Scenario generation for nongaussian time series via Quantile Regression

Marcelo Ruas and Alexandre Street

Oct 19, 2017

Introduction

Motivation

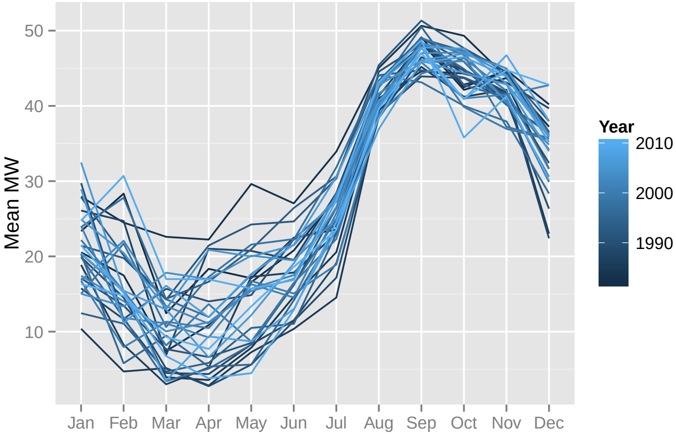
- Renewable energy scenarios are important in many fields in Power Systems:
  1. Energy trading;
  2. unit commitment;
  3. grid expansion planning;
  4. investment decisions
- In stochastic optimization problems, a set of scenarios is a needed input.
- Robust optimization requires bounds for probable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

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

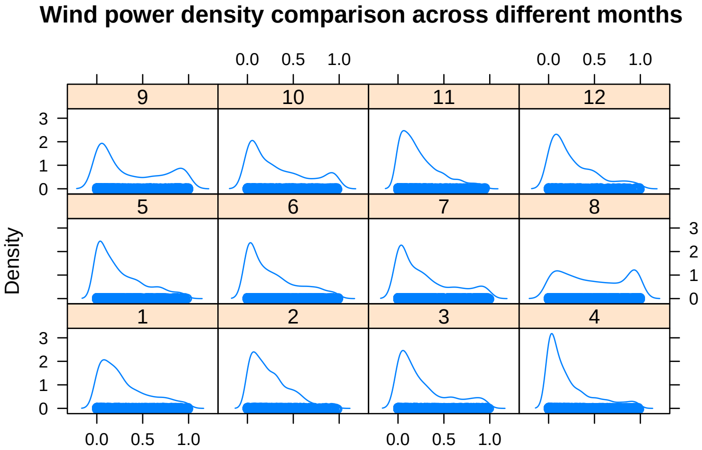
Probability Forecasting Approaches

- *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 (Koenker and Bassett Jr (1978)), Kernel Density Estimation (Gallego-Castillo et al. (2016)), Artificial Intelligence (Wan et al. (2017))*

Wind Power Time Series - Icaraizinho



The nongaussianity of Wind Power



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., $Q_{Y|X} : [0,1] \times \mathbb{R}^d \to \mathbb{R}$, can be defined as:

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

Conditional Quantile from a sample

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

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

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 $L(\cdot)$:

$$\hat{Q}_{Y|X}(\alpha, \cdot) \quad \in \quad \arg\min_{q(\cdot) \in \mathcal{Q}} L_\alpha(q) = \sum_{t \in T} \rho_\alpha(y_t - q(x_t)).,\tag{2}$$

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, the problem (2) can be solved by the following Linear Programming problem:

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

- The output are the coefficients β_0 and β (which is the same dimension as x_t), that describe describes the quantile function as an affine function.

The non-crossing issue

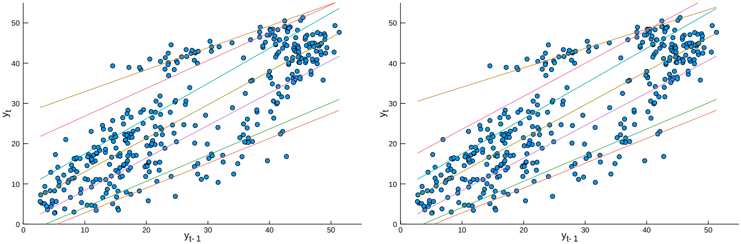


Figure 1: Estimated quantile functions, for different values of y_{t-1}. On the left using a linear model and using a nonparametric approach on the right.

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 $J \backslash \{1\}$
α_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
λ	The Lasso penalization on the coefficients ℓ_1-norm
γ	The penalization on the coefficients second-derivative with respect of the quantiles

Conditional Quantile as a Linear Programming Problem

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

$$\begin{aligned}
&\text{s.t.} \\
&\varepsilon^+_{tj} - \varepsilon^-_{tj} = y_t - \beta_{0j} - \beta_j^T x_t, && \forall t \in T, \forall j \in J, \\
&\varepsilon^+_{tj}, \varepsilon^-_{tj} \geq 0, && \forall t \in T, \forall j \in J, \\
&\beta_{0,j-1} + \beta_{j-1}^T x_t \leq \beta_{0j} + \beta_j^T x_t, && \forall t \in T, \forall j \in J_{(-1)},
\end{aligned}$$

We apply QR to estimate the conditional distribution $\hat{Q}_{Y_{t+h}|X_{t+h},Y_t,Y_{t-1},\ldots}(\alpha,\cdot)$ 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.

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 β_{pj} 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

$$
\begin{array}{lll}
\min\limits_{\beta_{0j},\beta_j,z_{pj}\varepsilon^+_{tj},\varepsilon^-_{tj}} & \sum_{j\in J}\sum_{t\in T}\left(\alpha_j\varepsilon^+_{tj} + (1-\alpha_j)\varepsilon^-_{tj}\right) & \\
\text{s.t} & \varepsilon^+_{tj} - \varepsilon^-_{tj} = y_t - \beta_{0j} - \sum_{p=1}^{P}\beta_{pj}x_{t,p}, & \forall t\in T, \forall j \in J, \\
& \varepsilon^+_{tj}, \varepsilon^-_{tj} \geq 0, & \forall t\in T, \forall j \in J, \\
& -Mz_{pj} \leq \beta_{pj} \leq Mz_{pj}, & \forall j\in J, \forall p \in P, \\
& \sum_{p=1}^{P} z_{pj} \leq K, & \forall j \in J, \\
& z_{pj} \in \{0,1\}, & \forall j\in J, \forall p \in P, \\
& \beta_{0,j-1} + \beta_{j-1}^T x_t \leq \beta_{0,j} + \beta_j^T x_t, & \forall t\in T, \forall j \in J_{(-1)},
\end{array}
$$

MILP - Defining groups for α-quantiles

$$
\begin{array}{lll}
\min\limits_{\beta_{0j},\beta_j,z_{pj},\varepsilon^+_{tj},\varepsilon^-_{tj}} & \sum_{j\in J}\sum_{t\in T}\left(\alpha_j\varepsilon^+_{tj}+(1-\alpha_j)\varepsilon^-_{tj}\right) & \\
\text{s.t} & \varepsilon^+_{tj}-\varepsilon^-_{tj}=y_t-\beta_{0j}-\beta_j^T x_{t,p}, & \forall t\in T,\forall j\in J, \\
& \varepsilon^+_{tj},\varepsilon^-_{tj}\geq 0, & \forall t\in T,\forall j\in J, \\
& -Mz_{pjg}\leq\beta_{pj}\leq Mz_{pjg}, & \forall j\in J,\forall p\in P, \\
& & \quad \forall g\in G \\
& z_{pjg}:=2-(1-z_{pg})-I_{gj} & \\
& \sum_{p=1}^{P} z_{pg}\leq K, & \forall j\in J, \\
& \beta_{0,j-1}+\beta_{j-1}^T x_t\leq\beta_{0j}+\beta_j^T x_t, & \forall t\in T,\forall j\in J_{(-1)}, \\
& I_{gj},z_{pg}\in\{0,1\}, & \forall p\in P,\forall g\in G, \\
& z_{pg}\in\{0,1\}, & \forall j\in J,\forall p\in P,
\end{array}
$$

MILP - Penalization of derivative

$$\min_{\beta_{0j},\beta_j,z_{pj}\varepsilon^+_{tj},\varepsilon^-_{tj}} \quad \sum_{j\in J}\sum_{t\in T}\left(\alpha_k\varepsilon^+_{tj} + (1-\alpha_k)\varepsilon^-_{t\alpha}\right) + \gamma\sum_{j\in J'} D2_{p\alpha} \tag{3}$$

$$\text{s.t} \quad \varepsilon^+_{tj} - \varepsilon^-_{tj} = y_t - \beta_{0j} - \sum_{p=1}^{P}\beta_{pj}x_{t,p}, \qquad \forall t\in T, \forall j\in J, \tag{4}$$

$$\varepsilon^+_{tj}, \varepsilon^-_{tj} \geq 0, \qquad \forall t\in T, \forall j\in J, \tag{5}$$

$$-Mz_{pj} \leq \beta_{pj} \leq Mz_{pj}, \qquad \forall j\in J, \forall p\in P, \tag{6}$$

$$\sum_{p=1}^{P} z_{pj} \leq K, \qquad \forall j\in J, \tag{7}$$

$$z_{pj} \in \{0,1\}, \qquad \forall j\in J, \forall p\in P, \tag{8}$$

$$\tilde{D}^2_{pj} = \frac{\left(\frac{\beta_{p,j+1}-\beta_{pj}}{\alpha_{j+1}-\alpha_j}\right) - \left(\frac{\beta_{p,j}-\beta_{p,j-1}}{\alpha_J-\alpha_{j-1}}\right)}{\alpha_{j+1}-2\alpha_j+\alpha_{j-1}} \tag{9}$$

$$D2_{pj} > \tilde{D}^2_{pj} \qquad \forall j\in J_{(-1)}, \forall p\in P, \tag{10}$$

$$D2_{pj} > -\tilde{D}^2_{pj} \qquad \forall j\in J_{(-1)}, \forall p\in P, \tag{11}$$

$$\beta_{0,j-1} + \beta^T_{j-1}x_t \leq \beta_{0j} + \beta^T_j x_t, \qquad \forall t\in T, \forall j\in J_{(-1)}, \tag{12}$$

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

$$\underset{\beta_0,\beta,\varepsilon^+_{tj},\varepsilon^-_{tj}}{\arg\min} \quad \sum_{j\in J}\sum_{t\in T}\left(\alpha_j\varepsilon^+_{tj} + (1-\alpha_j)\varepsilon^-_{tj}\right) + \lambda\sum_{p=1}^{P}\xi_{pj} + \gamma\sum_{j\in J'} D2_{pj} \tag{13}$$

$$\text{s.t.} \quad \varepsilon^+_{tj} - \varepsilon^-_{tj} = y_t - \beta_{0j} - \sum_{p=1}^{P}\beta_{pj}\tilde{x}_{t,p}, \qquad \forall t \in T, \forall j \in J, \tag{14}$$

$$\varepsilon^+_{tj}, \varepsilon^-_{tj} \geq 0, \qquad \forall t \in T, \forall j \in J, \tag{15}$$

$$\xi_{p\alpha} \geq \beta_{pj}, \qquad \forall p \in P, \forall j \in J, \tag{16}$$

$$\tilde{D}^2_{pj} = \frac{\left(\frac{\beta_{p,j+1}-\beta_{pj}}{\alpha_{j+1}-\alpha_j}\right) - \left(\frac{\beta_{p,j}-\beta_{p,j-1}}{\alpha_J-\alpha_{j-1}}\right)}{\alpha_{j+1}-2\alpha_j+\alpha_{j-1}} \tag{17}$$

$$D2_{pj} > \tilde{D}^2_{pj} \qquad \forall j \in J_{(-1)}, \forall p \in P, \tag{18}$$

$$D2_{pj} > -\tilde{D}^2_{pj} \qquad \forall j \in J_{(-1)}, \forall p \in P, \tag{19}$$

$$\beta_{0,j-1} + \beta^T_{j-1}x_t \leq \beta_{0j} + \beta^T_j x_t, \qquad \forall t \in T, \forall j \in J_{(-1)}, \tag{20}$$

$$\xi_{p\alpha} \geq -\beta_{pj}, \qquad \forall p \in P, \forall j \in J. \tag{21}$$

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

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

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

$$\beta^{*LASSO}_{\theta,p} = 0 \Longrightarrow \beta^{*}_{\theta,p} = 0.$$

- 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

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

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

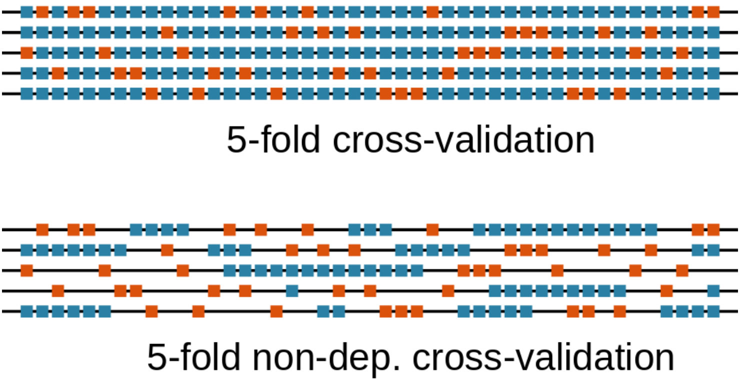


Figure 2: $\mathcal{K}$-fold CV and $\mathcal{K}$-fold with non-dependent data. Observations in blue are used to estimation and in orange for evaluation. Note that non-dependent data doesn't use all dataset in each fold.

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:

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

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

Nonparametric model

Nonparametric model - Formulation

$$
\begin{array}{lll}
\min\limits_{q_{\alpha t}, \delta_t^+, \delta_t^-, \xi_t} & \sum_{\alpha \in A} \sum_{t \in T'} \left(\alpha \delta_{t\alpha}^+ + (1-\alpha)\delta_{t\alpha}^- \right) & \\
& + \lambda_1 \sum_{t \in T'} \gamma_{t\alpha} + \lambda_2 \sum_{t \in T'} \xi_{t\alpha} & \\
s.t. & \delta_t^+ - \delta_{t\alpha}^- = y_t - q_{t\alpha}, & \forall t \in T', \forall \alpha \in A, \\
& D_{t\alpha}^1 = \frac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t}, & \forall t \in T', \forall \alpha \in A, \\
& D_{t\alpha}^2 = \frac{\left(\frac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t}\right) - \left(\frac{q_{\alpha t} - q_{\alpha t-1}}{x_t - x_{t-1}}\right)}{x_{t+1} - 2x_t + x_{t-1}}. & \forall t \in T', \forall \alpha \in A, \\
& \gamma_{t\alpha} \geq D_{t\alpha}^1, & \forall t \in T', \forall \alpha \in A, \\
& \gamma_{t\alpha} \geq -D_{t\alpha}^1, & \forall t \in T', \forall \alpha \in A, \\
& \xi_{t\alpha} \geq D_{t\alpha}^2, & \forall t \in T', \forall \alpha \in A, \\
& \xi_{t\alpha} \geq -D_{t\alpha}^2, & \forall t \in T', \forall \alpha \in A, \\
& \delta_{t\alpha}^+, \delta_{t\alpha}^-, \gamma_{t\alpha}, \xi_{t\alpha} \geq 0, & \forall t \in T', \forall \alpha \in A, \\
& q_{t\alpha} \leq q_{t\alpha'}, & \forall t \in T', \forall (\alpha, \alpha') \in A \times A, \alpha <
\end{array}
$$

Nonparametric vs. Linear Model

- The nonparametric approach is more flexible to capture heteroscedasticity.

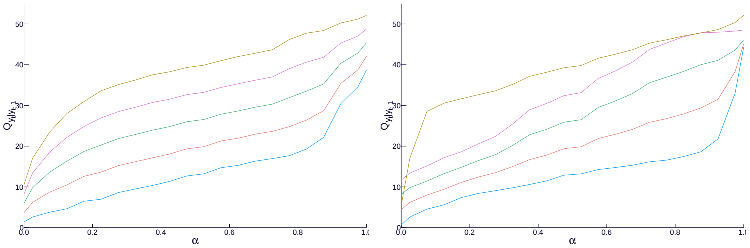


Figure 3: Estimated quantile functions, for different values of y_{t-1}. On the left using a linear model and using a nonparametric approach on the right.

Nonparametric vs. Linear Model

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

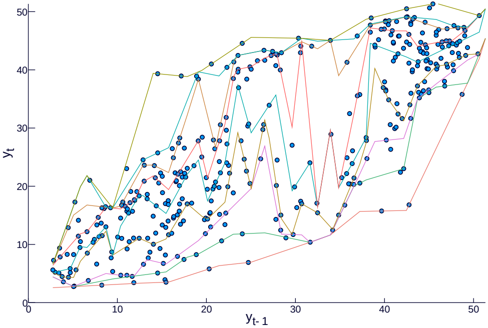


Figure 4: Example of a overfitted quantile function

Final

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.