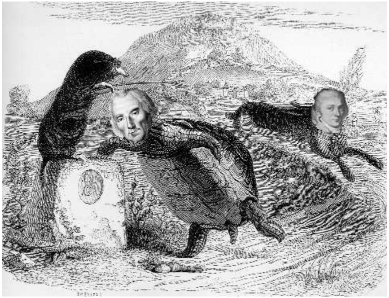
Quantile Regression: Past and Prospects

Roger Koenker

University of Illinois, Urbana-Champaign

Northern Illinois University: 31 October 2014

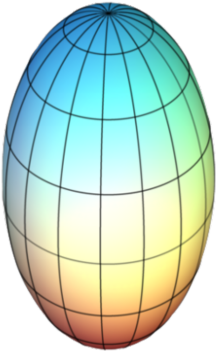


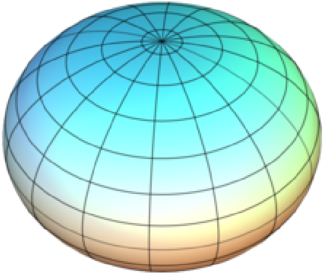
An Outline

- A Brief Prehistory of Regression
- Regression to Mediocrity
- Revenge of the Medians
- Why Conditional Quantiles?
- Some Recent Developments
- Quo Vadis?

A Burning Question of Early 18th Century Science

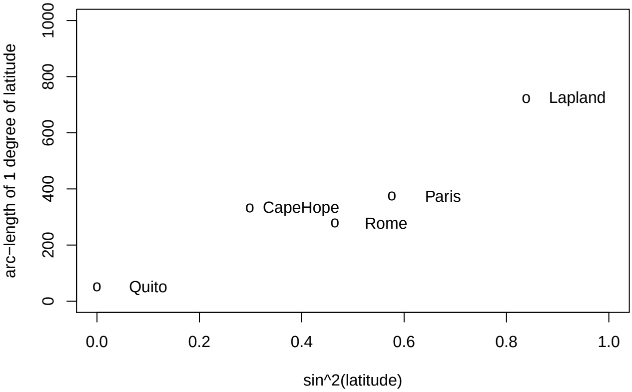
Does the earth look more like a lemon or a grapefruit? Prolate or Oblate?





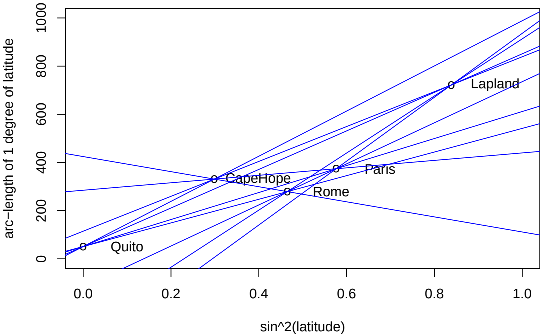
The French, based on extensive survey work by Cassini, maintained the prolate view while the English, based on gravitational theory of Newton, maintained the oblate view.

The First Scatterplot?



Grapefruit or Lemon? Boscovich (1755) considered the five measurements depicted above determined by arduous survey efforts. An upward slope in this figure indicates that the earth is oblate (like a grapefruit) rather than prolate (like a lemon).

The First Regression



Boscovich (1755) computed all the pairwise slopes and initially reported a trimmed mean of the pairwise slopes as a point estimate of the earth's ellipticity.

Boscovich's Second Approach

Ruder Josip Bošković (1711-1787)

Boscovich (1757) proposed estimating the ellipticity of the earth by solving:

$$\min \sum_{i=1}^{n} |y_i - \alpha - x_i'\beta| \text{ s. t. } \sum_{i=1}^{n} (y_i - \alpha - x_i'\beta) = 0,$$

The constraint allowed the problem to be reduced to the weighted median problem:

$$\min \sum_{i=1}^{n} |\tilde{y}_i - \tilde{x}_i'\beta|$$

where $\tilde{y}$ and $\tilde{x}$ denote deviations from their respective means.

Laplace's *Méthode de Situation*

Pierre-Simon Laplace (1749-1827)

Laplace showed that Boscovich's problem: could be solved by ordering the candidate slopes $\{\tilde{y}_i/\tilde{x}_i\}$ and finding the weighted median using weights $w_i = |\tilde{x}_i|$, i.e. finding the smallest j such that,

$$\sum_{i=1}^{j} w_{(i)} > \frac{1}{2} \sum_{i=1}^{n} w_{(i)},$$

where $w_{(i)}$ denotes weights ordered according to the order of the slopes.

Laplace "does the asymptotics" based on the DeMoivre-Laplace CLT.

Gauss Spoils the Party

Carl Friedrich Gauss (1777-1855)

Gauss's discovery of least-squares replaced absolute error by squared error,

$$\min \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

To explain why errors were squared he invented the Gaussian Law of Errors:
Errors have density,

$$\varphi(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(u^2/2\sigma^2)$$

looking like Napolean's hat.

Napolean's Hat

Is it possible that the modern fixation on the "normal" distribution isn't a leap of Gaussian faith, but instead an act of obeisance to Napolean I?

The Father of the Average Man

Adolphe Quetelet (1796 - 1874)

Meanwhile, Quetelet's (1835) *Sur l'Homme (On Man)* iconified the "Average Man".

Through systematic measurement and relentless averaging Quetelet sought to extract man's essential qualities: social, economic, aesthetic, and moral.

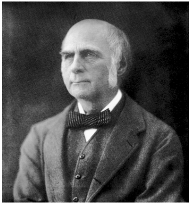
The Mother of the Average Man

Florence Nightingale (1820 - 1910)

Heroine of the Crimean War, Patron Saint of Nurses, admirer of Quetelet, and champion of the scientific, i.e. statistical, study of society.

To Nightingale every piece of legislation was an experiment in the laboratory of society deserving study and demanding evaluation.

The [Re]Discoverer of the Average Man



Francis Galton (1822 - 1911)

Galton's (1885) discovery of regression and correlation paved the way for a more sophisticated (conditional) view of the Average Man as a social scientific construct. But especially in his early work Galton extensively employed median and interquantile range methods.

Galton Had Doubts about the Average Man

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, p. 62]*

The Revenge of the Median Man

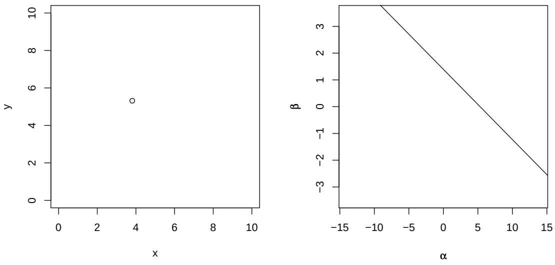
Francis Ysidro Edgeworth (1845 - 1926)

Edgeworth disdained the growing reliance on the *gens d'armes hat* (Gaussian Law of Errors) and revived the median regression methods pioneered by Boscovich and Laplace in the 18th century. But Edgeworth removed Boscovich's intercept constraint.

Edgeworth's (1888) Dual Plot: Anticipating Simplex

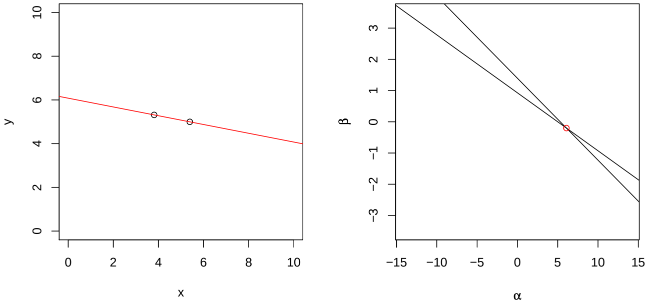
Points in sample space map to lines in parameter space.

$$(x_i, y_i) \mapsto \{(\alpha, \beta) : \alpha = y_i - x_i\beta\}$$



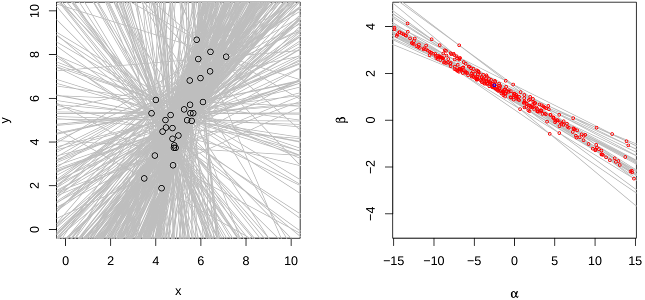
Edgeworth's (1888) Dual Plot: Anticipating Simplex

Lines through pairs of points in sample space map to points in parameter space.



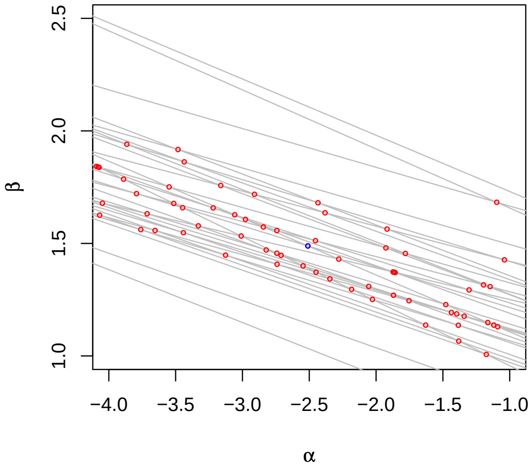
Edgeworth's (1888) Dual Plot: Anticipating Simplex

All pairs of observations produce $\binom{n}{2}$ points in dual plot.



Edgeworth's (1888) Dual Plot: Anticipating Simplex

Follow path of steepest descent through points in the dual plot.



But Not Everyone was Convinced

George Waddel Snedecor (1881 - 1974)

The first edition of Snedecor's influential textbook contains:

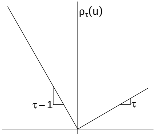
> *The median-quartile description of a sample falls into the same category as the range, furnishing a rough-and-ready means of summarizing the data. It has never gained favor among biologists both because it provides no accurate test of significance, and because it leads into a blind alley so far as more advanced statistical methods are concerned.*

Sample Quantiles via Optimization

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

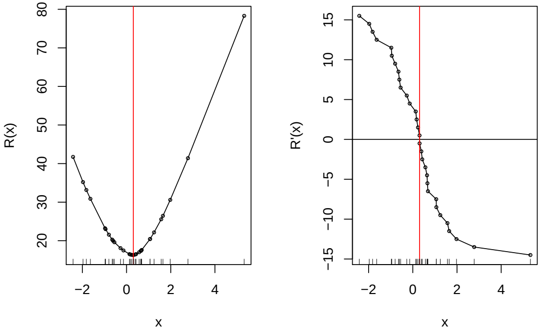
$$\hat{\alpha}(\tau) = \text{argmin}_a \sum_{i=1}^{n} \rho_\tau(y_i - a)$$

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



The τ-tilt biases the argmin toward making the lower cost error; e.g. forecasting flood levels.

Optimization for Sample Quantiles



The quantile objective function is piecewise linear, nicely convex, so its derivative is monotone and optimization is easy.

The Least Squares Meta-Model

Once one realizes that quantiles may be defined via convex optimization the rest of the quantile regression story is almost trivial.
The unconditional mean solves

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

The Least Squares Meta-Model

Once one realizes that quantiles may be defined via convex optimization the rest of the quantile regression story is almost trivial.

The unconditional mean solves

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

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

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

The Least Squares Meta-Model

Once one realizes that quantiles may be defined via convex optimization the rest of the quantile regression story is almost trivial.
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, α_τ, solves

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

The Least Squares Meta-Model

Once one realizes that quantiles may be defined via convex optimization the rest of the quantile regression story is almost trivial.
The unconditional mean solves

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

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

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

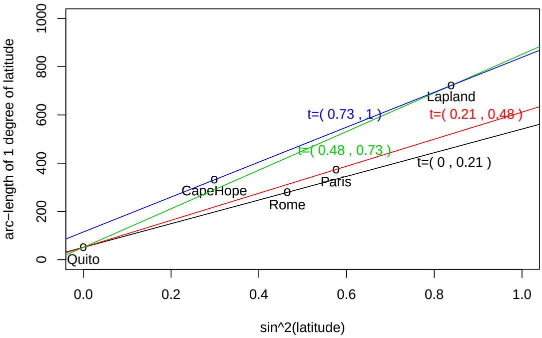
Similarly, the unconditional τth quantile, α_τ, solves

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

and the conditional τth quantile, $\alpha_\tau(x)$, solves

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

Earth's Ellipticity: A Quantile Regression View



The quantile regression analysis of the Boscovich data identifies four distinct pairs of points that solve the weighted ℓ_1 problem for various intervals of the parameter τ.

Why Conditional Quantiles?

The initial response to the "Regression Quantiles" paper with Gib Bassett was not terribly auspicious:

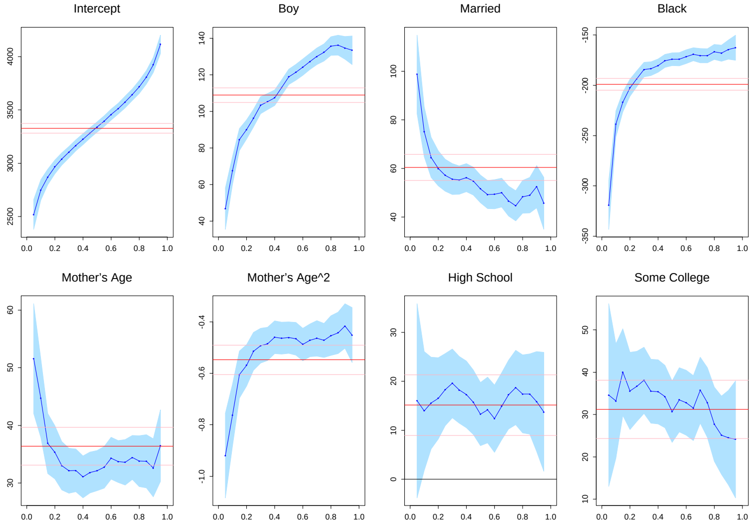
> *I regret that I cannot see any point in this paper, and therefore cannot recommend its publication. It may be of interest to compute regression analyses to minimize the sum of absolute deviations between the observed and fitted responses, and there is a fair amount of literature on this topic. But why should one consider $\tau \neq 1/2$?*
>
> *Annals of Statistics* Referee Report, November, 1975.

Fortunately, others were kinder and more understanding: Notably Steve Portnoy, Joe Gastwirth, Jana Jurečkova, Ed Leamer.

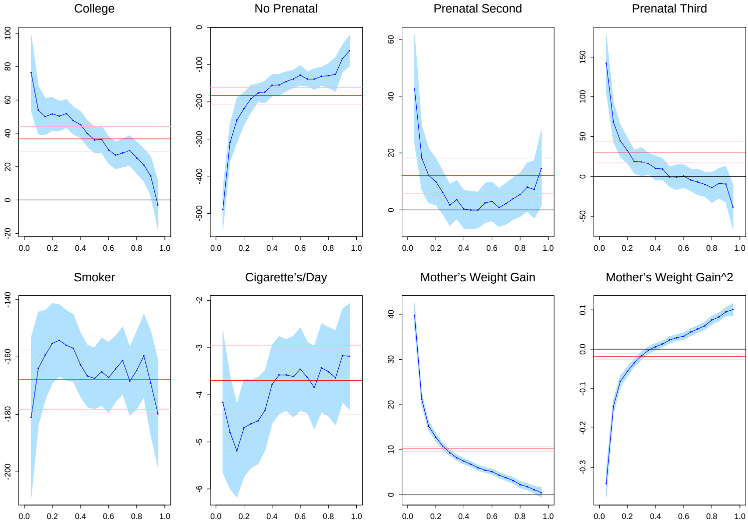
A Model of Infant Birthweight

- 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



Developments in the Twilight Zone

In my 2005 monograph I relegated several topics to a penultimate chapter called “The Twilight Zone of Quantile Regression” including:

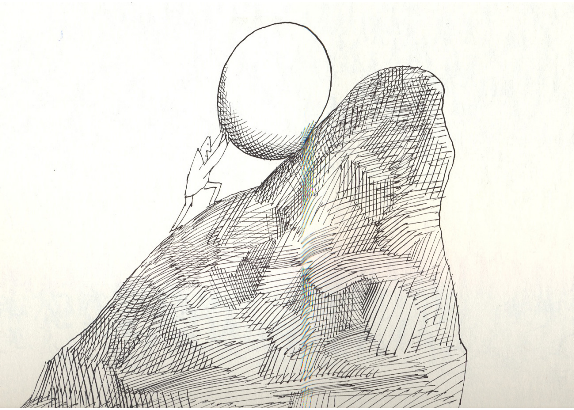
- Survival Analysis
- Discrete Response
- Quantile Autoregression
- Multivariate Quantiles
- Longitudinal Data Analysis
- Causal Models and Endogenous Treatment
- Choquet Risk and Portfolio Allocation

Beyond the Twilight Zone

There are many new directions that deserve further exploration:

- High dimension model selection and inference
- Heterogeneous treatment effects and mixture models
- Frequency domain time series
- Functional data analysis
- Measurement error and selection models
- Gradient descent computational methods
- Survival models for recurrent events and competing risks
- ...

Statistics in the 20th Century



Statistics in the 21st Century