Quantile Regression:An Introduction

Roger Koenker
Kevin Hallock

Abstract

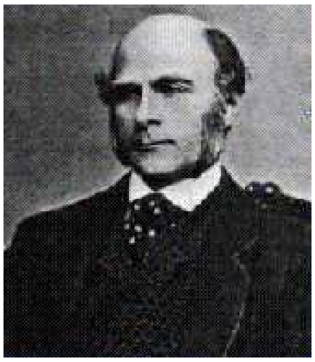
There is more to econometric life than is dreamt of in the philosophies of Gaussian regression and least squares estimation.

Journal of Economic Perspectives
Symposium on Econometrics Tools
New Orleans: January, 2001

Regression is Demeaning

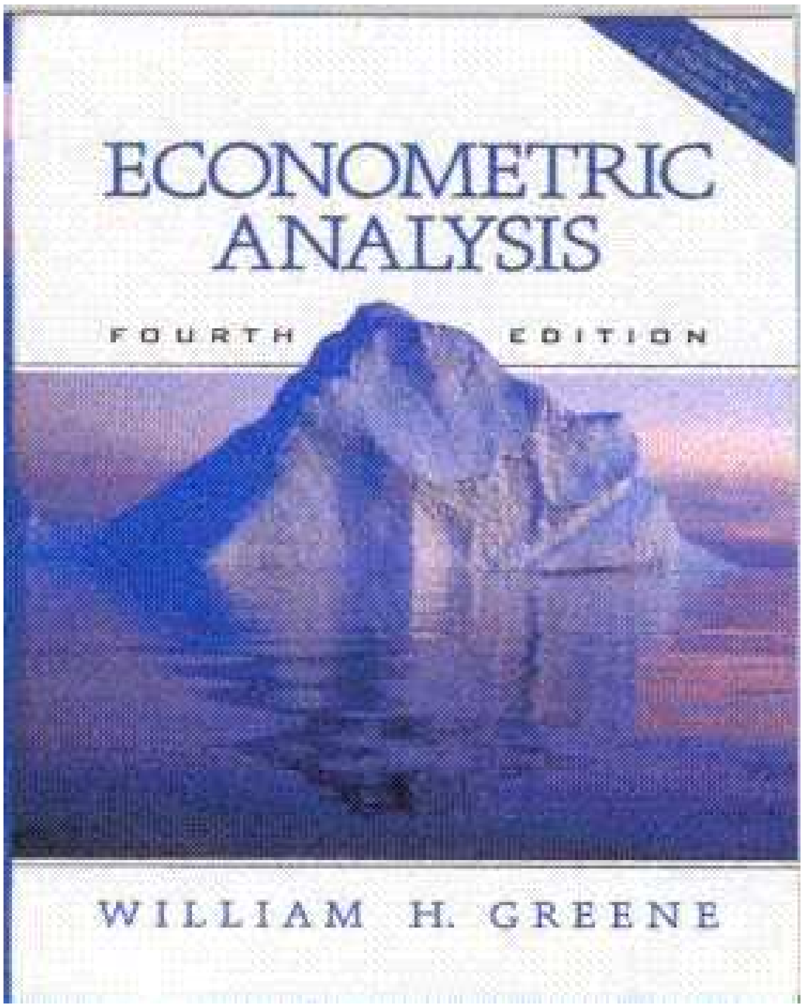
 'De mean is 'de meaning.

 Regression is demeaning.

 Regression is de-meaning

Quantile Regression

Just Enough to be Dangerous

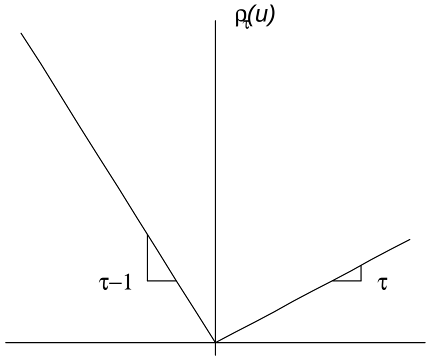


Quantile Regression

Univariate τth Quantile

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

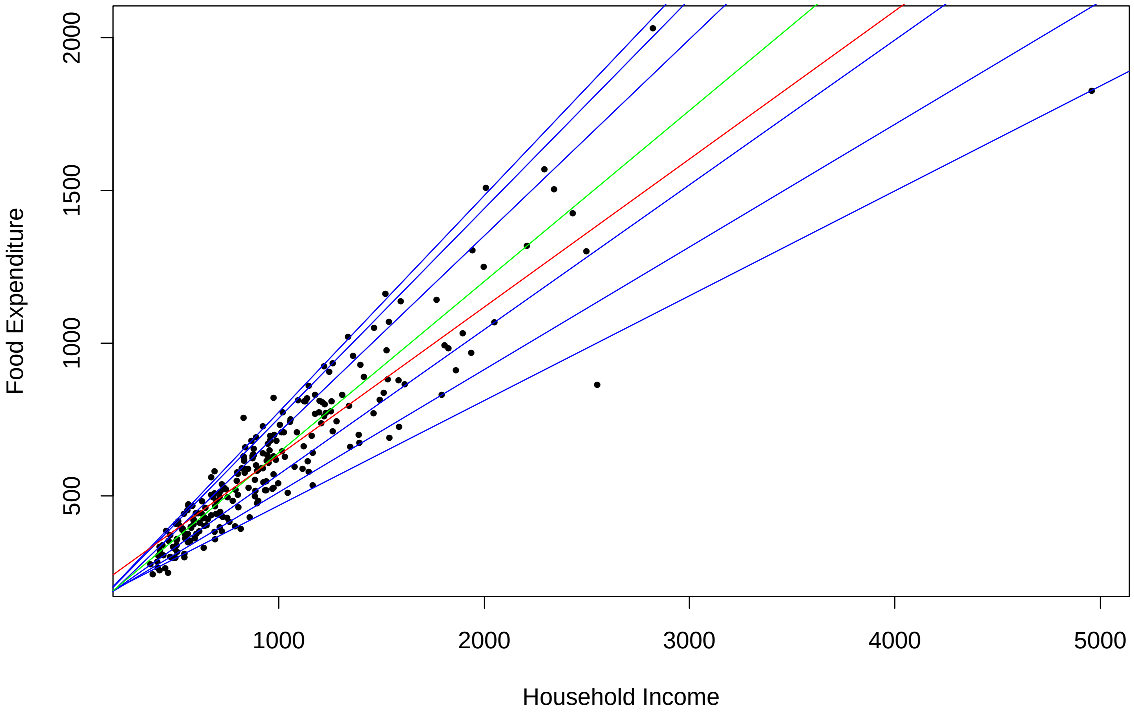
$$\text{where} \qquad \rho_\tau(u) = (\tau - I(u < 0))u$$



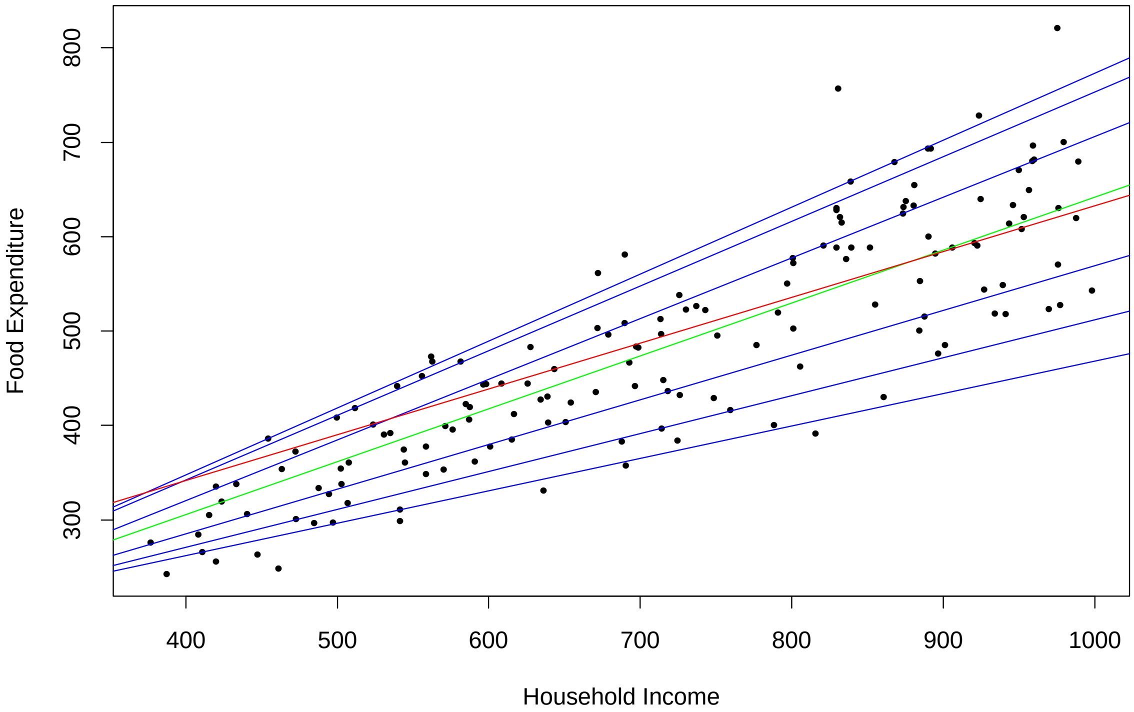
Regression τth Quantile

$$\hat{\beta}(\tau) = \mathrm{argmin}_{b \in \Re^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top b)$$

Engel's Food Expenditure Data

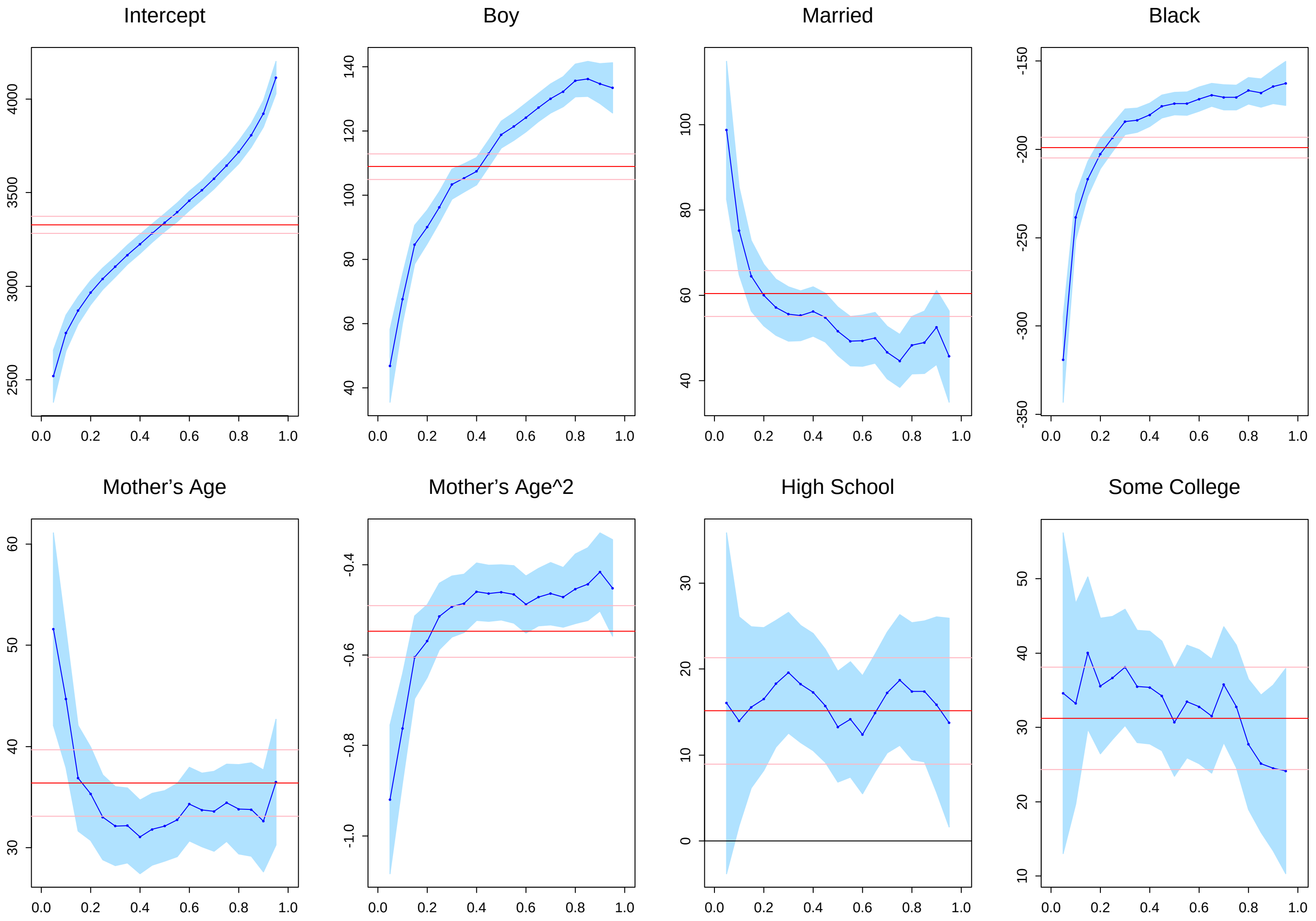


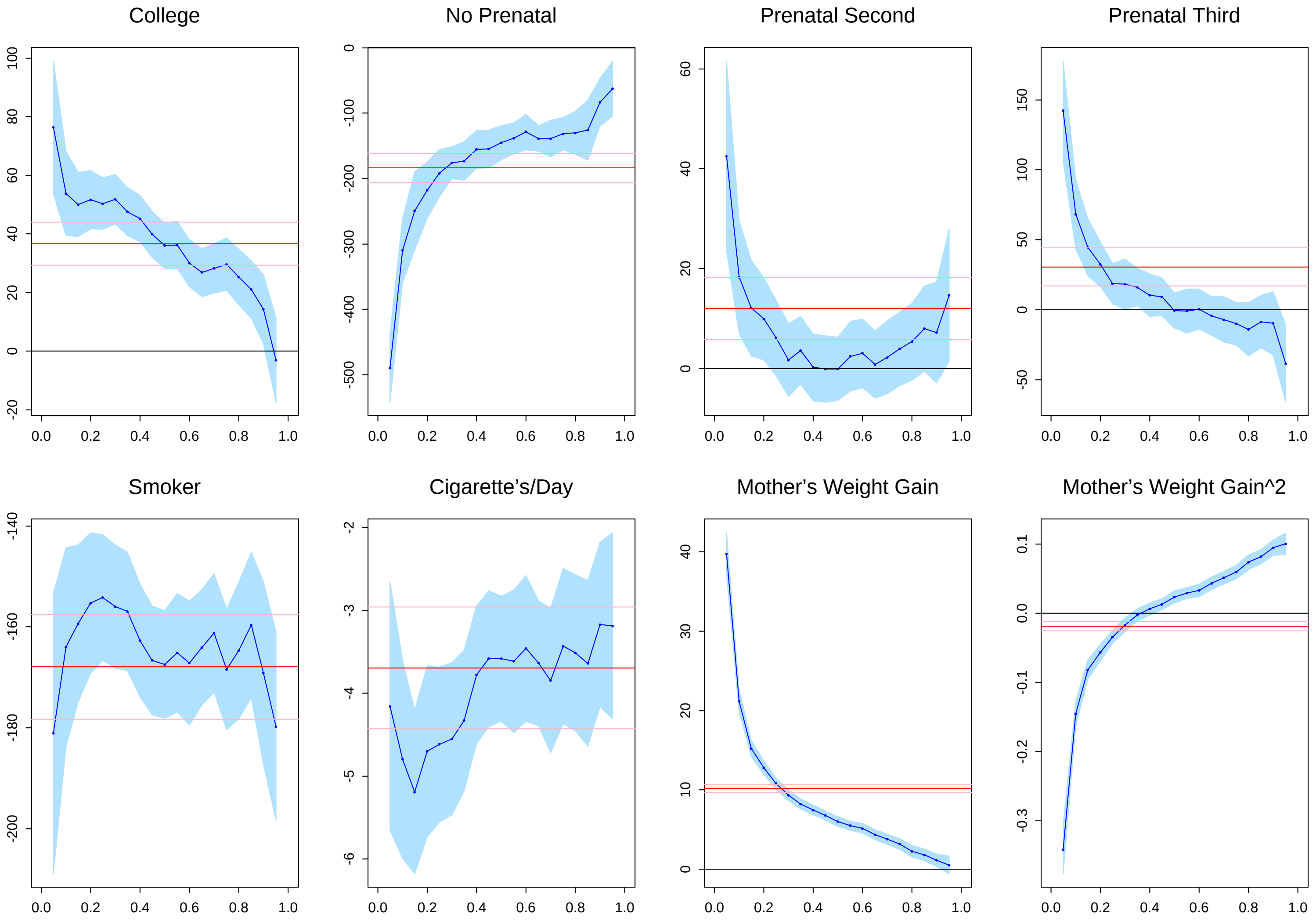
Engel's Food Expenditure Data



A Model of Infant Birthweight

- Reference: Abreveya (2001)
- Data: June, 1997, Detailed Natality Data of the US. Live, singleton births, with mother's 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


Intercept
Boy
Married
Black
Mother's Age
Mother's Age^2
High School
Some College
2500
3000
3500
4000
40
60
80
100
120
140
40
60
80
100
-350
-300
-250
-200
-150
30
40
50
60
-1.0
-0.8
-0.6
-0.4
0
10
20
30
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0


College
No Prenatal
Prenatal Second
Prenatal Third
Smoker
Cigarette's/Day
Mother's Weight Gain
Mother's Weight Gain^2

Conclusions

- Quantile regression methods complement established mean regression (least-squares) methods.
- By focusing on local slices of the conditional distribution, they offer a useful deconstruction of conditional mean models.
- They provide a more flexible role for covariate effects allowing them to influence location, scale *and shape* of the response distribution.
- In applications a variety of unobserved heterogeneity phenomena are rendered observable.