

# Application of quantile regression for Metheorological Time Series modelling

Marcelo

November 12, 2016

Introduction

Linear Models for the Quantile Autoregression

Quantile Autoregression with a nonparametric approach

Next steps

# Introduction

# Introduction

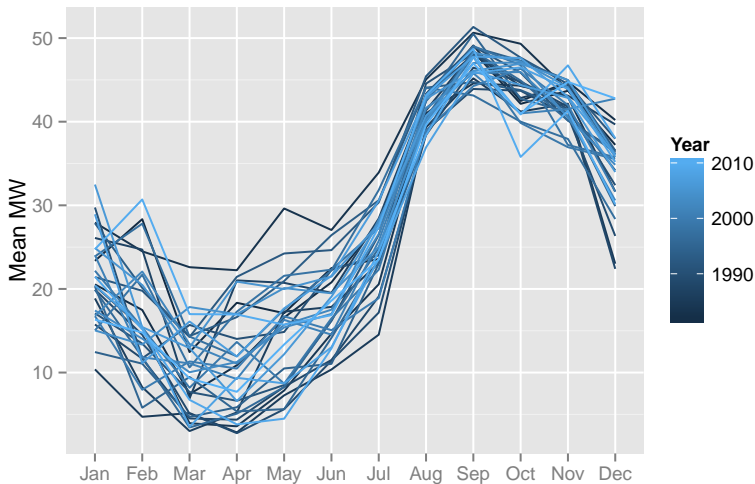
- ▶ Wind Firm Energy Certificate (FEC) (Porrúa, 2010) estimation imposes several challenges:
  - ▶ It is a quantile function of an aleatory quantity, named here on wind capacity factor (WP). Due to its non-dispatchable profile, accurate scenario generation models could reproduce a fairly dependence structure in order to the estimation of FEC.
  - ▶ Second, as it is a quantile functions, the more close to the extremes of the interval, the more sensitive to sampling error.

# Introduction

- ▶ The main frameworks we investigate are parametric linear models and a non-parametric regression: both for quantile regressions.
- ▶ In all approaches we use only time series lags as the regression covariates.
- ▶ To study our methods performance, we use the mean power monthly data of Icaraizinho, a wind farm located in the Brazilian northeast.

# The Icaraizinho dataset

- ▶ The Icaraizinho dataset consists of monthly observations from 1981 to 2011 of mean power measured in Megawatts.



## Quantile Regression

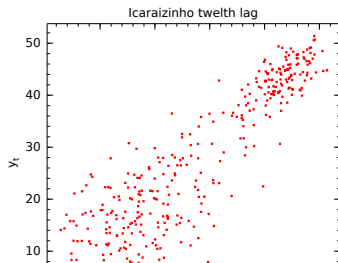
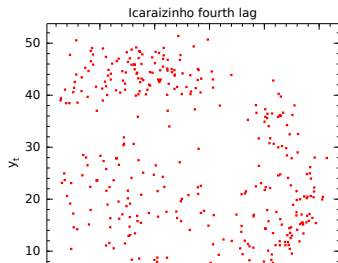
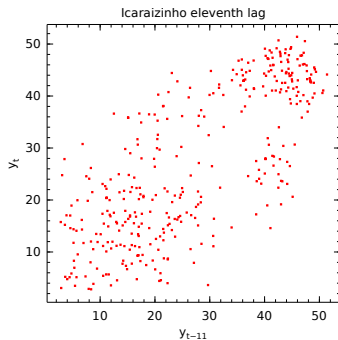
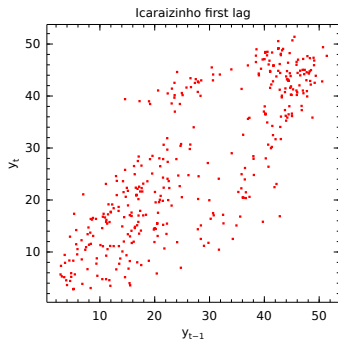
A Quantile Regression for the  $\alpha$ -quantile is the solution of the following optimization problem:

$$\min_q \sum_{t=1}^n \alpha |y_t - q(x_t)|^+ + (1 - \alpha) |y_t - q(x_t)|^-,$$

where  $q(x_t)$  is the estimated quantile value at a given time  $t$  and  $|x|^+ = \max\{0, x\}$  and  $|x|^- = -\min\{0, x\}$ . To model this problem as a Linear Programming problem, thus being able to use a modern solver to fit our model, we can create variables  $\varepsilon_t^+$  e  $\varepsilon_t^-$  to represent  $|y - q(x_t)|^+$  and  $|y - q(x_t)|^-$ , respectively. So we have:

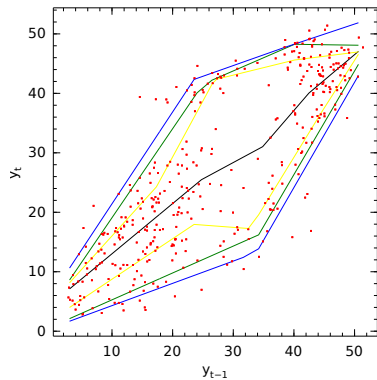
$$\begin{aligned} \min_{q, \varepsilon_t^+, \varepsilon_t^-} \quad & \sum_{t=1}^n \left( \lambda \varepsilon_t^+ + (1 - \lambda) \varepsilon_t^- \right) \\ \text{s.t.} \quad & \varepsilon_t^+ - \varepsilon_t^- = y_t - q(x_t), & \forall t \in \{1, \dots, n\}, \\ & \varepsilon_t^+, \varepsilon_t^- \geq 0, & \forall t \in \{1, \dots, n\}. \end{aligned}$$

# Relationship between $y_t$ and some lags

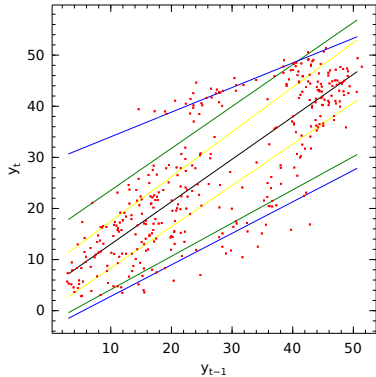




## Relationship between $y_t$ and some lags



$\lambda = 10$



Degenerated case with  $\lambda = 200$

## Linear Models for the Quantile Autoregression

# Best subset selection with Mixed Integer Programming

- We investigate the usage of Mixed Integer Programming to select which variables are included in the model, up to a limit of inclusions imposed *a priori*. The optimization problem is described below:

$$\begin{aligned} \min_{\beta_0, \beta, z, \varepsilon_t^+, \varepsilon_t^-} \quad & \sum_{t=1}^n \left( \alpha \varepsilon_t^+ + (1 - \alpha) \varepsilon_t^- \right) \\ \text{s.t.} \quad & \varepsilon_t^+ - \varepsilon_t^- = y_t - \beta_0 - \sum_{p=1}^P \beta_p x_{t,p}, & \forall t \in \{1, \dots, n\}, \\ & \varepsilon_t^+, \varepsilon_t^- \geq 0, & \forall t \in \{1, \dots, n\}, \\ & -M_U z_p \leq \beta_p \leq M_U z_p, & \forall p \in \{1, \dots, P\}, \\ & \sum_{p=1}^P z_p \leq K, \\ & z_p \in \{0, 1\}, & \forall p \in \{1, \dots, P\}. \end{aligned}$$

## Best subset selection with Mixed Integer Programming

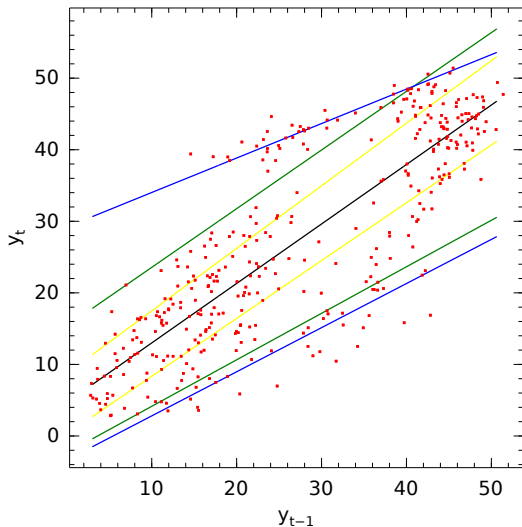


Figure 3: Linear Quantile Regression when only  $y_{t-1}$  is used

# Best subset selection with Mixed Integer Programming

	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8
$\beta_0$	-15.33	9.38	1.48	1.34	8.72	-1.68	4.94	0.65
$\beta_1$	-0.00	0.79	0.66	0.58	0.46	0.40	0.48	0.46
$\beta_2$	-0.00	-0.00	-0.00	-0.00	-0.00	0.33	-0.00	-0.00
$\beta_3$	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.20
$\beta_4$	-0.00	-0.47	-0.28	-0.27	-0.29	-0.35	-0.31	-0.40
$\beta_5$	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
$\beta_6$	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.11	0.08
$\beta_7$	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
$\beta_8$	-0.00	-0.00	-0.00	-0.00	-0.15	-0.00	-0.31	-0.26
$\beta_9$	-0.00	-0.00	-0.00	-0.00	-0.00	0.14	0.16	0.20
$\beta_{10}$	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
$\beta_{11}$	-0.00	-0.00	0.26	0.17	0.21	0.08	0.16	0.19
$\beta_{12}$	1.17	-0.00	-0.00	0.18	0.15	0.19	0.22	0.20

Table 1: Coefficients for quantile  $\alpha = 0.05$

## Best subset selection with a $\ell_1$ penalty

- ▶ Another way of doing regularization is including the  $\ell_1$ -norm of the coefficients on the objective function.
- ▶ By lowering the penalty we impose on the  $\ell_1$ -norm, more variables are being added to the model.
- ▶ This is the same strategy of the LASSO, and its usage for the quantile regression is discussed in Li and Zhu (2012).
- ▶ The proposed optimization problem to be solved is:

$$\min_{\beta_0, \beta} \sum_{t=1}^n \alpha |y_t - q(x_t)|^+ + (1 - \alpha) |y_t - q(x_t)|^- + \lambda \|\beta\|_1$$

$$q(x_t) = \beta_0 - \sum_{p=1}^P \beta_p x_{t,p},$$

## Best subset selection with a $\ell_1$ penalty

- In order to represent the above problem to be solved with linear programming solver, we restructure the problem as below:

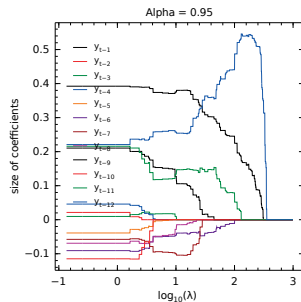
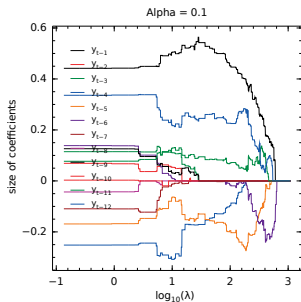
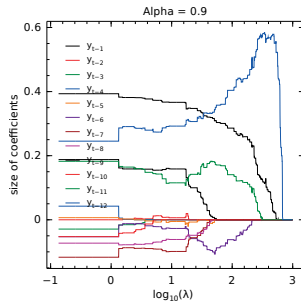
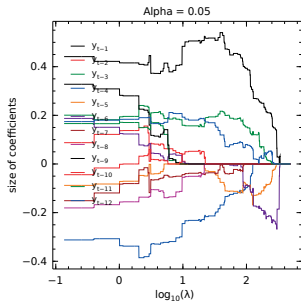
$$\begin{aligned}\beta_{\lambda}^{*LASSO} = \operatorname{argmin}_{\beta_0, \beta, \varepsilon_t^+, \varepsilon_t^-} & \sum_{i=1}^n \left( \alpha \varepsilon_t^+ + (1 - \alpha) \varepsilon_t^- \right) + \lambda \sum_{p=1}^P \xi_p \\ \varepsilon_t^+, \varepsilon_t^- & \geq 0, \quad \forall t \in \{1, \dots, n\}, \\ \varepsilon_t^+, \varepsilon_t^- & \geq 0, \quad \forall t \in \{1, \dots, n\}, \\ \xi_p & \geq \beta_p, \quad \forall p \in \{1, \dots, P\}, \\ \xi_p & \geq -\beta_p, \quad \forall p \in \{1, \dots, P\},\end{aligned}$$

## Best subset selection with a $\ell_1$ penalty

- For low values of  $\lambda$ , the penalty is small and thus we have a model where all coefficients have a nonzero value. - When  $\lambda$  is increased, the coefficients shrink towards zero (as an extreme case we have a constant model) - The linear coefficient  $\beta_0$  is not penalized. - We make this experiment For the same quantiles values of  $\alpha$  we experimented on the previous section ( $\alpha \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$ ).



# Best subset selection with a $\ell_1$ penalty



# Simulation Study

- ▶ We propose simulating an AR(1) model

$$y_t = \phi_0 + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (1)$$

and test two approaches to predict the one-step ahead quantile.

- ▶ On the first one, we consider known this process true model, thus estimating values for  $\hat{\phi}_0$ ,  $\hat{\phi}$  and  $\hat{\sigma}_\varepsilon^2$ .
- ▶ On the second approach, we use the quantile regression to make a direct estimation of the quantiles.

# Simulation Study

# Simulation Study

## Quantile Autoregression with a nonparametric approach

## Formulation

Let  $\{\tilde{y}_t\}_{t=1}^n$  be the sequence of observations in time  $t$ . Now, let  $\tilde{x}_t$  be the  $p$ -lagged time series of  $\tilde{y}_t$ , such that  $\tilde{x}_t = L^p(\tilde{y}_t)$ , where  $L$  is the lag operator. Matching each observation  $\tilde{y}_t$  with its  $p$ -lagged correspondent  $\tilde{x}_t$  will produce  $n - p$  pairs  $\{(\tilde{y}_t, \tilde{x}_t)\}_{t=p+1}^n$  (note that the first  $p$  observations of  $y_t$  must be discarded). When we order the observation of  $x$  in such way that they are in growing order

$$\tilde{x}^{(p+1)} \leq \tilde{x}^{(p+2)} \leq \dots \leq \tilde{x}^{(n)},$$

we can then define  $\{x_i\}_{i=1}^{n-p} = \{\tilde{x}^{(t)}\}_{t=p+1}^n$  and  $\{y_i\}_{i=1}^{n-p} = \{\tilde{y}^{(t)}\}_{t=p+1}^n$  and  $T = \{2, \dots, n - p - 1\}$ . As we need the second difference of  $q_i$ ,  $I$  has to be shortened by two elements.

# The optimization problem

Our optimization model to estimate the nonparametric quantile is as follows:

$$\mathcal{Q}_{y_t|y_{t-1}}^\alpha(t) = \arg \min_{q_t} \sum_{t \in T} \left( |y_t - q_t|^+ \alpha + |y_t - q_t|^- (1 - \alpha) \right) \\ + \lambda \sum_{t \in T} |D_{x_t}^2 q_t|,$$

where  $D^2 q_t$  is the second derivative of the  $q_t$  function, calculated as follows:

$$D_{x_t}^2 q_t = \left( \frac{q_{t+1} - q_t}{x_{t+1} - x_t} \right) - \left( \frac{q_t - q_{t-1}}{x_t - x_{t-1}} \right).$$

# The optimization problem

The full model can be rewritten as a LP problem as bellow:

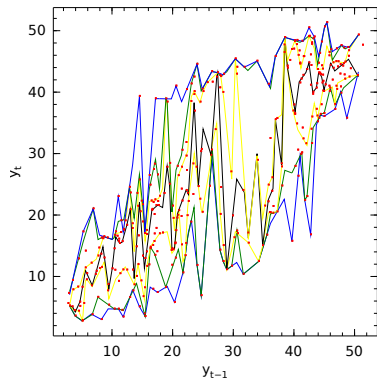
$$\begin{aligned} \min_{q_t} \quad & \sum_{t=1}^n \left( \alpha \delta_t^+ + (1 - \alpha) \delta_t^- \right) + \lambda \sum_{t=1}^n \xi_t \\ \text{s.t.} \quad & \delta_t^+ - \delta_t^- = y_t - q_t, & \forall t \in \{3, \dots, n-1\}, \\ & D_t = \left( \frac{q_{t+1} - q_t}{x_{t+1} - x_t} \right) - \left( \frac{q_t - q_{t-1}}{x_t - x_{t-1}} \right), & \forall t \in \{3, \dots, n-1\}, \\ & \xi_t \geq D_t, & \forall t \in \{3, \dots, n-1\}, \\ & \xi_t \geq -D_t, & \forall t \in \{3, \dots, n-1\}, \\ & \delta_t^+, \delta_t^-, \xi_t \geq 0, & \forall t \in \{3, \dots, n-1\}. \end{aligned}$$



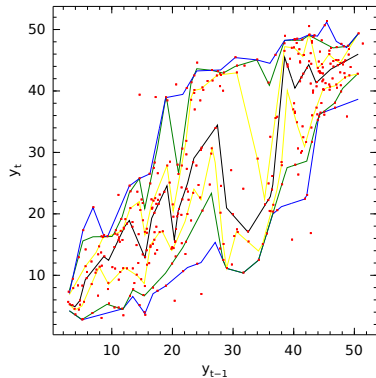
# Tuning

- ▶ The output of our optimization problem is a sequence of ordered points  $\{(x_t, q_t)\}_{t \in T}$ . The next step is to interpolate these points in order to provide an estimation for any other value of  $x$ . To address this issue, we propose using a B-splines interpolation, that will be developed in another study.
- ▶ The quantile estimation is done for different values of  $\lambda$ . By using different levels of penalization on the second difference, the estimation can be more or less adaptive to the fluctuation. It is important to notice that the usage of the  $\ell_1$ -norm as penalty leads to a piecewise linear solution  $q_t$ .
- ▶ Quantile estimation for a few different values of  $\lambda$  are shown on the next slides.

## Quantile estimation for different values of $\lambda$

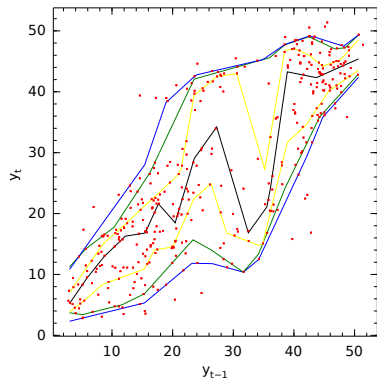


Overfitting with  $\lambda = 0.1$

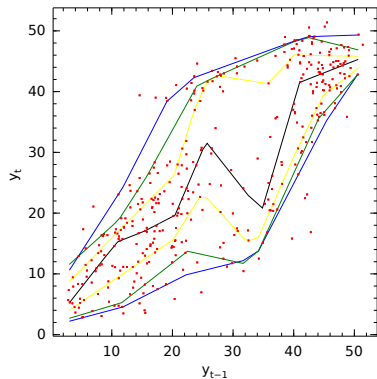


$\lambda = 0.3$

## Quantile estimation for different values of $\lambda$

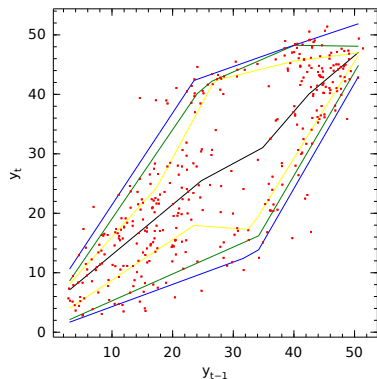


$\lambda = 1$

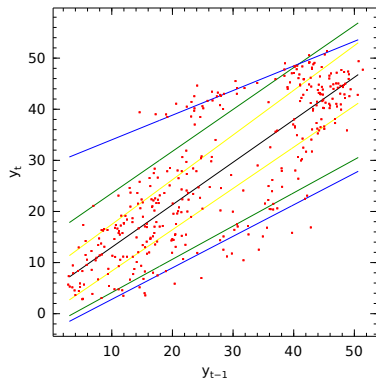


$\lambda = 3$

## Quantile estimation for different values of $\lambda$



$\lambda = 10$



Degenerated case with  $\lambda = 200$

## Open issues about the nonparametric quantile autoregression

- ▶ The first issue is **how to select an appropriate value for  $\lambda$** . A simple way is to do it by inspection, which means to test many different values and pick the one that suits best our needs by looking at them. The other alternative is to use a metric to which we can select the best tune. We can achieve this by using a cross-validation method, for example.
- ▶ The other issue occurs when **we try to add more than one lag to the analysis at the same time**. This happens because the problem solution is a set of points that we need to interpolate. This multivariate interpolation, however, is not easily solved, in the sense that we can either choose using a very naive estimator such as the K-nearest neighbors or just find another method that is not yet adopted for a wide range of applications.

Next steps

## Testing methods with high-frequency data

- ▶ All methods already discussed applied to high frequency data on different locations

# Local Quantile Regression

- Based on Bremmes (2004). Being the estimation of the  $\theta$  quantile  $q_\theta()$  :

$$q_\theta(x; \alpha_0, \alpha) = \alpha_0 + \alpha^T x$$

we define the local quantile regression by solving the following optimization problem:

$$\operatorname{argmin}_{(\alpha_0, \alpha)} \sum_{i=1}^n \rho_\theta(e_{\rho,i} - q_\theta(x_i - x; \alpha_0, \alpha)) w \left( \frac{\|x_i - x\|_2}{h_\lambda(x)} \right)$$

where the loss function is defined by

$$\rho_\theta(u) = \begin{cases} u\theta & \text{if } u \geq 0 \\ u(\theta - 1) & \text{otherwise} \end{cases}$$

and the weight function

$$w(u) = \begin{cases} (1 - u^3)^3 & \text{if } u \in [0, 1) \\ 0 & \text{otherwise} \end{cases}$$



# Generating Scenarios of Wind Power Prediction

- ▶ Following recommendation of Pinson and Madsen (2009). This method is employed for short term monte carlon simulation.
- ▶ 1. One must have a different model for each horizon  $k \in \{1, \dots, K\}$ . Residuals  $\mathbf{X}$  come from a multivariate Gaussian distribution  $\mathbf{X} \sim N(\mu_0, \Sigma_{t-K})$ . As we want to generate  $d$  different scenarios, we have to take  $d$  samples.
- ▶ 2. By applying the inverse probit function  $\Phi$  to each component of  $\mathbf{X}^{(i)}$ , we obtain the random variable  $Y_k^{(i)} = \Phi(\mathbf{X}_k^{(i)})$ ,  $\forall k, i$ .
- ▶ 3. The predicted value of for time  $t + k$  for the  $i$ th scenario is given by:

$$\hat{p}_{t+k|t}^{(i)} = \hat{F}_{t+k|t}^{-1}(Y_k^{(i)}), \quad \forall k, i,$$

where  $\hat{F}$  is the distribution function.

- ▶ Quantile regression is used to approximate distribution function  $F$  and using splines.

## Multiple states defining multiple models

- ▶ Explore the possibility of having different models for different states of climate conditions.
- ▶ One approach is using unsupervised classification techniques, such as support vector machines