Nongaussian time series model via Quantile Regression

Marcelo Ruas and Alexandre Street, *Member, IEEE*

***Abstract*—Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.**

***Keywords*—Quantile Regression, Model Identification, Nongaussian time series model**

TODO LIST

- Melhorar esta parte? ”é importante prever bem quantis altos e baixos p analise de risco - fica prejudicada pela dificuldade de previsao destes quantis” 1
- melhorar 1
- Definir Y_t no lugar de Y? 2

I. INTRODUCTION

Renewable energy power is a hot topic in the recent years. It is a much cleaner way of producing energy than by using other sources such as coal and gas, and with less hazard potential than nuclear power plants. The installed capacity of renewable plants has been increasing in a fast pace and projections point that wind power alone will account to 18% of global power by 2050 [1]. In spite of being a welcoming technology, it brings new challenges for power system planners, operators, and agents. It is important to have good forecasts of either high and low quantiles to analyze the risk. the complex behavior of wind is very difficult to model and predict. Having better prediction models can help the planner to make better and less risky decisions, increasing the attractiveness of renewable energy to the energy system. In this work we will investigate how to model dynamics of renewable energy time series in both short and long terms.

Conventional statistics are often focused on estimating the conditional mean of a given random variable. This is not very useful when dealing with renewable energy, as the variability and the notion of risk is extremely important for planning. One of the first works in wind power prediction, [3] treated the nonlinearity of wind power by applying a transformation on the prediction of wind speed, which is modeled by an autoregressive process. The data is standardized to account for the normal variation during the day. [4] estimated the k-step-ahead conditional density function using the Chapman-Kolmorov relation. The method is applied on a non-linear autoregressive time series. Our attention is, then, turned to probabilistic forecasting models. Among these, there are many possibilities. [2] reviews the commonly used methodologies for this task, separating them in parametric and nonparametric classes. Main characteristics of **parametric models** are (i) assuming a distribution shape and (ii) low computational costs. ARIMA-GARCH, for example, model the renewable series by assuming the distribution *a priori*. On the other hand, **nonparametric models** (i) don't require a distribution to be specified, (ii) needs mode data to produce a good approximation and (iii) have a higher computational cost. Popular methods are Quantile Regression, Kernel Density Estimation, Artificial Intelligence or a mix of them.

The seminal work [5] defines QR as we use today. Instead of using the classical regression to estimate the conditional mean, the QR determines any quantile from the conditional distribution. Applications are enormous, ranging from risk measuring at financial funds (the Value-at-Risk) to ... QR is a powerful tool for measuring quantile others than the median. By estimating many quantiles on a thin grid of probabilities, one can have as many points as desired of the estimated conditional distribution function. As we are working with time series, we rely on the results of [13], which defines the quantile autoregression, extending the application of QR on cases where the covariates are lagged values of y_t. In this work, beyond autoregressive terms, there are also exogenous variables as covariates.

The approach by [6] is to use QR with a nonparametric methodology. The authors add a penalty term based on the Reproducing Kernel Hilbert Space, which allows a nonlinear relationship between the explanatory variables and the output. This paper also develops an on-line learning technique, where the model is easily updated after each new observation. In [7], wind power probabilistic forecasts are made by using QR with a special type of Neural Network (NN) with one hidden layer, called extreme learning machine. In this setup, each quantile is a different linear combination of the features of the hidden layer. Although assuming a distribution for the errors is essential for the majority of models, the work by [8] proposes the usage of QR to circumvent this assumption. The authors use the weighted Nadaraya-Watson to estimate the conditional function in the time series. Another important feature is the ability to select important covariates in a QR. The work by [9] defines the proprieties and convergence rates for QR when adding a penalty proportional to the ℓ_1-norm to

Melhorar esta parte? ”é importante prever bem quantis altos e baixos p analise de risco - fica prejudicada pela dificuldade de previsao destes quantis”

melhorar

perform variable selection, using the same idea as the LASSO [10]. The ADALASSO equivalent to QR is proposed by [11]. In this variant, the penalty for each variable has a different weight, and this modification ensures that the oracle propriety is being respected.

A common finding is that wind and solar time series don't have a Gaussian behavior. Furthermore, we can't point any distribution that fits on the data undoubtedly. For this reason, we choose to use a nonparametric approach, more specifically the Quantile Regression (QR), as defined in [13]. However, when estimating a distribution function, as each quantile is estimated independently, the monotonicity of the distribution function may be violated. This issue is a known as crossing-quantiles. To get around it we propose to either add a constraint on the optimization model or make a transformation as in [14], which can be estimated independently.

Using QR as in [12] is not a novelty when predicting the conditional distribution of wind power time series, as is already done in [7], [15]–[17]. In this article, we combine QR with regularization techniques. On the Mixed Integer Linear Programming (MILP) approach, we use the best subset selection, which [18] does for minimizing the quadratic error and the ℓ_1 penalization as in [9] or [11] (the AdaLasso variant, where each coefficient may have a different weight on the objective function to ensure oracle properties). On the advantages of QR we highlight the fact that it is a direct way of finding a given quantile directly and it is data driven, so it doesn't depend on distribution assumptions.

For the best of the authors knowledge, no other work has developed a methology where regularization and estimation of the conditional quantile regression is carried on at the same time. We propose to attack both problems at the same time by using Mixed Integer Linear Programming (MILP). This approach follows the work by [18], where the best subset selection for a given number K of covariates is selected using MILP. In our approach, we estimate the conditional distribution from a thin grid of quantiles, all at the same time. This method is then computationally intensive, but provides the global optimal solution for the optimization problem that determines the QR, restricted to the degree of parsimony chosen. The objective of this paper is, then, to propose a new methodology to address nonparametric time-series focused on renewable energy. In our analysis, we develop both nonlinear and linear models for QR. The main contributions are:

- A nonparametric methodology to model the conditional distribution of a given time series.
- On the linear case, we propose a parsimonious methodology that selects the global optimal solution.
- Regularization techniques applied to an ensemble of quantile regressions to estimate the conditional distribution
- A nonlinear QR

The remaining of the paper is organized as follows. In section II, we present both the linear parametric and the nonlinear QR based time series models. In section III, we discuss the estimation procedures for them. The regularization strategies are also presented on this section. Finally, in section IV, a case study using real data from both solar and wind power is presented in order to test our methodology. Section V will conclude this article.

II. Quantile Regression based time series model

Let the conditional quantile function $Q_{Y|X=x} : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ (in short, from now on, $Q_{Y|X}(\cdot,\cdot)$) be defined as

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

where Y is a random variable and X a random vector of dimension d. Let a dataset be composed of n observations of $\{y_t, x_t\}$. To calculate the sample quantile, the optimization based on this finite number of observations, instead of integrating over all domain of random variable Y. The sample quantile is given by

Definir Y_t no lugar de Y?

$$\hat{Q}_{Y|X}(\alpha, \cdot) \quad \in \quad \underset{q_\alpha(\cdot)}{\arg\min} \sum_{t \in T} \rho_\alpha(y_t - q_\alpha(\cdot)), \quad (2)$$

$$q_\alpha \in \mathcal{Q}, \quad (3)$$

and ρ is the check function defined as

$$\rho_\alpha(x) = \begin{cases} \alpha x & \text{if } x \geq 0 \\ (1-\alpha)x & \text{if } x < 0 \end{cases}. \quad (4)$$

Let Y be a real valued random variable. The quantile function $Q_Y : [0,1] \to \mathbb{R}$ is defined pointwise by its α-quantile, which is given by where F_Y is the distribution function of random variable Y and $\alpha \in [0,1]$. Equation 1 defines what we call from now on as the quantile function $Q_Y(\cdot)$, in relation to random variable Y. In this article, we are interested in a conditional quantile function $Q_{Y|X=x} : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ (in short, from now on, $Q_{Y|X}(\cdot,\cdot)$), where X can be a vector.

In this work, we use the sample quantile, where we calculate the optimization based on a finite number of observations, instead of integrating over all domain of random variable Y. For the specific case where the random variable is a time series y_t, quantiles are estimated from a n size sample of observations of y_t and a explanatory variable x_t for each t, such that our random sample is formed by the sequence $\{y_t, x_t\}_{t=1}^n$. To estimate the α-quantile from a sample, we change **??** for the following optimization problem:

where $T = \{1, \ldots, n\}$, $|x|^+ = \max\{0, x\}$ and $|x|^- = -\min\{0, x\}$. The solution from the above problem is an estimator $\hat{Q}_{Y|X}$ for the quantile function $Q_{Y|X}$.

To model this problem as a Linear Programming problem, thus being able to use a modern solver to fit our model, we create variables ε_t^+ e ε_t^- to represent $|y - q(\cdot)|^+$ and $|y - q(\cdot)|^-$, respectively. The optimal argument $q^*_\alpha(\cdot)$ on the Linear Programming problem 5 is the estimated α-quantile for the given random sample.

$$\begin{aligned} q^*_\alpha(.) \in \underset{q_\alpha(\cdot), \varepsilon_t^+, \varepsilon_t^-}{\arg\min} & \sum_{t \in T} \left(\alpha\varepsilon_t^+ + (1-\alpha)\varepsilon_t^-\right) \\ \text{s.t. } & \varepsilon_t^+ - \varepsilon_t^- = y_t - q_\alpha(x_t), \qquad \forall t \in T, \\ & \varepsilon_t^+, \varepsilon_t^- \geq 0, \qquad \forall t \in T. \end{aligned} \quad (5)$$

Let A be a set containing a sequence of probabilities α_i such that $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_Q < 1$. This set represents a

finite discretization of the interval $[0, 1]$. One of our goals with quantile regression is to estimate a quantile function $\hat{Q}_{Y|X}$ of a given real valued random variable X from a sequence of quantiles $q_{\alpha_1}(x_t) \leq q_{\alpha_2}(x_t) \leq \cdots \leq q_{\alpha_{|A|}}(x_t)$, with $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{|A|} < 1$, for any given t. The process of fitting $\hat{Q}_{Y|X}$ is by mapping every α_i with its estimated quantile $\hat{q}_{\alpha_i}(x_t)$. The denser the grid of values in A, better is the approximation of $Q_{Y|X}$. Thus, the distribution found for Y is nonparametric, as no previous assumptions are made about its shape, and its form is fully recovered by the data we have.

A typical problem, however, arises when working with quantile regression. When quantiles are estimated independently, it is possible to find $q_\alpha(x_t) > q_{\alpha'}(x_t)$, for a given t, when $\alpha_1 < \alpha_2$. An example can be seen on Figure **??**, where quantiles $\alpha = 0.95$ and $\alpha = 0.9$ cross. This problem, called *crossing quantiles*, can be prevented by estimating all quantiles with a single minimization problem.

As opposed to traditional parametric models where one has to assume the error distribution on the model

$$y_t = h(F_t) + \varepsilon_t, \tag{6}$$

as we choose the nonparametric methodology we are released from this step. A distribution for ε_t is defined within the estimation procedure. The predictive distribution is derived from the individual quantiles

$$\hat{q}^{\alpha}_{t+k|k} = \hat{y}_{t+k|t} + \hat{F}^{\varepsilon}_{t,k}(\alpha). \tag{7}$$

[7] defines a model based on extreme machine learning to estimate every α-quantile in a thin grid $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_Q < 1$.

A. Linear parametric explanatory model

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

B. General non-linear nonparametric model

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

REFERENCES

[1] International energy agency. [Online]. Available: https://www.iea.org/newsroom/news/2013/october/wind-power-seen-generating-up-to-18-of-global-power-by-2050.html

[2] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, Apr. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032114000446

[3] B. G. Brown, R. W. Katz, and A. H. Murphy, "Time Series Models to Simulate and Forecast Wind Speed and Wind Power," *Journal of Climate and Applied Meteorology*, vol. 23, no. 8, pp. 1184–1195, Aug. 1984. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%281984%29023%3C1184%3ATSMTSA%3E2.0.CO%3B2

[4] R. Moeanaddin and H. Tong, "Numerical Evaluation of Distributions in Non-Linear Autoregression," *Journal of Time Series Analysis*, vol. 11, no. 1, pp. 33–48, Jan. 1990. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.1990.tb00040.x/abstract

[5] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

[6] C. Gallego-Castillo, R. Bessa, L. Cavalcante, and O. Lopez-Garcia, "On-line quantile regression in the rkhs (reproducing kernel hilbert space) for operational probabilistic forecasting of wind power," *Energy*, vol. 113, pp. 355–365, 2016.

[7] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct Quantile Regression for Nonparametric Probabilistic Forecasting of Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, Jul. 2017.

[8] Z. Cai, "Regression Quantiles for Time Series," *Econometric Theory*, vol. 18, no. 1, pp. 169–192, 2002. [Online]. Available: http://www.jstor.org/stable/3533031

[9] A. Belloni and V. Chernozhukov, "L1-Penalized Quantile Regression in High-Dimensional Sparse Models," *arXiv:0904.2931 [math, stat]*, Apr. 2009, arXiv: 0904.2931. [Online]. Available: http://arxiv.org/abs/0904.2931

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[11] G. Ciuperca, "Adaptive LASSO model selection in a multiphase quantile regression," *Statistics*, vol. 50, no. 5, pp. 1100–1131, Sep. 2016. [Online]. Available: http://dx.doi.org/10.1080/02331888.2016.1151427

[12] R. Koenker, *Quantile regression*. Cambridge university press, 2005, no. 38.

[13] R. Koenker, Z. Xiao, J. Fan, Y. Fan, M. Knight, M. Hallin, B. J. M. Werker, C. M. Hafner, O. B. Linton, and P. M. Robinson, "Quantile Autoregression [with Comments, Rejoinder]," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 980–1006, 2006. [Online]. Available: http://www.jstor.org/stable/27590777

[14] V. Chernozhukov, I. Fernández-Val, and A. Galichon, "Quantile and Probability Curves Without Crossing," *Econometrica*, vol. 78, no. 3, pp. 1093–1125, May 2010. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.3982/ECTA7880/abstract

[15] J. K. Møller, H. A. Nielsen, and H. Madsen, "Time-adaptive quantile regression," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292–1303, Jan. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947307002502

[16] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 95–108, 2006.

[17] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, Jan. 2004. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/we.107/abstract

[18] D. Bertsimas, A. King, and R. Mazumder, "Best Subset Selection via a Modern Optimization Lens," *arXiv:1507.03133 [math, stat]*, Jul. 2015, arXiv: 1507.03133. [Online]. Available: http://arxiv.org/abs/1507.03133