Proposta de Tese

Marcelo Castiel Ruas[*], Alexandre Street[†]

June 21, 2017

[*]Aluno de doutorado do Departamento de Engenharia Elétrica da PUC-RIO.
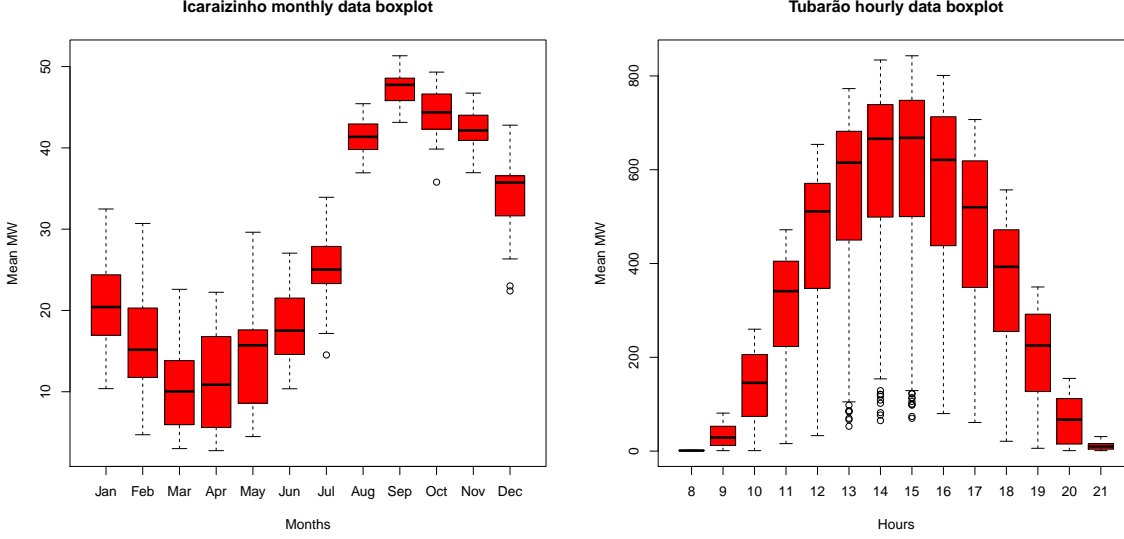[†]Professor do Departamento de Engenharia Elétrica da PUC-RIO.

Figure 1.1: Boxplots showing seasonality for monthly and hourly data.

1 Introduction

Quantile Regression is a powerful tool for measuring quantiles others than the median or predicting the mean. A quantile of a random variable is important in risk measuring, as we can measure the probability of occurrence of extreme events, and in many other fields. While working with energy forecasts, quantile regression can produce interesting results when working with both short term (hourly) or long term (monthly) data. As an example, we present a solar time series for the short term and a wind time series for long term. The first set of data is measured at the location of Tubarao (Brazil) on the year of 2014, while the latter is a dataset of mean power monthly observations from Icaraizinho (Brazil) between 1981 to 2011 of measured in Megawatts. Figure 1.1 illustrate the seasonality present in these datasets.

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

$$Q_Y(\alpha) = F_Y^{-1}(\alpha) = \inf\{y : F_Y(y) \geq \alpha\}, \tag{1.1}$$

where F_Y is the distribution function of random variable Y and $\alpha \in [0,1]$. Equation 1.1 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.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. The conditional quantile function can be found as the result of the following optimization problem:

$$Q_{Y|X}(\alpha, \cdot) \quad \in \quad \underset{q_\alpha(\cdot)}{\arg\min} \quad (1-\alpha) \int_{\omega \in \Omega} |Y(w) - q_\alpha(X(w))|^- P(dw) \tag{1.2}$$

$$+(\alpha) \int_{\omega \in \Omega} |Y(w) - q_\alpha(X(w))|^+ P(dw)$$

$$q_\alpha \in \mathcal{Q}, \tag{1.3}$$

The argument of optimization problem described on equation 1.2 is the function q_α, which belongs to a function space $\mathcal{Q}$. We might have different assumptions for the space $\mathcal{Q}$, depending on the type of function we want to find for q_α. A few properties, however, must be achieved by our choice. The conditional quantile function $Q_{Y|X}(\alpha)$ must be monotone on α, and its first derivative must be limited.
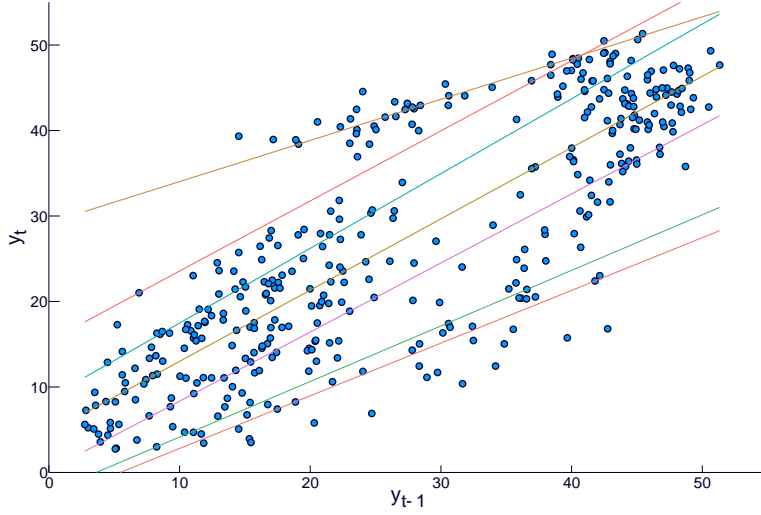
Figure 1.2: Linear quantile estimator with crossing quantiles for $\alpha = 0.95$ and $\alpha = 0.9$

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 1.2 for the following optimization problem:

$$\hat{Q}_{Y|X}(\alpha, \cdot) \quad \in \quad \underset{q_\alpha(\cdot)}{\arg\min} \sum_{t \in T} \alpha |y_t - q_\alpha(\cdot)|^+ + \sum_{t \in T} (1 - \alpha)|y_t - q_\alpha(\cdot)|^-, \tag{1.4}$$

$$q_\alpha \in \mathcal{Q}, \tag{1.5}$$

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 1.6 is the estimated α-quantile for the given random sample.

$$\begin{aligned}
q_\alpha^*(.) \in \quad & \underset{q_\alpha(\cdot), \varepsilon_t^+, \varepsilon_t^-}{\arg\min} \sum_{t \in T} \left(\alpha \varepsilon_t^+ + (1 - \alpha) \varepsilon_t^- \right) \\
& \text{s.t. } \varepsilon_t^+ - \varepsilon_t^- = y_t - q_\alpha(x_t), \qquad \forall t \in T, \\
& \qquad \varepsilon_t^+, \varepsilon_t^- \geq 0, \qquad \forall t \in T.
\end{aligned} \tag{1.6}$$

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 1.2, 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.

In order to estimate all quantiles simultaneously, the new objective function will be the sum of all individual objective functions, as well as include all constraints from all individual problems. The only

3

difference is the inclusion of an equation to guarantee that quantiles won't cross. When modifying problem 1.6 to account for all quantiles, we have the following new problem:

$$\{q_\alpha^*(\cdot)\}_{\alpha \in A} \in \underset{q_\alpha(\cdot), \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-}{\arg\min} \quad \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1-\alpha) \varepsilon_{t\alpha}^- \right) \tag{1.7}$$

$$\text{s.t.} \qquad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - q_\alpha(x_t), \qquad \forall t \in T, \forall \alpha \in A, \tag{1.8}$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \qquad \forall t \in T, \forall \alpha \in A, \tag{1.9}$$

$$q_\alpha(x_t) \leq q_{\alpha'}(x_t), \qquad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha' > \alpha, \tag{1.10}$$

where constraint 1.10 assures that no lower quantile will have a bigger value than a higher quantile.

The next section discusses with bigger details how to fit a distribution function $Q_{Y|X}(\alpha, x)$ from a sequence of estimated quantiles, as well as showing two different strategies to estimate them: linear models and nonparametric models. In the former, q_α is a linear function of an explanatory variable x_t. In the latter, we let $q_\alpha(x_t)$ assume any functional form. To prevent overfitting, however, we penalize the function's roughness by incorporating a penalty on the second derivative.

As opposed to the conventional work of doing a model to forecast the conditional mean, our work focus on finding a distribution for y_t on each t.

We find a time series model, based on quantile autoregression (as in Konker 2005).

As we are interested in the whole distribution of $\hat{y}_{t+k|t}$, we estimate a phin grid of quantiles in $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{|A|} < 1$, such that the distribution can be well approximated.

As a Quantile Autoregression model, we are interested in selecting the best subset of variables do model the time series.

As we are trying to model the whole k-step ahead distribution, we estimate many quantiles. We didn't find any previous work where a given α-quantile model influenced another model.

In all works found, each quantile is estimated separately.

Regularization can be done by introducing a penalty on the ℓ_1-norm of the coefficients. The work by [1] defines proprieties and convergence analysis. The AdaLasso variant, where each coefficient may have a different weight on the objective function to ensure oracle proprieties, is developed on [2].

References

[1] Alexandre Belloni and Victor Chernozhukov. L1-Penalized Quantile Regression in High-Dimensional Sparse Models. *arXiv:0904.2931 [math, stat]*, April 2009. arXiv: 0904.2931.

[2] Gabriela Ciuperca. Adaptive LASSO model selection in a multiphase quantile regression. *Statistics*, 50(5):1100–1131, September 2016.