marginal Distributions
  - Proportion τ of students perform better than the τth quantile.
- Regression Quantiles Describe Conditional Distributions
  - Given characteristics X, proportion τ of students of type X perform better than τth conditional quantile.
- Quantiles minimize asymmetric linear loss
  - Sorting can be replaced by convex optimization.
- Regression Quantiles also minimize asymmetric linear loss
  - Optimization generalizes nicely to the regression setting, unlike sorting.

Sample Quantiles via Optimization

The τth sample quantile can be defined as any solution to:

$$\hat{\alpha}(\tau) = \text{argmin}_{a \in \mathfrak{R}} \sum_{i=1}^{n} \rho_\tau(y_i - a)$$

where $\rho_\tau(u) = (\tau - I(u < 0))u$ as illustrated below.



Biases the argmin toward making the lower cost error; e.g. forecasting flood levels.

The Least Squares Meta-Model

The unconditional mean solves

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

The Least Squares Meta-Model

The unconditional mean solves

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

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

$$\mu(x) = \operatorname{argmin}_m E_{Y|X=x}(Y - m(x))^2.$$

The Least Squares Meta-Model

The unconditional mean solves

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

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

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

Similarly, the unconditional τth quantile solves

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

The Least Squares Meta-Model

The unconditional mean solves

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

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

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

Similarly, the unconditional τth quantile solves

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

and the conditional τth quantile solves

$$\alpha_\tau(x) = \text{argmin}_q E_{Y|X=x}\rho_\tau(Y - q(x))$$

Parametric Quantile Regression

Linear parametric models are simplest:

$$Q_Y(\tau|x) = q(x) = x^\top \beta(\tau)$$

estimable by solving the linear program:

$$\hat{\beta}(\tau) = \text{argmin}_b \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top b)$$

Solutions have p zero residuals when $\beta \in \mathbb{R}^p$.

Parametric Quantile Regression

Linear parametric models are simplest:

$$Q_Y(\tau|x) = q(x) = x^\top \beta(\tau)$$

estimable by solving the linear program:

$$\hat{\beta}(\tau) = \text{argmin}_b \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top b)$$

Solutions have p zero residuals when $\beta \in \mathbb{R}^p$.
Nonlinear (in parameters) models can also be estimated:

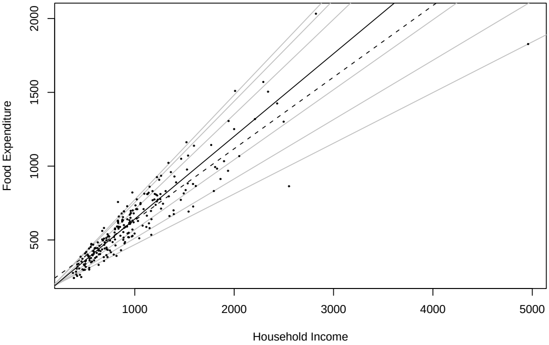
$$\hat{\beta}(\tau) = \text{argmin}_b \sum_{i=1}^{n} \rho_\tau(y_i - g(x_i, b))$$

for some fully specified function, g.

Four Introductory Applications

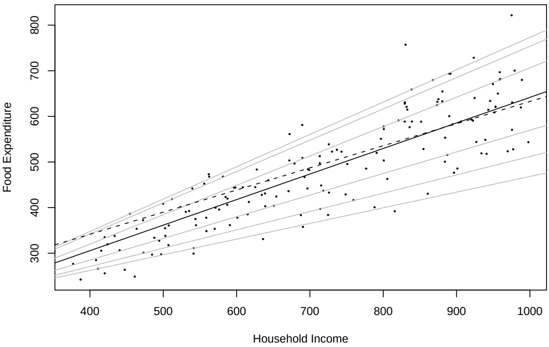
- Engel's Law: A Classical Economic Example
- CEO Pay: Boxplots as nonparametric Quantile Regression
- Infant Birthweight: A Public Health Example
- Melbourne Daily Temperature: A Time Series Example

Engel's Food Expenditure Data



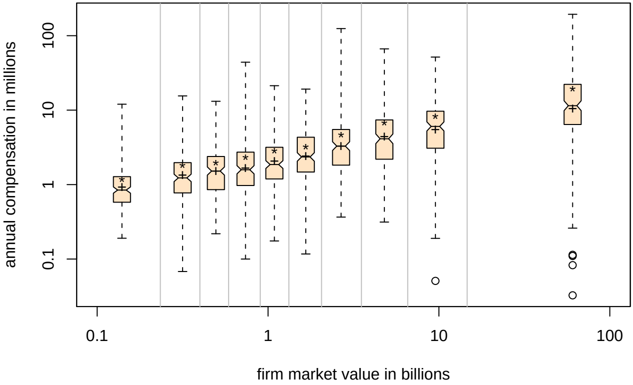
Engel Curves for Food: This figure plots data taken from Engel's (1857) study of the dependence of households' food expenditure on household income. Seven estimated quantile regression lines for $\tau \in \{.05, .1, .25, .5, .75, .9, .95\}$ are superimposed on the scatterplot. The median $\tau = .5$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

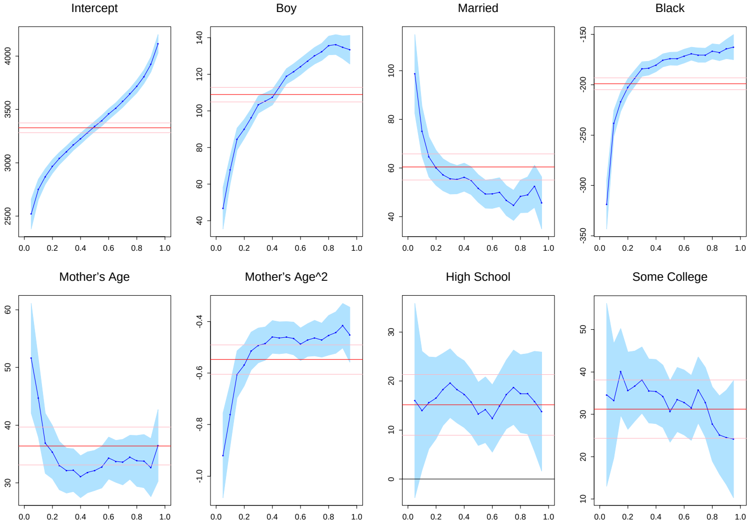
Boxplot of CEO Pay by Firm Size



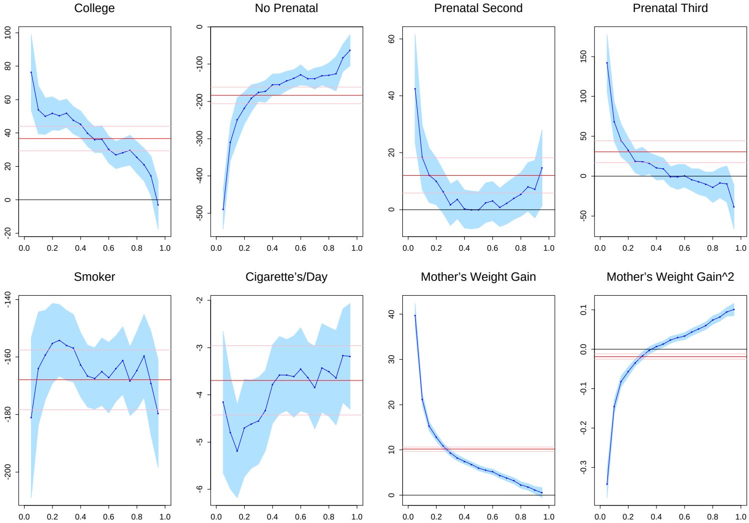
A Model of Infant Birthweight

- 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

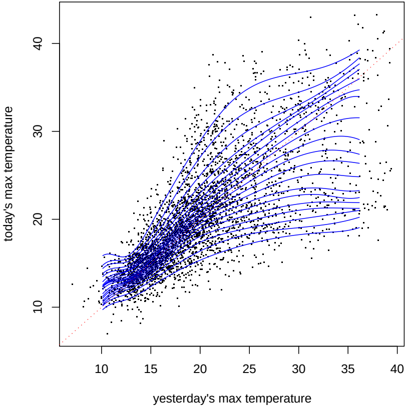
Quantile Regression Birthweight Model I



Quantile Regression Birthweight Model II

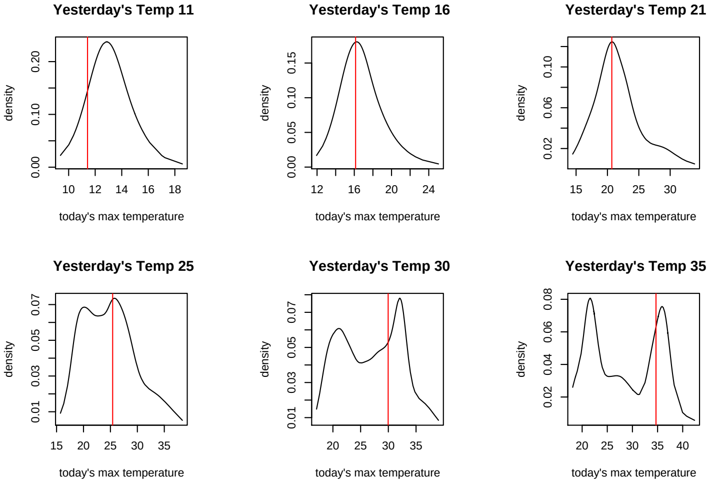


AR(1) Model of Melbourne Daily Temperature



The plot illustrates 10 years of daily maximum temperature data for Melbourne, Australia as an AR(1) scatterplot. Superimposed are estimated conditional quantile functions for $\tau \in \{.05, .10, ..., .95\}$. parameterized via B-splines.

Predictive Densities for Melbourne Temperature



Quantile Autoregression and Irrational Exuberance

- Simple linear QAR models

  $$Q_{Y_t|Y_{t-1}}(\tau|y_{t-1}) = \alpha(\tau) + \beta(\tau)y_{t-1}$$

  can exhibit strong unit-root or even explosive episodic tendencies, but still be stationary, and mean reverting, provided that $\beta(\tau)$ is square integrable, K and Xiao (2006).
- Copulas offer a rich source of convenient *nonlinear* specifications of QAR models, Chen, K and Xiao (2009).
- Similar methods yield more flexible GARCH type models, K and Xiao (2009).

Nonparametric Quantile Regression

Locally Polynomial (Kernel) Method

$$\begin{aligned}\hat{\alpha}(\tau, x) &= \text{argmin}_\alpha \sum_{i=1}^n \rho_\tau(y_i - \alpha_0 - \alpha_1(x_i - x) - \ldots - \frac{1}{p!}\alpha_p(x_i - x)^p) \\ \hat{g}(\tau, x) &= \hat{\alpha}_0(\tau, x)\end{aligned}$$

Series Methods

$$\begin{aligned}\hat{\alpha}(\tau) &= \text{argmin}_\alpha \sum_{i=1}^n \rho_\tau(y_i - \sum_j \varphi_j(x_i)\alpha_j) \\ \hat{g}(\tau, x) &= \sum_{j=1}^p \varphi_j(x)\hat{\alpha}_j\end{aligned}$$

Penalty Methods

$$\hat{g}(\tau, x) = \text{argmin}_g \sum_{i=1}^n \rho_\tau(y_i - g(x_i)) + \lambda P(g)$$

Total Variation Regularization I

There are many possible penalties, but total variation of the first derivative of g is particularly attractive:

$$P(g) = V(g') = \int |g''(x)| dx$$

As $\lambda \to \infty$ we constrain g to be closer to linear in x. Solutions of

$$\min_{g \in \mathcal{G}} \sum_{i=1}^{n} \rho_\tau(y_i - g(x_i)) + \lambda V(g')$$

are continuous and piecewise linear, K, Ng and Portnoy (1994)

Total Variation Regularization II

For bivariate functions we consider the analogous problem:

$$\min_{g \in \mathcal{G}} \sum_{i=1}^{n} \rho_\tau(y_i - g(x_{1i}, x_{2i})) + \lambda V(\nabla g)$$

where the total variation variation penalty is now:

$$V(\nabla g) = \int \|\nabla^2 g(x)\| dx$$

Solutions are again continuous, but now they are piecewise linear on a triangulation of the observed x observations. Again, as $\lambda \to \infty$ solutions are forced toward linearity, K and Mizera (2004).

Additive Models: Putting it all together

We can combine such models:

$$\min_{g \in \mathcal{G}} \sum_{i=1}^{n} \rho_\tau(y_i - \sum_j g_j(x_{ij})) + \sum_j \lambda_j V(\nabla g_j)$$

- Components g_j can be univariate, or bivariate, but beyond lies dragons.
- Additivity is intended to muffle the curse of dimensionality.
- Linear terms are easily allowed.
- And shape restrictions like monotonicity and convexity/concavity can also be imposed.

Implementation in R

- Problems are typically large, very sparse linear programs.
- Optimization via interior point methods are quite efficient,
- Provided sparsity of the linear algebra is exploited, quite large problems can be estimated.
- The nonparametric `qss` components can be either univariate, or bivariate
- Each `qss` component has its own λ specified
- Linear covariate terms enter formula in the usual way
- The `qss` components can be shape constrained.

```
fit <- rqss(y ~ qss(x1,3) + qss(x2,8) + x3, tau = .6)
```

Tuning Parameter Selection

One way to interpret λ parameters is to note that they control the number of effective parameters of the fitted model.

$$p(\lambda) = \|\hat{\beta}(\lambda)\|_0 = \text{card}\{i : \hat{\beta}_i(\lambda) = 0\}$$

This is equivalent to the number of interpolated observations, the number of zero residuals.

$$p(\lambda) = \text{div } \hat{g}_{\lambda,\tau}(y_1, \ldots, y_n) = \sum_{i=1}^{n} \partial \hat{y}_i / \partial y_i$$

Childhood Malnutrition in India

A larger scale problem illustrating the use of these methods is a model of childhood malnutrition considered by Fenske, Kneib and Hothorn (2009).

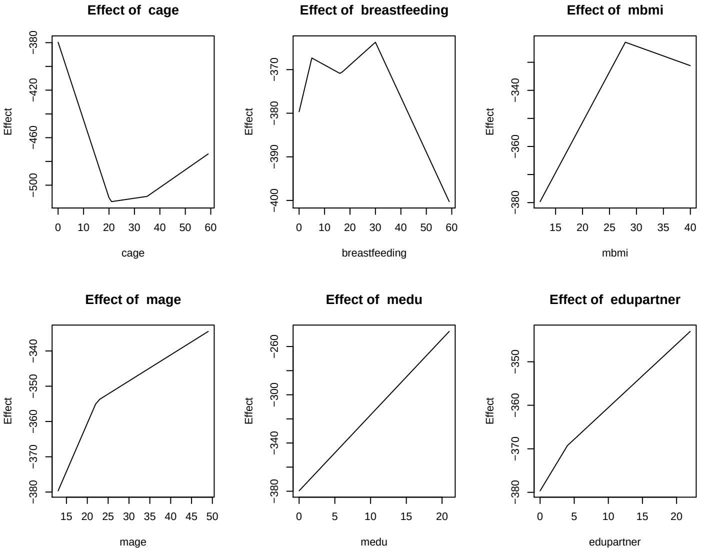
- They motivate the use of models for low conditional quantiles of height as a way to explore influences on malnutrition,
- They employ boosting as a model selection device,
- Their model includes six univariate nonparametric components and 15 other linear covariates.
- There are 37,623 observations on children's heights

R Formulation

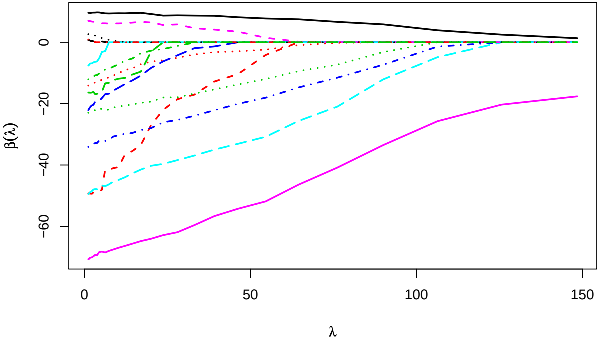
```
  rqss(stunting ~ csex + ctwin + cbirthorder + munemployed +
mreligion + mresidence + deadchildren + wealth + electricity +
radio + television + frig + bicycle + motorcycle + car +
qss(cage, lambda = lamss[1]) + qss(bfed, lambda = lamss[2]) +
qss(mbmi, lambda = lamss[3]) + qss(mage, lambda = lamss[4]) +
qss(medu, lambda = lamss[5]) + qss(fedu, lambda = lamss[6]) +
tau = 0.10, method = "lasso", lambda = lambda, data = india)
```

- The six coordinates of `lamss` control the smoothness of the nonparametric components,
- `lambda` controls the degree of shrinkage in the linear (lasso) coefficients.
- The estimated model has roughly 40,000 observations, including the penalty contribution, and has 2201 parameters.
- Fitting the model for a single choice of λ's takes approximately 5 seconds.

Nonparametric Components



Lasso Shrinkage of Linear Components



Of the 15 original covariates that were introduced linearly, the lasso selection with λ chosen to be 100 selects four: the gender of the child, whether the mother is employed, her religion, and whether she is "urban" or "rural." The final selected model has dimension $p(\lambda) = 43$.

Conclusions

- Quantile regression provides a unified approach to the estimation of conditional quantile functions just as least squares and related robust methods estimate models for conditional central tendency.
- In some applications it is useful to focus attention on covariate effects at low (or high) conditional quantiles of the response without assuming that these effects are the same at other quantiles.
- There are challenging new approaches to time-series analysis using quantile regression methods,
- Total variation roughness penalties offer an attractive approach to nonparametric quantile regression, and additive models are easily implemented combining nonparametric and linear parametric components.
- Conventional lasso penalties can be used as a model selection device for linear components, and AIC-like methods used to select model dimension.