Advanced Time Series Analysis

Quantile Regression

Henrik Madsen

Lecture

This Presentation

- Quantiles
- Conditional quantiles
- Quantile regression
- Linear models
- Non-linear models
- Example: Quantile regression based probabilistic wind power forecasting
- Performance of quantiles
- Adaptive quantile regression

Quantiles

- Let Y be a random variable with distribution function F_Y and τ a real number between zero and one.
- The τ^{th} quantile of F_Y is

$$q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\}$$

  Thus 100τ pct. of the probability mass of Y is below $q_Y(\tau)$
- Special quantiles are the first quartile $q_Y(0.25)$, the second quartile (median) $q_Y(0.5)$, the third quartile $q_Y(0.75)$, and the percentiles $q_Y(0.01), q_Y(0.02), ..., q_Y(0.99)$.
- The interquartile range (IQR) is $q_Y(0.75) - q_Y(0.25)$.

Determination of quantiles

- Consider the following objective function

$$
\begin{aligned}
V(q) &= \tau \int_{y>q} |y-q| dF_Y(y) + (1-\tau) \int_{y<q} |y-q| dF_Y(y) && (1) \\
&= \tau \int_{y>q} (y-q) dF_Y(y) - (1-\tau) \int_{y<q} (y-q) dF_Y(y) && (2)
\end{aligned}
$$

- The solution to $\min_q V(q)$ is the τ^{th} quantile of F_Y.
- *Proof:* The first order condition for a minimization of V is

$$
\begin{aligned}
0 &= -\tau \int_{y>q} dF_Y(y) + (1-\tau) \int_{y<q} dF_Y(y) && (3) \\
&= -\tau(1 - F_Y(q)) + (1-\tau) F_Y(q) && (4) \\
&= -\tau + F_Y(q) && (5)
\end{aligned}
$$

Conditional quantiles

- Consider now $Y|X$ with the conditional distribution $F_{Y|X}(y)$, its τ^{th} quantile is clearly a function of X.
- This quantile solves

$$\min_q \{\tau \int_{y>q} |y-q| dF_{Y|X}(y) + (1-\tau) \int_{y<q} |y-q| dF_{Y|X}(y)\} \quad (6)$$

- If q is considered as a linear function $q(X) = X'\beta$, unknown up to the parameter vector β, (6) is equivalent to the following minimization problem:

$$\min_\beta \{\tau \int_{y \geq X'\beta} |y - X'\beta| dF_{Y|X}(y) + (1-\tau) \int_{y<X'\beta} |y - X'\beta| dF_{Y|X}(y)\} \quad (7)$$

The solution to (7) is denoted as β_τ, from which we can obtain the τ^{th} quantile as

$$q_\tau(X) = X'\beta_\tau \quad (8)$$

Quantile Regression

- Given the data $(y_t, x_t')', t = 1, ..., N$ where x_t is $k \times 1$, and consider the linear model

$$y_t = x_t'\beta + e_t \tag{9}$$

for the particular conditional quantile of y_t.

- Considering the sample counterpart of the previous slides.

  The τ^{th} quantile regression estimator of β can be found by minimizing the sample counterpart of Eq. (7) as the asummetrically weighted absolute errors with weight τ on positive errors and weight $(\tau - 1)$ on negative errors:

$$V_N(\beta, \tau) = \frac{1}{N}\{\tau \sum_{t:y \geq x_t'\beta} |y - x_t'\beta| + (1-\tau) \sum_{t:y < x_t'\beta} |y - x_t'\beta| \tag{10}$$

- For $\tau = 0.5$, 2 times (10) is exactly the objective function for Least Absolute Deviation (LAD) estimation or 'median regression'.

Quantile Regression – Compact form

- Introduce the "check" function ρ_τ such that $\rho_\tau(e) = \tau e$ if $e \geq 0$ and $\rho_\tau(e) = (\tau - 1)e$ if $e < 0$.
- Now the criterion can be written in compact form

$$
\begin{aligned}
V_N(\beta, \tau) &= \frac{1}{N} \sum_{t=1}^{N} \rho_\tau(y_t - x_t'\beta) && (11) \\
&= \frac{1}{N} \sum_{t=1}^{N} (\tau - 1_{y_t - x_t'\beta < 0})(y_t - x_t'\beta) && (12)
\end{aligned}
$$

- Once $\hat{\beta}_\tau$ is obtained, the estimated quantile regression hyperplane is computed as $x_t'\hat{\beta}_\tau$, and the quantile regression residuals as $\hat{e}_t(\tau) = y_t - x_t'\hat{\beta}_\tau$.

Quantile (Linear) Regression

In this context quantile regression is a linear model

$$y_t = \hat{Q}(\tau, \mathbf{x}_t) + r_t = \mathbf{x}_t^T \hat{\beta} + r_t \tag{13}$$

with a peicewise linear and asymmetric loss function

$$\rho_\tau(r) = \begin{cases} \tau r & , \quad r \geq 0 \\ (\tau - 1) r & , \quad r < 0 \end{cases} \tag{14}$$

given N obeservations of y and $\mathbf{x}$, the best estimate of β is

$$\hat{\beta}(\tau) = \arg\min_\beta \sum_{i=1}^{N} \rho_\tau(r_i) = \arg\min_\beta S(\beta; \tau, \mathbf{r}) \tag{15}$$

As shown previously this construction give the τ quantile.

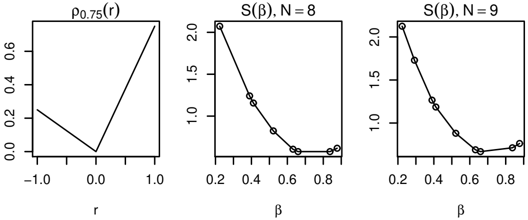
Some issues related to Quantile

A quantile in which sense?

$$\frac{1}{N}\#(r < 0) \quad \leq \tau \leq \quad \frac{1}{N}\#(r \leq 0)$$

$$\frac{1}{N}\#(r > 0) \quad \leq 1 - \tau \leq \quad \frac{1}{N}\#(r \geq 0)$$

Or quantile regression divide the response in a proper way, and generalize the sample quantile.



The Quantile Model

We do not always believe that linear regression give a realistic model a general with $\mathbf{x} \in \mathbb{R}^p$ model is

$$Q(\mathbf{x}; \tau) = g(\mathbf{x}; \tau) \tag{16}$$

An additive model

$$Q(\mathbf{x}; \tau) = \alpha(\tau) + \sum_{j=1}^{p} f_j(x_j; \tau) \tag{17}$$

An additive model with splines

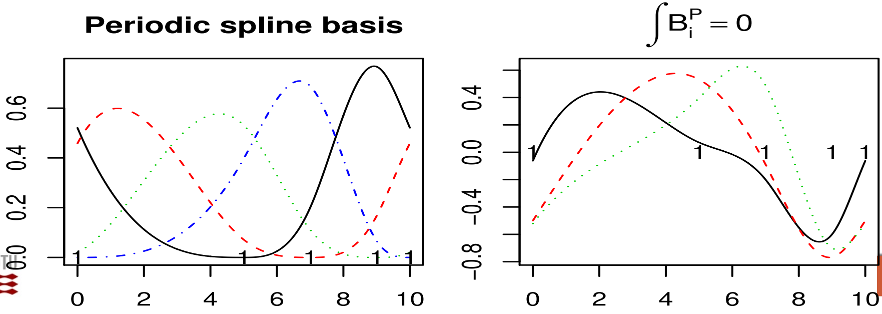
$$Q(\mathbf{x}; \tau) = \alpha(\tau) + \sum_{j=1}^{p} \sum_{k=1}^{n_k} b_{jk}(x_j)\beta_{jk} \tag{18}$$

b_{jk} are spline basis functions. The model is now linear in the known basis functions.

Splines in Short

But what is a spline basis function?

- Splines are polynomials of degree m on intervals defined by a knot sequence and it is $m-1$ times differentiable over the knots.
- The knots defines the splines, the more knots the greater the flexibility.
- Different end point conditions can be imposed, by controlling the knotssequence and using linear combination of the basis functions.



Example – The Data Set

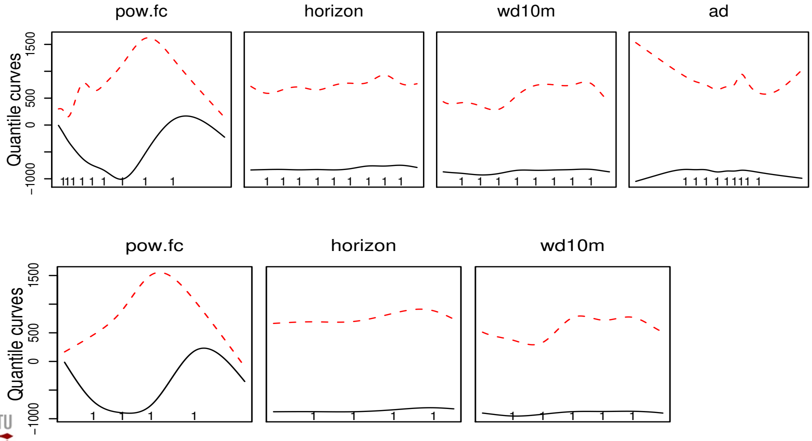
The data set consists of data from a wind power plant placed at Tunø Knob. The data are

- Forecated power form WindFor (previously called WPPT)
- Prediction error from WindFor
- Meteorological forecasts from DMI (Danish MET service)

The aim is now to model quantile curves of the prediction error as a function of meteorological forecasts and forecasted power.
This data set is divided into a training set and a test set of approximal equal sizes. Will be used later for adaptive quantile regression.

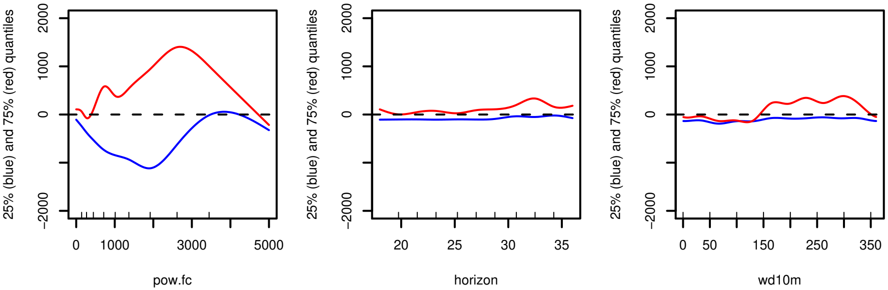
Quantile Regression with Splines

Two models:



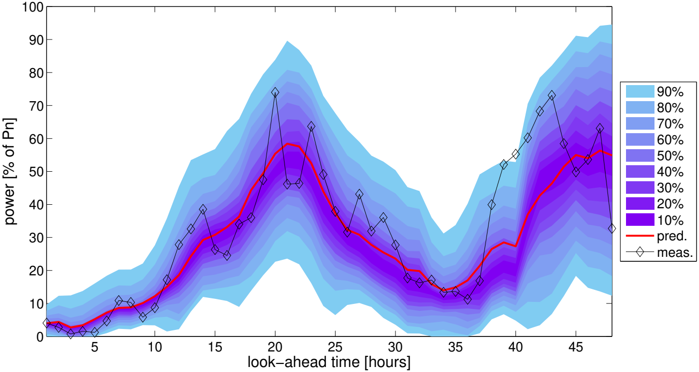
Quantile regression - Example - Klim wind farm

Effect of variables (- the functions are approximated by Spline basis functions):



- Forecasted power has a large influence.
- The effect of horizon is of less importance.
- Some increased uncertainty for Westerly winds.

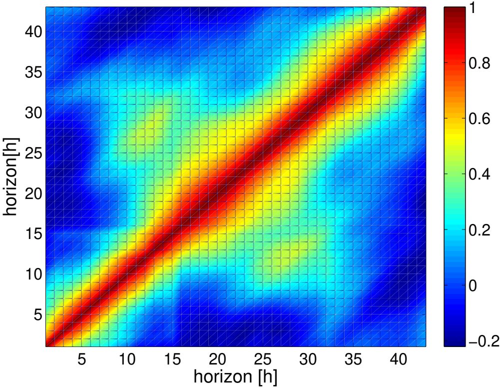
Example: Probabilistic forecasts - Klim wind farm



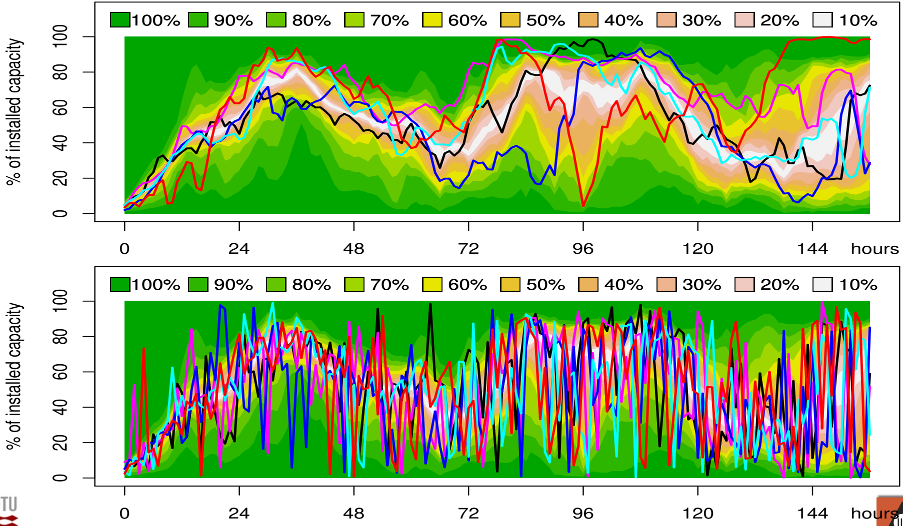
- Notice how the confidence intervals varies ...
- But the correlation in forecasts errors is not described so far.

Correlation structure of forecast errors

- It is important to model the **interdependence structure** of the prediction errors.
- An example of interdependence covariance matrix:



Correct (top) and naive (bottom) scenarios



An Adaptive Approach

The idea is to use knowledge about the solution at time t to find a solution at time $t+1$, based on some forgetting factor.
What is needed is

- An LO formulation for the QR problem
- The simplex algorithm
- An algorithm for updating the design matrix

For details please see:
Møller, J. K., Nielsen, H. A., & Madsen, H. (2008). Time-adaptive quantile regression. Computational Statistics and Data Analysis, 52(3), 1292-1303.

The Linear Optimization Formulation

The LO formulation of the Quantile Regression problem

$$\min\{\tau \mathbf{e}^T \mathbf{r}^+ + (1-\tau)\mathbf{e}^T \mathbf{r}^- : \mathbf{X}\beta + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{y}, (\mathbf{r}^+, \mathbf{r}^-) \in \mathbb{R}_+^{2N}, \beta \in \mathbb{R}^K\}$$

or in a more compact notation

$$\min\{\mathbf{c}^T \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{y}, (\mathbf{r}^+, \mathbf{r}^-) \in \mathbb{R}_0^{2N}, \beta \in \mathbb{R}^K\} \tag{19}$$

the equality constraints can be written as

$$\mathbf{A}\mathbf{x} = \mathbf{B}\mathbf{x}_{\mathcal{B}} + \mathbf{C}\mathbf{x}_{\mathcal{C}} = \mathbf{y} \tag{20}$$

with $\mathbf{x}_{\mathcal{C}} = \mathbf{0}$. $\mathbf{B}$ is a square matrix with several of thousand rows, the expensive part of the simplex algorithm is to find $\mathbf{B}^{-1}$.

Simplex and Quantile Regression

By interchanging of row $\mathbf{B}$ can be written as

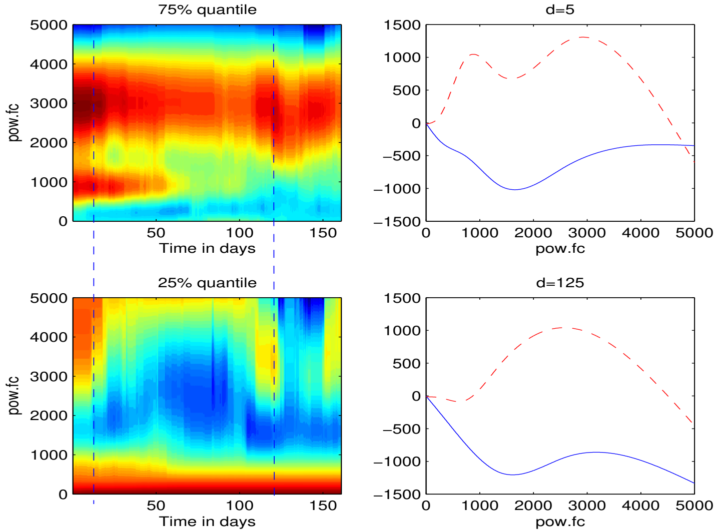
$$\mathbf{B} = \begin{bmatrix} \mathbf{X}(h) & \mathbf{0} \\ \mathbf{X}(\bar{h}) & \mathbf{P} \end{bmatrix} \tag{21}$$

and the inverse of $\mathbf{B}$ can be written as

$$\mathbf{B}^{-1} = \begin{bmatrix} \mathbf{X}(h)^{-1} & \mathbf{0} \\ -\mathbf{P}\mathbf{X}(\bar{h})\mathbf{X}(h)^{-1} & \mathbf{P} \end{bmatrix} \tag{22}$$

The number of rows in $\mathbf{X}(h)$ is less than 40 for all the models considered here. This gives large improvements w.r.t. both calculation time and numerical stability.

An Adaptive Model



Reliability Again

Overall Reliability

Model	**Reference**	**Basic 1**	**Basic 2**	**Basic 3**	**Basic4**
Below **75% (test)**	83.9%	78.6%	**77.2%**	77.6%	**77.2%**
Below **25% (test)**	22.6%	22.9%	25.7%	25.1%	**25.0%**

Local Reliability