

Scenario generation for nongaussian time series via Quantile Regression

Marcelo Ruas and Alexandre Street

OCt 05, 2017

Introduction

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, Kernel Density Estimation, Artificial Intelligence*

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 α -quantiles at once and forming a data-driven conditional distribution

Quantile Regression

Definition of the Conditional Quantile

Let the α -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:

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

Conditional Quantile from a sample

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

$$\rho_{\alpha}(x) = \begin{cases} \alpha x & \text{if } x \geq 0 \\ (1 - \alpha)x & \text{if } x < 0 \end{cases}, \quad (2)$$

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

$$\hat{Q}_{Y|X}(\alpha, \cdot) \in \arg \min_{q \in \mathcal{Q}} L_{\alpha}(q) = \sum_{t \in T} \rho_{\alpha}(y_t - q(x_t)). \quad (3)$$

Conditional Quantile as a Linear Programming Problem

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} (\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^-)$$

subject to

$$\varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \beta_{\alpha}^T x_t, \quad \forall t \in T, \forall \alpha \in A,$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A,$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t,$$

$$\forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha',$$

- We apply QR to estimate the conditional distribution

$\hat{Q}_{Y_{t+k}|X_{t+k}, Y_t, Y_{t-1}, \dots}(\alpha, \cdot)$ for a k -step ahead forecast of time series $\{y_t\}$, where X_{t+k} is a vector of exogenous variables at the time we want to forecast.

Regularization

Best Subset selection via MILP

- ▶ Mixed Integer Linear Programming (MILP) models allow only K variables to be used for each α -quantile. This means that only K coefficients $\beta_{p\alpha}$ may have nonzero values, for each α -quantile. It must be guaranteed by constraints on the optimization problem.
- ▶ We present three forms of regularization using MILP

MILP - One model for each α -quantile

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, z_{p\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) \quad (4)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \sum_{p=1}^P \beta_{p\alpha} x_{t,p}, \quad \forall t \in T, \forall \alpha \in A, \quad (5)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (6)$$

$$-Mz_{p\alpha} \leq \beta_{p\alpha} \leq Mz_{p\alpha}, \quad \forall \alpha \in A, \forall p \in P, \quad (7)$$

$$\sum_{p=1}^P z_{p\alpha} \leq K, \quad \forall \alpha \in A, \quad (8)$$

$$z_{p\alpha} \in \{0, 1\}, \quad \forall \alpha \in A, \forall p \in P, \quad (9)$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t, \quad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \quad (10)$$

MILP - Defining groups for α -quantiles

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, z_{p\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) \quad (11)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \sum_{p=1}^P \beta_{p\alpha} x_{t,p}, \quad \forall t \in T, \forall \alpha \in A, \quad (12)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (13)$$

$$-Mz_{p\alpha g} \leq \beta_{p\alpha} \leq Mz_{p\alpha g}, \quad \forall \alpha \in A, \forall p \in P, \forall g \in G \quad (14)$$

$$z_{p\alpha g} := 2 - (1 - z_{pg}) - I_{g\alpha} \quad (15)$$

$$\sum_{p=1}^P z_{pg} \leq K, \quad \forall g \in G, \quad (16)$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t, \quad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \quad (17)$$

$$\sum_{g \in G} I_{g\alpha} = 1, \quad \forall \alpha \in A, \quad (18)$$

$$I_{g\alpha}, z_{pg} \in \{0, 1\}, \quad \forall p \in P, \quad \forall g \in G, \quad (19)$$

MILP - Penalization of derivative

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, z_{p\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) + \gamma \sum_{\alpha \in A'} D2_{p\alpha} \quad (20)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \sum_{p=1}^P \beta_{p\alpha} x_{t,p}, \quad \forall t \in T, \forall \alpha \in A, \quad (21)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (22)$$

$$-Mz_{p\alpha} \leq \beta_{p\alpha} \leq Mz_{p\alpha}, \quad \forall \alpha \in A, \forall p \in P, \quad (23)$$

$$\sum_{p=1}^P z_{p\alpha} \leq K, \quad \forall \alpha \in A, \quad (24)$$

$$z_{p\alpha} \in \{0, 1\}, \quad \forall \alpha \in A, \forall p \in P, \quad (25)$$

$$\tilde{D}_{p\alpha'}^2 = \frac{\left(\frac{\beta_{p\alpha''} - \beta_{p\alpha'}}{\alpha'' - \alpha'} \right) - \left(\frac{\beta_{p\alpha'} - \beta_{p\alpha}}{\alpha' - \alpha} \right)}{\alpha'' - 2\alpha' + \alpha} \quad (26)$$

$$D2_{p\alpha'} > \tilde{D}_{p\alpha'}^2 \quad \forall \alpha' \in A', \forall p \in P \quad (27)$$

$$D2_{p\alpha'} > -\tilde{D}_{p\alpha'}^2 \quad \forall \alpha' \in A', \forall p \in P \quad (28)$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t, \quad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \quad (29)$$

Variable Selection via LASSO

- ▶ Another way of doing regularization is including the coefficients ℓ_1 -norm on the objective function
- ▶ In this method, coefficients are shrunk towards zero by changing a continuous parameter λ , which penalizes the size of the ℓ_1 -norm.
- ▶ When the value of λ gets bigger, fewer variables are selected to be used.
- ▶ The optimization problem for a single quantile is presented below:

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

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

Variable Selection via LASSO

- At first, we select variables using LASSO

$$\arg \min_{\beta_0, \beta, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) + \lambda \sum_{p=1}^P \xi_{p\alpha} + \gamma \sum_{\alpha \in A'} D2_{p\alpha} \quad (30)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \sum_{p=1}^P \beta_{p\alpha} \tilde{x}_{t,p}, \quad \forall t \in T, \forall \alpha \in A, \quad (31)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (32)$$

$$\xi_{p\alpha} \geq \beta_{p\alpha}, \quad \forall p \in P, \forall \alpha \in A, \quad (33)$$

$$\tilde{D}_{p\alpha'}^2 = \frac{\left(\frac{\beta_{p\alpha''} - \beta_{p\alpha'}}{\alpha'' - \alpha'} \right) - \left(\frac{\beta_{p\alpha'} - \beta_{p\alpha}}{\alpha' - \alpha} \right)}{\alpha'' - 2\alpha' + \alpha} \quad (34)$$

$$D2_{p\alpha'} > \tilde{D}_{p\alpha'}^2 \quad \forall \alpha' \in A', \forall p \in P \quad (35)$$

$$D2_{p\alpha'} > -\tilde{D}_{p\alpha'}^2 \quad \forall \alpha' \in A', \forall p \in P \quad (36)$$

$$\xi_{p\alpha} \geq -\beta_{p\alpha}, \quad \forall p \in P, \forall \alpha \in A. \quad (37)$$

Variable Selection via LASSO

- ▶ We then define S_θ (where $\theta = [\lambda \quad \gamma]^T$) as the set of indexes of selected variables given by

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

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_λ

Estimation and Evaluation

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

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

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

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

Time-series Cross-Validation



5-fold cross-validation



5-fold non-dep. cross-validation

Figure 1: \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:

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

- ▶ To optimize CV function in θ , we use the Nelder-Mead algorithm, which is very efficient for searching in a two-dimension parametric space.

Nonparametric model

Nonparametric model - Formulation

$$\min_{q_{\alpha t}, \delta_t^+, \delta_t^-, \xi_t}$$

s.t.

$$\sum_{\alpha \in A} \sum_{t \in T'} \left(\alpha \delta_{t\alpha}^+ + (1 - \alpha) \delta_{t\alpha}^- \right)$$

$$+ \lambda_1 \sum_{t \in T'} \gamma_{t\alpha} + \lambda_2 \sum_{t \in T'} \xi_{t\alpha}$$

$$\delta_t^+ - \delta_{t\alpha}^- = y_t - q_{t\alpha},$$

$$D_{t\alpha}^1 = \frac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t},$$

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

$$\gamma_{t\alpha} \geq D_{t\alpha}^1,$$

$$\gamma_{t\alpha} \geq -D_{t\alpha}^1,$$

$$\xi_{t\alpha} \geq D_{t\alpha}^2,$$

$$\xi_{t\alpha} \geq -D_{t\alpha}^2,$$

$$\delta_{t\alpha}^+, \delta_{t\alpha}^-, \gamma_{t\alpha}, \xi_{t\alpha} \geq 0,$$

$$q_{t\alpha} \leq q_{t\alpha'},$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall (\alpha, \alpha') \in A \times A, \alpha <$$

Nonparametric vs. Linear Model

- ▶ The nonparametric approach is more flexible to capture heteroscedasticity.

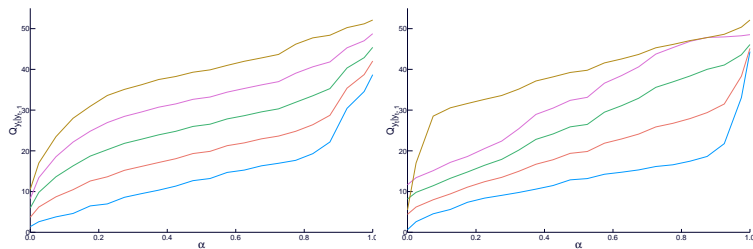


Figure 2: Estimated quantile functions, for different values of y_{t-1} . On the left using a linear model and using a nonparametric approach on the right.

Nonparametric vs. Linear Model

- This flexibility might lead to overfitting, if we don't select a proper penalty, as shown below:

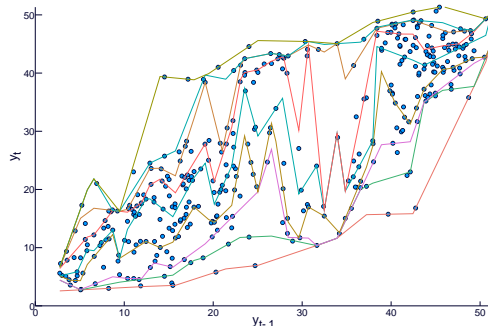


Figure 3: Example of a overfitted quantile function