

# Scenario generation for nongaussian time series via Quantile Regression

Marcelo Ruas e Alexandre Street

November 23, 2017

# Motivation

- Renewable energy scenarios are important in many fields in Power Systems:
  - ① Energy trading;
  - ② unit commitment;
  - ③ grid expansion planning;
  - ④ investment decisions
- In stochastic optimization problems, a set of scenarios is a needed input.
- Robust optimization requires bounds for probable values.

**Change in paradigm: from predicting the conditional mean to predicting the conditional distribution**

# Probability Forecasting Approaches

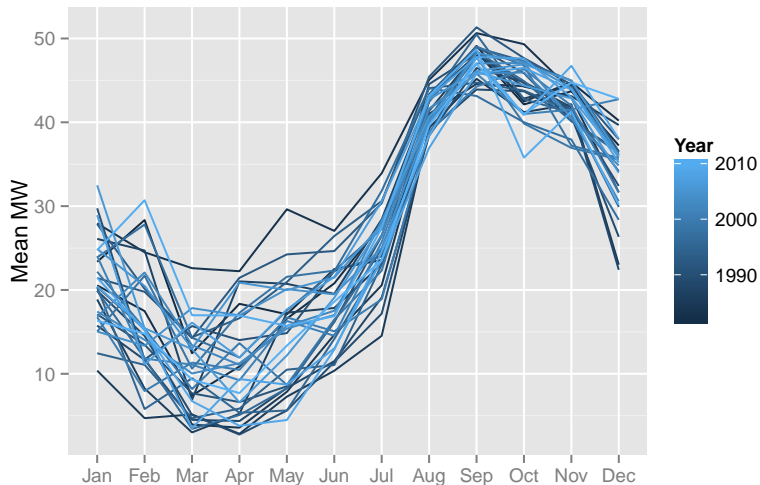
- *Parametric Models*

- Assume a distributional shape
- Low computational costs
- Faster convergence
- *Examples: Arima-GARCH, GAS*

- *Nonparametric Models*

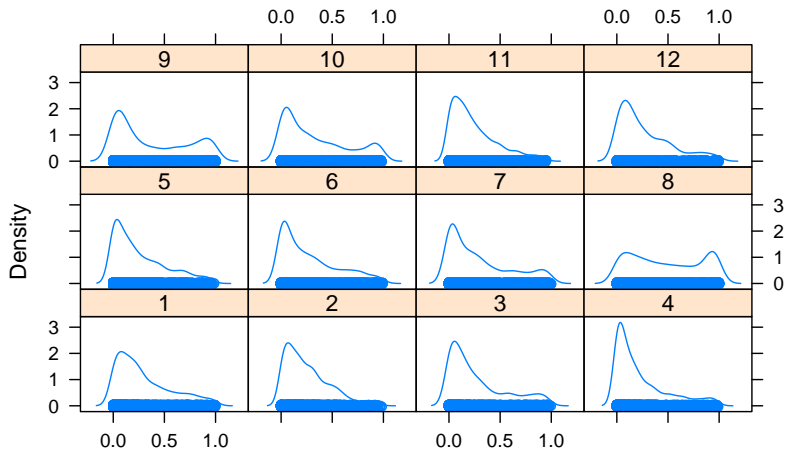
- Don't require a distribution to be specified
- High computational cost
- Needs more data to produce a good approximation
- *Examples: Quantile Regression (Koenker and Bassett Jr (1978)), Kernel Density Estimation (Gallego-Castillo et al. (2016)), Artificial Intelligence (Wan et al. (2017))*

# Wind Power Time Series - Icaraizinho monthly data



# Wind Power Time Series - Kaggle forecasting competition hourly data

## Wind power density comparison across different months



# The nongaussianity of Wind Power

- Renewables, such as wind and solar power have reportedly nongaussian behaviour
- Convenience of using a nonparametric approach, which doesn't rely on assuming a distribution
- Quantile regression is the chosen technique available to model this time series dynamics, by estimating a thin grid of  $\alpha$ -quantiles at once and forming a data-driven conditional distribution

# Objectives

- A nonparametric methodology to model the conditional distribution of renewables time series to produce scenarios.
- We propose a methodology that selects the global optimal solution with parsimony both on the selection of covariates as on the quantiles. Regularization methods are based on two techniques: Best Subset Selection (MILP) and LASSO (Linear Programming)
- Regularization techniques applied to an ensemble of quantile functions to estimate the conditional distribution, solving the issue of non-crossing quantiles. On regularizing quantiles, we propose a smoothness on the coefficients values across the sequence of quantiles.

# Definition of the Conditional Quantile

Let the conditional quantile function of  $Y$  for a given value  $x$  of the  $d$ -dimensional random variable  $X$ , i.e.,  $Q_{Y|X} : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$ , can be defined as:

$$Q_{Y|X}(\alpha, x) = F_{Y|X=x}^{-1}(\alpha) = \inf\{y : F_{Y|X=x}(y) \geq \alpha\}.$$



# Conditional Quantile from a sample

Let a dataset be composed from  $\{y_t, x_t\}_{t \in T}$  and let  $\rho$  be the check function

$$\rho_\alpha(x) = \begin{cases} \alpha x & \text{if } x \geq 0 \\ (1 - \alpha)x & \text{if } x < 0 \end{cases}, \quad (1)$$

The sample quantile function for a given probability  $\alpha$  is then based on a finite number of observations and is the solution to minimizing the loss function  $L(\cdot)$ :

$$\hat{Q}_{Y|X}(\alpha, \cdot) \in \arg \min_{q(\cdot) \in \mathcal{Q}} L_\alpha(q) = \sum_{t \in T} \rho_\alpha(y_t - q(x_t)),$$

$$q(x_t) = \beta_0 + \beta^T x_t,$$

where  $\mathcal{Q}$  is a space of functions. In this paper, we use  $\mathcal{Q}$  as an **affine functions space**.

# Conditional Quantile from a sample

- For a single quantile, this problem can be solved by the following Linear Programming problem:

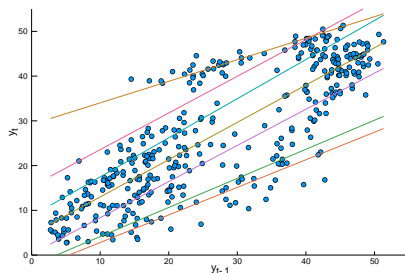
$$\begin{array}{ll}
 \min_{\beta_0, \beta, \varepsilon_t^+, \varepsilon_t^-} & \sum_{t \in T} (\alpha \varepsilon_t^+ + (1 - \alpha) \varepsilon_t^-) \\
 \text{s.t.} & \varepsilon_t^+ - \varepsilon_t^- = y_t - \beta_0 - \beta^T x_t, \quad \forall t \in T, \\
 & \varepsilon_t^+, \varepsilon_t^- \geq 0, \quad \forall t \in T.
 \end{array}$$

- The output are the coefficients  $\beta_0$  and  $\beta$  (which is the same dimension as  $x_t$ ), that describe the quantile function as an affine function.

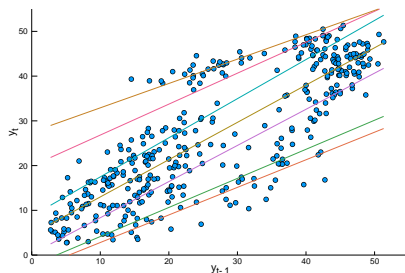
# The non-crossing issue

- The following condition must always hold:

$$q_{\alpha}(x_t) \leq q_{\alpha'}(x_t), \text{ when } \alpha \leq \alpha'$$



(a) Each  $\alpha$ -quantile estimated independently



(b) Estimation with non-crossing constraint

**Figure:** These graphs show how the addition of a constraint can contour the crossing quantile issue

# Notation

Expression	Meaning
$Q_{Y X}(\alpha, x)$	The conditional quantile function
$y_t$	the time series we are modelling
$x_t$	explanatory variables of $y_t$ in $t$
$T$	the set containing all observations indexes
$J$	the set containing all quantile indexes
$J_{(-1)}$	the set $J \setminus \{1\}$
$\alpha_j$	a probability, might be indexed by $j$
$A$	the set of probabilities $\{\alpha_j \mid j \in J\}$
$K$	Maximum number of covariates on MILP regularization
$\lambda$	The Lasso penalization on the coefficients $\ell_1$ -norm
$\gamma$	The penalization on the coefficients second-derivative with respect of the quantiles

# Conditional Quantile as a Linear Programming Problem

$$\min_{\beta_{0j}, \beta_j, \varepsilon_{tj}^+, \varepsilon_{tj}^-} \sum_{j \in J} \sum_{t \in T} \left( \alpha_j \varepsilon_{tj}^+ + (1 - \alpha_j) \varepsilon_{tj}^- \right)$$

s.t.

$$\varepsilon_{tj}^+ - \varepsilon_{tj}^- = y_t - \beta_{0j} - \beta_j^T x_t, \quad \forall t \in T, \forall j \in J,$$

$$\varepsilon_{tj}^+, \varepsilon_{tj}^- \geq 0, \quad \forall t \in T, \forall j \in J,$$

$$\beta_{0,j-1} + \beta_{j-1}^T x_t \leq \beta_{0j} + \beta_j^T x_t, \quad \forall t \in T, \forall j \in J_{(-1)},$$

- Coefficients  $\beta_{0j}$  and  $\beta_j$  refer to the  $j^{\text{th}}$  quantile
- We apply QR to estimate the conditional distribution  $\hat{Q}_{Y_{t+h}|X_{t+h}, Y_t, Y_{t-1}, \dots}(\alpha, \cdot)$  for a  $k$ -step ahead forecast of time serie  $\{y_t\}$ , where  $X_{t+h}$  is a vector of exogenous variables at the time we want to forecast.

# Best Subset selection via MILP

- Mixed Integer Linear Programming (MILP) models allow only  $K$  variables to be used for each  $\alpha$ -quantile.
- Only  $K$  coefficients  $\beta_{pj}$  may have nonzero values, for each  $\alpha$ -quantile.
- It is guaranteed by constraints on the optimization model.
- One model for each  $\alpha$ -quantile

# Best Subset selection via MILP

$$\begin{aligned}
 & \min_{\beta_{0j}, \beta_j, z_{pj}, \varepsilon_{tj}^+, \varepsilon_{tj}^-} \sum_{j \in J} \sum_{t \in T} \left( \alpha_j \varepsilon_{tj}^+ + (1 - \alpha_j) \varepsilon_{tj}^- \right) \\
 \text{s.t.} \quad & \varepsilon_{tj}^+ - \varepsilon_{tj}^- = y_t - \beta_{0j} - \beta_j^T x_t, & \forall t \in T, \forall j \in J, \\
 & \varepsilon_{tj}^+, \varepsilon_{tj}^- \geq 0, & \forall t \in T, \forall j \in J, \\
 & -Mz_{pj} \leq \beta_{pj} \leq Mz_{pj}, & \forall j \in J, \forall p \in P, \\
 & \sum_{p \in P} z_{pj} \leq K, & \forall j \in J, \\
 & z_{pj} \in \{0, 1\}, & \forall j \in J, \forall p \in P, \\
 & \beta_{0,j-1} + \beta_{j-1}^T x_t \leq \beta_{0j} + \beta_j^T x_t, & \forall t \in T, \forall j \in J_{(-1)},
 \end{aligned}$$

- $z_{pj}$  is a binary variable which indicates when  $\beta_{pj} > 0$ .

# Variable Selection via LASSO

- Regularization by including the coefficients  $\ell_1$ -norm on the objective function.
- In this method, coefficients are shrunk towards zero by changing a continuous parameter  $\lambda$ , which penalizes the size of the  $\ell_1$ -norm.
- When the value of  $\lambda$  gets bigger, fewer variables are selected to be used.
- The optimization problem for a single quantile is presented below:

$$\min_{\beta_0, \beta} \sum_{t \in T} \rho_\alpha(y_t - (\beta_0 + \beta^T x_t)) + \lambda \|\beta\|_1,$$



# Variable Selection via LASSO

- At first, we select variables using LASSO

$$\arg \min_{\beta_0, \beta, \varepsilon_{tj}^+, \varepsilon_{tj}^-} \sum_{j \in J} \sum_{t \in T} \left( \alpha_j \varepsilon_{tj}^+ + (1 - \alpha_j) \varepsilon_{tj}^- \right) + \lambda \sum_{p \in P} \xi_{pj}$$

subject to

$$\begin{aligned} \varepsilon_{tj}^+ - \varepsilon_{tj}^- &= y_t - \beta_0 j - \beta_j^T x_t, & \forall t \in T, \forall j \in J, \\ \varepsilon_{tj}^+, \varepsilon_{tj}^- &\geq 0, & \forall t \in T, \forall j \in J, \\ \xi_{pj} &\geq \beta_{pj}, & \forall p \in P, \forall j \in J, \\ \xi_{pj} &\geq -\beta_{pj}, & \forall p \in P, \forall j \in J, \\ \beta_{0,j-1} + \beta_{j-1}^T x_t &\leq \beta_{0j} + \beta_j^T x_t, & \forall t \in T, \forall j \in J_{(-1)}, \end{aligned}$$

# Variable Selection via LASSO

- We then define  $S_\lambda$  as the set of indexes of selected variables given by

$$S_\lambda = \{p \in \{1, \dots, P\} \mid |\beta_{\lambda,p}^{*LASSO}| \neq 0\}.$$

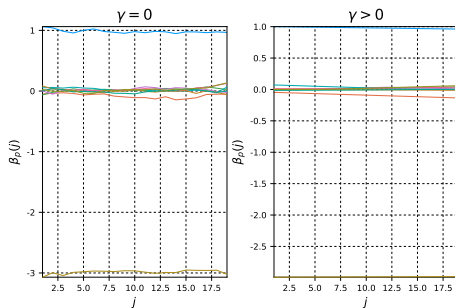
Hence, we have that, for each  $p \in \{1, \dots, P\}$ ,

$$\beta_{\theta,p}^{*LASSO} = 0 \implies \beta_{\theta,p}^* = 0.$$

- On the second stage, we estimate coefficients using a regular QR where input variables are only the ones which belonging to  $S_\lambda$

# Regularization on the quantiles - Motivation

In practice, often when we regularize



# MILP - Defining groups for $\alpha$ -quantiles

$$\begin{aligned}
 & \min_{\beta_{0j}, \beta_j, z_{pg}, \varepsilon_{tj}^+, \varepsilon_{tj}^-} \sum_{j \in J} \sum_{t \in T} \left( \alpha_j \varepsilon_{tj}^+ + (1 - \alpha_j) \varepsilon_{tj}^- \right) \\
 \text{s.t} \quad & \varepsilon_{tj}^+ - \varepsilon_{tj}^- = y_t - \beta_{0j} - \beta_j^T x_t, \quad \forall t \in T, \forall j \in J, \\
 & \varepsilon_{tj}^+, \varepsilon_{tj}^- \geq 0, \quad \forall t \in T, \forall j \in J, \\
 & \mathcal{Z}_{pg} := 2 - (1 - z_{pg}) - l_{gj}, \\
 & -M \mathcal{Z}_{pg} \leq \beta_{pj} \leq M \mathcal{Z}_{pg}, \quad \forall j \in J, \forall p \in P, \forall g \in G \\
 & \sum_{p \in P} z_{pg} \leq K, \quad \forall j \in J, \\
 & \beta_{0,j-1} + \beta_{j-1}^T x_t \leq \beta_{0j} + \beta_j^T x_t, \forall t \in T, \forall j \in J_{(-1)}, \\
 & l_{gj}, z_{pg} \in \{0, 1\}, \quad \forall p \in P, \forall g \in G, \\
 & z_{pg} \in \{0, 1\}, \quad \forall j \in J, \forall p \in P,
 \end{aligned}$$

# MILP with quantile regularization

$$\min_{\beta_{0j}, \beta_j, z_{pj}, \varepsilon_{tj}^+, \varepsilon_{tj}^-} \sum_{j \in J} \sum_{t \in T} \left( \alpha_j \varepsilon_{tj}^+ + (1 - \alpha_j) \varepsilon_{tj}^- \right) + \gamma \sum_{j \in J'} (D2_{pj}^+ + D2_{pj}^-)$$

subject to

$$\varepsilon_{tj}^+ - \varepsilon_{tj}^- = y_t - \beta_{0j} - \beta_j^T x_{t,p}, \quad \forall t \in T, \forall j \in J,$$

$$-Mz_{p\alpha} \leq \beta_{pj} \leq Mz_{p\alpha}, \quad \forall j \in J, \forall p \in P,$$

$$\sum_{p \in P} z_{p\alpha} \leq K, \quad \forall j \in J,$$

$$D2_{pj}^+ - D2_{pj}^- = \frac{\left( \frac{\beta_{p,j+1} - \beta_{pj}}{\alpha_{j+1} - \alpha_j} \right) - \left( \frac{\beta_{p,j} - \beta_{p,j-1}}{\alpha_j - \alpha_{j-1}} \right)}{\alpha_{j+1} - 2\alpha_j + \alpha_{j-1}},$$

$$\forall j \in J_{(-1)}, \forall p \in P,$$

$$\beta_{0j} + \beta_j^T x_t \leq \beta_{0,j+1} + \beta_{j+1}^T x_t, \quad \forall t \in T, \forall j \in J_{(-1)},$$

$$z_{p\alpha} \in \{0, 1\}, \quad \forall j \in J, \forall p \in P,$$

# LASSO with quantile regularization

$$\begin{aligned} \tilde{\beta}_{\lambda}^{*LASSO} = \arg \min_{\beta_0, \beta, \varepsilon_{tj}^+, \varepsilon_{tj}^-} & \sum_{j \in J} \sum_{t \in T} (\alpha_j \varepsilon_{tj}^+ + (1 - \alpha_j) \varepsilon_{tj}^-) \\ & + \lambda \sum_{p \in P} \xi_{pj} + \gamma \sum_{j \in J'} (D2_{pj}^+ + D2_{pj}^-) \end{aligned}$$

subject to

$$\varepsilon_{tj}^+ - \varepsilon_{tj}^- = y_t - \beta_{0j} - \beta_j^T x_{t,p}, \quad \forall t \in T, \forall j \in J,$$

$$\xi_{pj} \geq \beta_{pj}, \quad \forall p \in P, \forall j \in J,$$

$$\xi_{pj} \geq -\beta_{pj}, \quad \forall p \in P, \forall j \in J,$$

$$D2_{pj}^+ - D2_{pj}^- = \frac{\left( \frac{\beta_{p,j+1} - \beta_{pj}}{\alpha_{j+1} - \alpha_j} \right) - \left( \frac{\beta_{p,j} - \beta_{p,j-1}}{\alpha_j - \alpha_{j-1}} \right)}{\alpha_{j+1} - 2\alpha_j + \alpha_{j-1}},$$

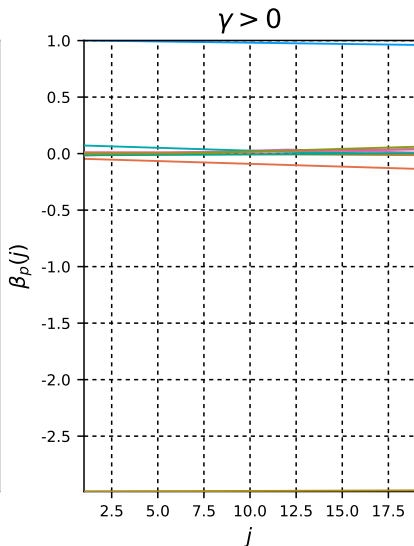
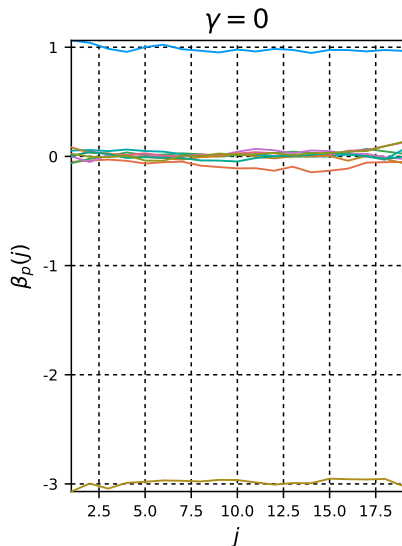
$$\forall j \in J_{(-1)}, \forall p \in P,$$

$$\beta_{0j} + \beta_j^T x_t \leq \beta_{0,j+1} + \beta_{j+1}^T x_t, \quad \forall t \in T, \forall j \in J_{(-1)},$$

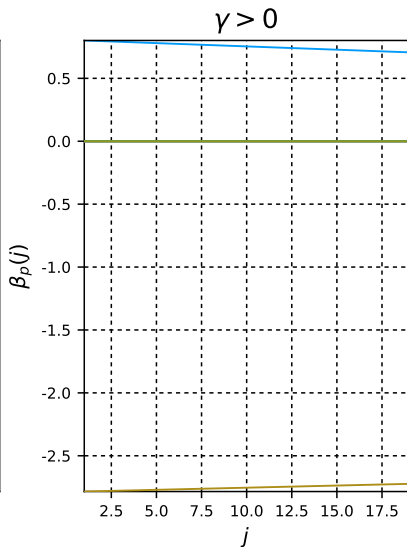
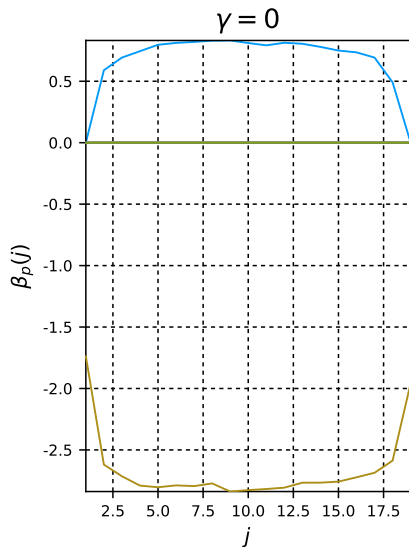
$$\varepsilon_{tj}^+, \varepsilon_{tj}^- \geq 0, \quad \forall t \in T, \forall j \in J,$$

$$D2_{pj}^+, D2_{pj}^- \geq 0, \quad \forall j \in J, \forall p \in P.$$

# Regularization on the quantiles



# LASSO - Penalization of derivative





# LASSO - Penalization of derivative

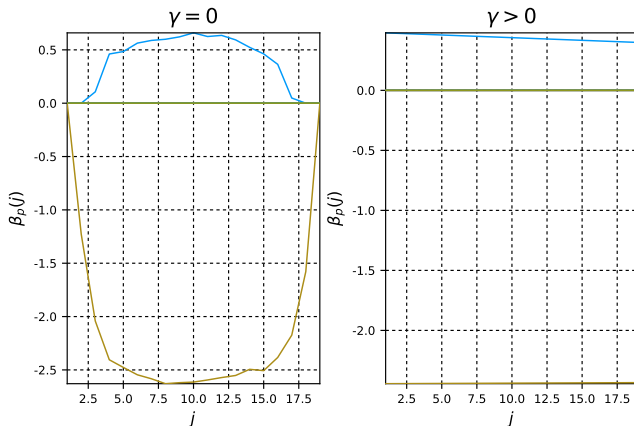


Figure: Testing caption

# ADALASSO

LASSO solutions are solutions that minimize

$$Q(\beta|X, y) = \frac{1}{2n} \|y - X\beta\|^2 + \lambda \sum_{p \in P} |\beta_p|.$$

The adaptive lasso adds weights to try to correct the known issue of bias on the LASSO estimation. The ADALASSO estimation is the second step LASSO given by:

$$Q_a(\beta|X, y, w) = \frac{1}{2n} \|y - X\beta\|^2 + \lambda \sum_{p \in P} w_p |\beta_p|,$$

where  $w_p$  weights the coefficients differently according to their first step value. Normally, we have

$$w_p(\lambda) = w(\tilde{\beta}_{p,t}(\lambda)).$$

# ADALASSO for quantile regression

When estimating the ADALASSO for quantile regression, we show a few adaptations and extensions of the original method. The full process consists of two steps, each consisting of a LASSO estimation:

- **First step:** First LASSO regularization

$$\min_{\beta_{0j}, \beta_j} \sum_{j \in J} \left( \sum_{t \in T} \rho_{\alpha_j}(y_t - (\beta_{0j} + \beta_j^T x_t)) + \lambda \sum_{p \in P} |\beta_{pj}| \right) + \gamma \sum_{j \in J'} (D2_{pj}^+ + D2_{pj}^-),$$

- **Second step:** Two forms of using initial estimation to determine  $w_{pj}$  are:

- 1  $w_{pj} = 1/\beta_{pj}$ .
- 2  $w_{pj} = 1/(\beta_{pj} \parallel \beta_j \parallel_1)$ ,

The weights  $w_j$  are input to a second-stage Lasso estimation:

$$\min_{\beta_{0j}, \beta_j} \sum_{j \in J} \left( \sum_{t \in T} \rho_{\alpha_j}(y_t - (\beta_{0j} + \beta_j^T x_t)) + \lambda \sum_{p \in P} w_{pj}^\delta |\beta_{pj}| \right) + \gamma \sum_{j \in J'} (D2_{pj}^+ + D2_{pj}^-)$$

where  $\delta$  is an exponential parameter, normally set to 1.

# Evaluation Metrics

- We use a performance measurement which emphasizes the correctness of each quantile. For each probability  $\alpha \in A$ , a loss function is defined by

$$L_{\alpha}(q) = \sum_{t \in T} \rho_{\alpha}(y_t - q_{\alpha}(x_t)).$$

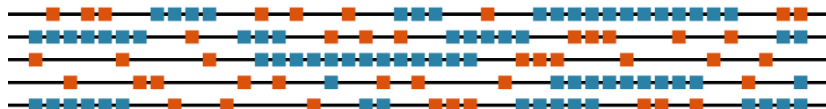
The loss score  $\mathcal{L}$ , which is the chosen evaluation metric to optimize, aggregates the score function over all elements of  $A$ :

$$\mathcal{L} = \frac{1}{|A|} \sum_{\alpha \in A} L_{\alpha}(q).$$

# Time-series Cross-Validation



5-fold cross-validation



5-fold non-dep. cross-validation

**Figure:**  $\mathcal{K}$ -fold CV and  $\mathcal{K}$ -fold with non-dependent data. Observations in blue are used to estimation and in orange for evaluation. Note that non-dependent data doesn't use all dataset in each fold.

# Time-series Cross-Validation

- The CV score is given by the sum of the loss function for each fold. The optimum value of  $t$  in this criteria is the one that minimizes the CV score:

$$\theta^* = \operatorname{argmin}_{\theta} CV(\theta) = \sum_{k \in \mathcal{K}} \sum_{\alpha \in A} L_{\alpha}(q).$$

- To optimize CV function in  $\theta$ , we use the Nelder-Mead algorithm, which is a known and widely used algorithm for black-box optimization.

# Nonparametric model

$$\hat{Q}_{Y|X}(\alpha, \cdot) \in \arg \min_{q(\cdot) \in \mathcal{Q}} L_{\alpha}(q) = \sum_{t \in T} \rho_{\alpha}(y_t - q(x_t)),$$

- On nonparametric models,  $q_{\alpha}$  belongs to a space of limited second derivative function  $\mathcal{Q}$ .
- The  $\alpha$ -quantile function is flexible enough to capture nonlinearities on the quantile function.

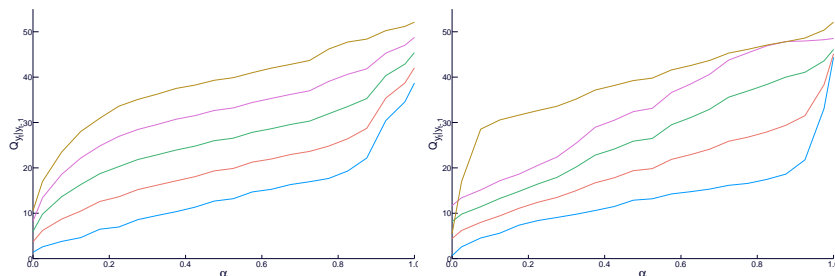
# Nonparametric model - Formulation

$$\begin{aligned}
 & \min_{q_{jt}, \varepsilon_t^+, \varepsilon_t^-, \xi_t} \sum_{j \in J} \sum_{t \in T'} \left( \alpha_j \varepsilon_{tj}^+ + (1 - \alpha_j) \varepsilon_{tj}^- \right) + \lambda \sum_{t \in T'} \xi_{tj} \\
 & \text{s.t.} \quad \varepsilon_t^+ - \varepsilon_{tj}^- = y_t - q_{tj}, \quad \forall t \in T', \forall j \in J, \\
 & \quad D_{tj}^1 = \frac{q_{jt+1} - q_{jt}}{x_{t+1} - x_t}, \quad \forall t \in T', \forall j \in J, \\
 & \quad D_{tj}^2 := \frac{\left( \frac{q_{jt+1} - q_{jt}}{x_{t+1} - x_t} \right) - \left( \frac{q_{jt} - q_{jt-1}}{x_t - x_{t-1}} \right)}{x_{t+1} - 2x_t + x_{t-1}} \\
 & \quad \xi_{tj} \geq D_{tj}^2, \quad \forall t \in T', \forall j \in J, \\
 & \quad \xi_{tj} \geq -D_{tj}^2, \quad \forall t \in T', \forall j \in J, \\
 & \quad \varepsilon_{tj}^+, \varepsilon_{tj}^-, \xi_{tj} \geq 0, \quad \forall t \in T', \forall j \in J, \\
 & \quad q_{tj} \leq q_{t,j+1}, \quad \forall t \in T', \forall j \in J,
 \end{aligned}$$



# Nonparametric vs. Linear Model

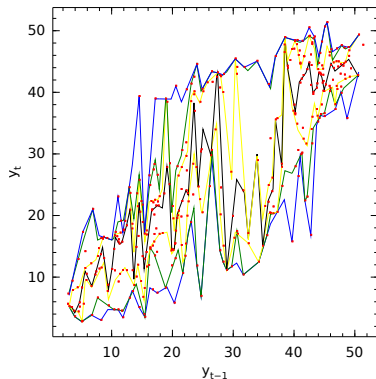
- The nonparametric approach is more flexible to capture heteroscedasticity.



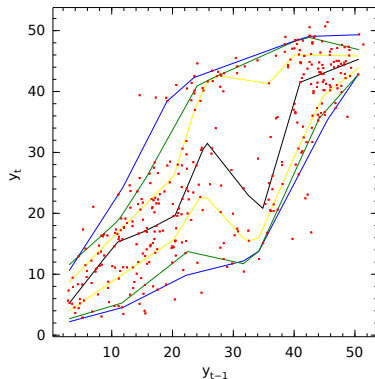
**Figure:** Estimated quantile functions, for different values of  $y_{t-1}$ . On the left using a linear model and using a nonparametric approach on the right.

# Control of smoothing parameter

- This flexibility might lead to overfitting, if we don't select a proper smoothing parameter.

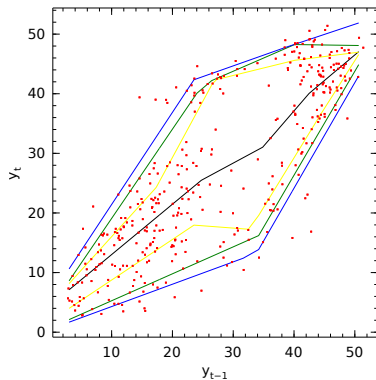


(a)  $\lambda = 0.1$

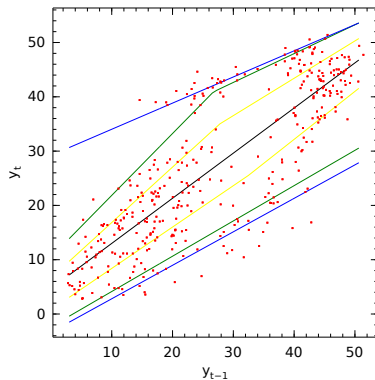


(b)  $\lambda = 3$

# Control of smoothing parameter



(a)  $\lambda = 10$



(b)  $\lambda = 100$

- On the limit, when  $\lambda \rightarrow \infty$ , the nonparametric model approaches a linear model.

# Present issues

- Difficult interpolation when  $x_t$  has dimension greater than 1.
- Control of smoothing parameter

# References

- Gallego-Castillo, Cristobal, Ricardo Bessa, Laura Cavalcante, and Oscar Lopez-Garcia. 2016. "On-Line Quantile Regression in the Rkhs (Reproducing Kernel Hilbert Space) for Operational Probabilistic Forecasting of Wind Power." *Energy* 113. Elsevier: 355–65.
- Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression Quantiles." *Econometrica: Journal of the Econometric Society*. JSTOR, 33–50.
- Wan, C., J. Lin, J. Wang, Y. Song, and Z. Y. Dong. 2017. "Direct Quantile Regression for Nonparametric Probabilistic Forecasting of Wind Power Generation." *IEEE Transactions on Power Systems* 32 (4): 2767–78. doi:10.1109/TPWRS.2016.2625101.