Pontíficia Universidade Católica do Rio de Janeiro
Departamento de Engenharia Elétrica

Quantile Regression

Marcelo Castiel Ruas*, Henrique Helfer Hoeltgebaum†, Alexandre Street‡,
Cristiano Fernandes§

March 3, 2017

*Aluno de doutorado do Departamento de Engenharia Elétrica da PUC-RIO.
†Aluno de doutorado do Departamento de Engenharia Elétrica da PUC-RIO.
‡Professor do Departamento de Engenharia Elétrica da PUC-RIO.
§Professor do Departamento de Engenharia Elétrica da PUC-RIO.

List of variables

$Q_Y(\cdot)$	Quantile function of real random variable Y	
$F_Y(\cdot)$	Distribution function of real random variable Y	
A	A set of probabilities. $A = \{\alpha_1, \ldots, \alpha_{\lvert A\rvert}$	$q_\alpha(x_t)$ α-quantile, given x_t
T_n	Set of size n of observation indexes, such that $T_n = \{1, 2, \ldots, n\}$	
$\{y_t\}_{t\in T_n}$	Sample of time series y_t	
$\{x_t\}_{t\in T_n}$	Sample of d-dimensional time series x_t	
S	Number of different paths in the simulation	
K	Size of each path in the simulation	

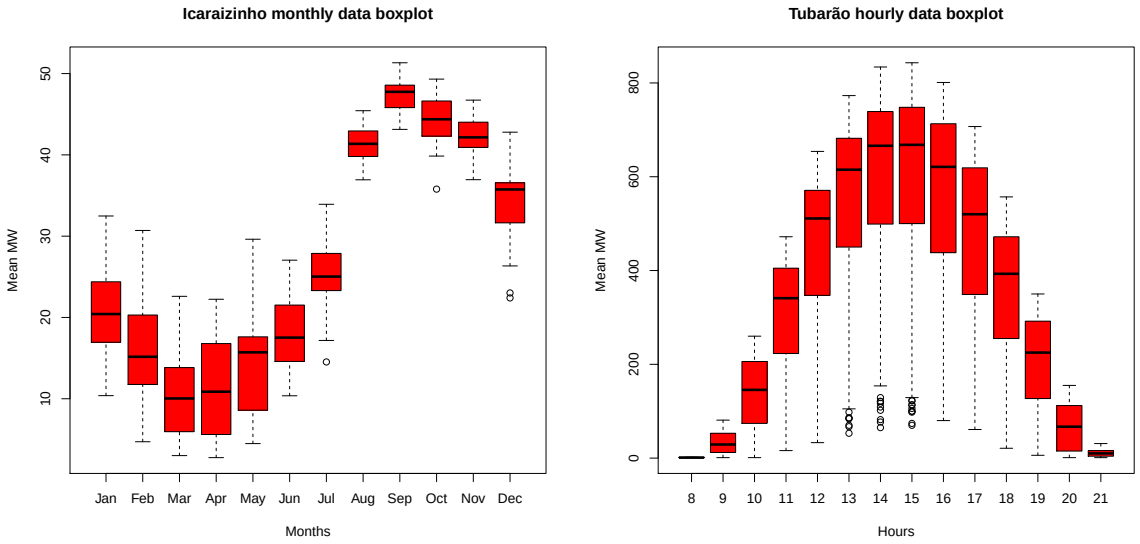

Figure 1.1: Boxplots showing seasonality for monthly and hourly data.

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.

In this work, we apply a few different techniques to forecast the quantile function a few steps ahead. The main frameworks we investigate are parametric linear models and a non-parametric regression. We also investigate how to apply quantile estimations to produce an empirical distribution for the k-step ahead forecasting by using a nonparametric approach.

To make good predictions of random variables, one must find good explanatory variables: it can be either autoregressive, exogenous terms or even a deterministic function that repeats itself. Figure 1.2 shows scatter plots relating y_t with its first lag for both short and long term. We can see that in both of them past values are good explanatory variables to use for forecasting.

In contrast to the linear regression model through ordinary least squares (OLS), which provides only an estimation of the dependent variable conditional mean, quantile regression model yields a much more detailed information concerning the complex relationship about the dependent variable and its covariates. Here we denote as parametric linear model the well-known quantile regression model [3].

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 $Q_{Y|X=x}(\alpha, x)$ (in short, from now on, $Q_{Y|X}(\alpha, x)$), 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

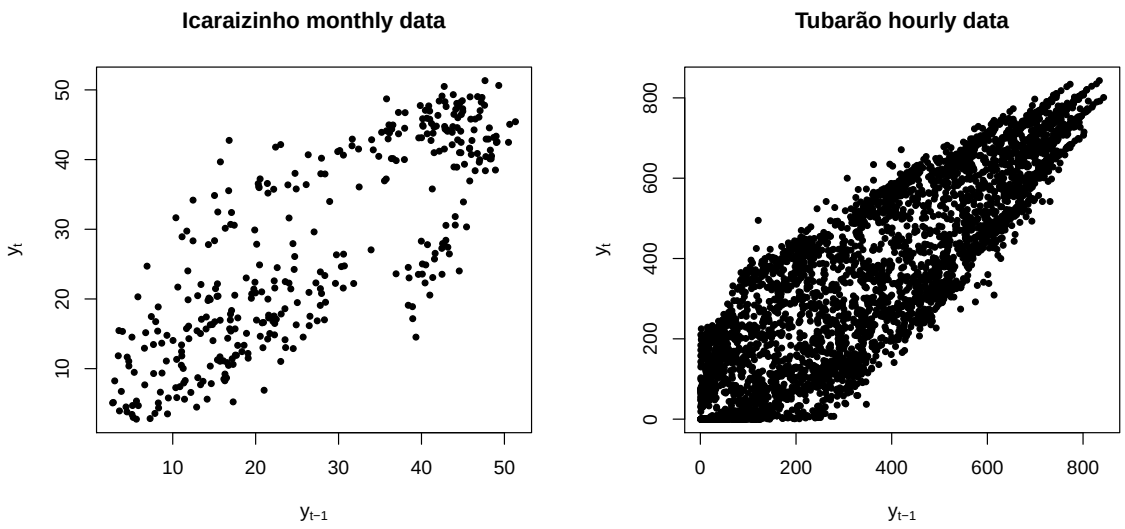

Figure 1.2: Relationship between y_t and its first lags for two selected series.

of the following optimization problem:

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

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

where A is 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 $[0, 1]$. These values must be close enough so that $Q_{Y|X}(\alpha)$ has a precise representation. 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.

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:

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

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

where $q_\alpha(x_t)$ is the estimated quantile value at a given time t and $|x|^+ = \max\{0, x\}$ 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 $|y - q(x_t)|^+$ and $|y - q(x_t)|^-$, respectively. The optimal argument $q_\alpha^*(\cdot)$ on the Linear Programming problem 1.6 is the estimated α-quantile for

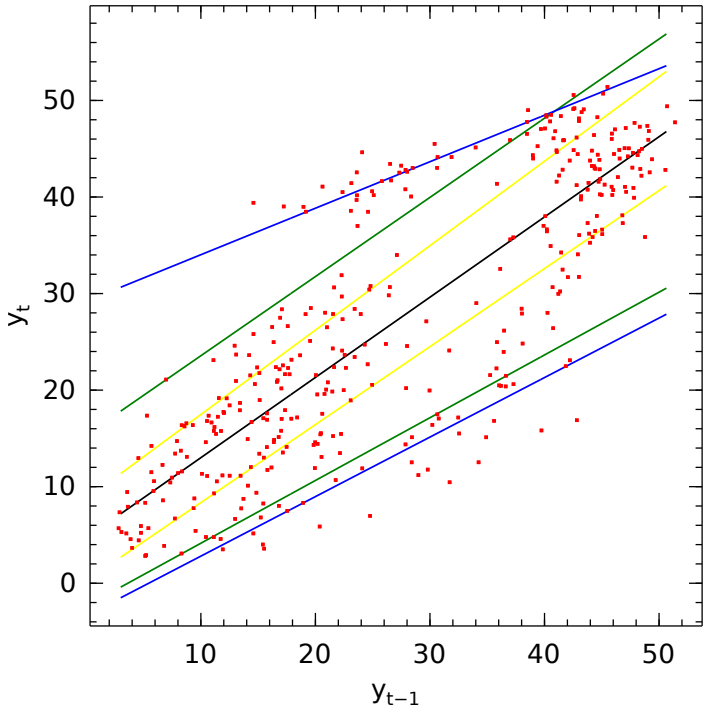

Figure 1.3: Linear quantile estimator with crossing quantiles for $\alpha = 0.95$ and $\alpha = 0.9$

the given random sample.

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

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 $\hat{Q}_Y$ is by mapping every α_i with its estimated quantile $\hat{q}_{\alpha_i}(x_t)$. When this sequence of chosen α_i is thin enough, we have a good 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 1.3, 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.

Let $T_n = \{1, 2, \dots, n\}$. 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 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:

$$\min_{q_\alpha(\cdot),\varepsilon_{t,\alpha}^+,\varepsilon_{t,\alpha}^-} \quad \textstyle\sum_{\alpha\in A}\sum_{t\in T_n}\left(\alpha\varepsilon_{t,\alpha}^+ + (1-\alpha)\varepsilon_{t,\alpha}^-\right) \tag{1.7}$$

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

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

$$q_\alpha(x_t) \leq q_{\alpha'}(x_t), \qquad \forall t \in T_\tau, \forall(\alpha,\alpha') \in AxA, \alpha' \geq \alpha, \tag{1.10}$$

where constraint 2.8 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.

In section 3 we investigate how to simulate S scenarios of y_t, considering a linear model and errors ε_t for which the distribution is unknown. To address this issue, we use quantile linear regression to calculate a thin grid of quantiles and fit a distribution function $\hat{F}_{y_t}$. This function will be used to simulate the innovations on the model.

2 Estimating distribution function from quantile regressions

In many applications where a time series model is employed, we often consider the innovations' distribution as known. Take, for example, the AR(p) model:

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i y_{t-1}$$

In this model, errors ε_t are assumed to have normal distribution with zero mean.

When we are dealing with meteorological time series, however, we can't always assume normality. In these cases, one can either find a distribution that has a better fit to the data or have a nonparametric method to estimate the distribution directly from the available data.

In a time series framework, where a time series y_t is given by a linear model of its regressors x_t

$$y_t = \beta^T x_t + \varepsilon_t,$$

we propose to estimate the k-step ahead distribution of y_t with a nonparametric approach. Let an empirical α-quantile $\hat{q}_\alpha \in \mathcal{Q}$ be a functional belonging to a functional space. In any given t, we can estimate the sequence of quantiles $\{q_\alpha(x_t)\}_{\alpha \in A}$ by solving the problem defined on equations (2.5)-(2.8). After evaluating this sequence, by making equal

$$\hat{Q}_{y_t|X}(\alpha) = \hat{q}_\alpha(x_t), \qquad \forall \alpha \in A, \tag{2.1}$$

we have a set of size $|A|$ of values to define the discrete function over the first argument $\hat{Q}_{y_t|x_t}(\alpha, X = x_t) : A \times \mathbb{R}^d \to \mathbb{R}$. The goal of having function $\hat{Q}$ is to use it as base to construct the estimated quantile function $\hat{Q}'_{y_t|X=x_t}(\alpha, x_t) : [0,1] \times \mathbb{R}^d \to \mathbb{R}$.

A problem arises for the distribution extremities, because when $\alpha = 0$ or $\alpha = 1$, the optimization problem becomes unbounded. In order to find values for $\hat{Q}(\alpha, x_t)$ when $\alpha \in \{0, 1\}$, we chose to linearly extrapolate its values. Note that as $A \subset [0,1]$, the domain of $\hat{Q}$ is also a subset of the domain of $\hat{Q}'$. The estimative of $\hat{Q}'$ is done by interpolating points of $\hat{Q}$ over the interval $[0,1]$. Thus, the distribution found for $\hat{y}_\tau$ is nonparametric, as no previous assumptions are made about its shape, and its form is fully recovered by the data we have.

We investigate two different approaches for Q_{y_t} by the functional structure of each individual $q_\alpha(x_t)$. In section 2.1, we explore the case where the individual quantiles $q_\alpha(x_t)$ are a linear function of its arguments:

$$\hat{q}_\alpha(x_t) = \beta_{0,\alpha} + \beta_\alpha^T x_t, \tag{2.2}$$

where β^α is a vector of coefficients for the explanatory variables.

In section 2.2 we introduce a Nonparametric Quantile Autoregressive model with a ℓ_1-penalty term, in order to properly simulate densities for several α-quantiles. In this nonparametric approach we don't assume any form for $q_\alpha(x_t)$, but rather let the function adjust to the data. To prevent overfitting, the ℓ_1 penalty for the second derivative (approximated by the second difference of the ordered observations) is included in the objective function. The result of this optimization problem is that each $q_\alpha(x_t)$ will be a function with finite second derivative.

In order to find good estimates for $Q_{y_t}(\alpha)$ when α approaches 0 or 1, as well as performing interpolation on the values that were not directly estimated, we can either use a kernel smoothing function, splines, linear approximation, or any other method. **This will be developed later.**

2.1 Linear Models for the Quantile Autoregression

Given a time series $\{y_t\}$, we investigate how to select which lags will be included in the Quantile Autoregression. We won't be choosing the full model because this normally leads to a bigger variance in our estimators, which is often linked with bad performance in forecasting applications. So our strategy will be to use some sort of regularization method in order to improve performance. We investigate two ways of accomplishing this goal. The first of them consists of selecting the best subset of variables through Mixed Integer Programming, given that K variables are included in the model.

Using MIP to select the best subset of variables is investigated in [1]. The second way is including a ℓ_1 penalty on the linear quantile regression, as in [2], and let the model select which and how many variables will have nonzero coefficients. Both of them will be built over the standard Quantile Linear Regression model. In the end of the section, we discuss a information criteria to be used for quantile regression and verify how close are the solutions in the eyes of this criteria.

When we choose $q_\alpha(x_t)$ to be a linear function

$$\hat{q}_\alpha(x_t) = \beta_{0\alpha} + \beta_\alpha^T x_t \tag{2.3}$$

we can substitute it on problem 1.6, getting the following LP problem:

$$\begin{aligned} \min_{\beta_0,\beta,\varepsilon_t^+,\varepsilon_t^-} \quad & \sum_{t\in T} \left(\alpha\varepsilon_t^+ + (1-\alpha)\varepsilon_t^-\right) \\ \text{s.t. } & \varepsilon_t^+ - \varepsilon_t^- = y_t - \beta_0 - \beta^T x_t, \qquad \forall t \in \{1,\dots,n\}, \\ & \varepsilon_t^+, \varepsilon_t^- \geq 0, \qquad \forall t \in \{1,\dots,n\}. \end{aligned} \tag{2.4}$$

To solve the crossing quantile issue discussed earlier, we employ a minimization problem with all quantiles at the same time and adding the non crossing constraints. The new optimization problem is shown below:

$$\min_{\beta_{0\alpha},\beta_\alpha,\varepsilon_{t\alpha}^+,\varepsilon_{t\alpha}^-} \quad \textstyle\sum_{\alpha\in A}\sum_{t\in T}\left(\alpha\varepsilon_{t,\alpha}^+ + (1-\alpha)\varepsilon_{t,\alpha}^-\right) \tag{2.5}$$

$$\text{s.t.} \quad \varepsilon_{t,\alpha}^+ - \varepsilon_{t,\alpha}^- = y_t - \beta_{0\alpha} - \beta_\alpha^T x_t, \qquad \forall t \in T_\tau, \forall \alpha \in A, \tag{2.6}$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \qquad \forall t \in T_\tau, \forall \alpha \in A, \tag{2.7}$$

$$q_\alpha(x_t) \leq q_{\alpha'}(x_t), \qquad \forall t \in T_\tau, \forall (\alpha,\alpha') \in A \times A, \alpha < \alpha', \tag{2.8}$$

When solving problem (2.5)-(2.8), the sequence $\{q_\alpha\}_{\alpha\in A}$ is fully defined by the values of $\beta_{0\alpha}^*$ and β_α^*, for every α.

2.2 Quantile Autoregression with a nonparametric approach

Fitting a linear estimator for the Quantile Auto Regression isn't appropriate when nonlinearity is present in the data. This nonlinearity may produce a linear estimator that underestimates the quantile for a chunk of data while overestimating for the other chunk. To prevent this issue from occurring we propose a modification which we let the prediction $q_\alpha(x_t)$ adjust freely to the data and its nonlinearities. To prevent overfitting and smoothen our predictor, we include a penalty on its roughness by including the ℓ_1 norm of its second derivative. For more information on the ℓ_1 norm acting as a filter, one can refer to [2].

This time, as opposed to when employing linear models, we don't suppose any functional form for $q_\alpha(x_t)$. This forces us to build each q_α differently: instead of finding a set of parameters that fully defines the function, we find a value for $q_\alpha(x_t)$ at each instant t. On the optimization problem, we will find optimal values for a variable $q_{\alpha t} \in \mathbb{R}$, each consisting of a single point. The sequence of $\{q_{\alpha t}^*\}$ will provide a discretization for the full function $\hat{q}_\alpha(x_t)$, which can be found by interpolating these points.

Let $\{\tilde{y}_t\}_{t=1}^n$ be the sequence of observations in time t. Now, let $\tilde{x}_t$ be the $p-$lagged time series of $\tilde{y}_t$, such that $\tilde{x}_t = L^p(\tilde{y}_t)$, where L is the lag operator. Matching each observation $\tilde{y}_t$ with its $p-$lagged correspondent $\tilde{x}_t$ will produce $n-p$ pairs $\{(\tilde{y}_t, \tilde{x}_t)\}_{t=p+1}^n$ (note that the first p observations of y_t must be discarded). When we order the observation of x in such way that they are in growing order

$$\tilde{x}^{(p+1)} \leq \tilde{x}^{(p+2)} \leq \cdots \leq \tilde{x}^{(n)},$$

we can then define $\{x_i\}_{i=1}^{n-p} = \{\tilde{x}^{(t)}\}_{t=p+1}^n$ and $\{y_i\}_{i=1}^{n-p} = \{\tilde{y}^{(t)}\}_{t=p+1}^n$ and $T' = \{2,\dots,n-p-1\}$. As we need the second difference of q_i, I has to be shortened by two elements.

Our optimization model to estimate the nonparametric quantile is as follows:

$$\begin{aligned}\hat{q}_\alpha(x_t) = \arg\min_{q_{\alpha t}} \sum_{t\in T'} \left(\alpha|y_t - q_{\alpha t}|^+ + (1-\alpha)|y_t - q_{\alpha t}|^-\right) \\ +\lambda_1 \sum_{t\in T'} |D^1_{x_t} q_{\alpha t}| + \lambda_2 \sum_{t\in T'} |D^2_{x_t} q_{\alpha t}|,\end{aligned} \tag{2.9}$$

where $D^1 q_t$ and $D^2 q_t$ are the first and second derivatives of the $q_\alpha(x_t)$ function, calculated as follows:

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

$$D^1_{t\alpha} = \frac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t}.$$

The first part on the objective function is the usual quantile regression condition for $\{q_{t\alpha}\}_{\alpha\in A}$. The second part is the ℓ_1-filter. The purpose of a filter is to control the amount of variation for our estimator $q_\alpha(x_t)$. When no penalty is employed we would always get $q_{\alpha t} = y_t$, for any given α. On the other hand, when $\lambda \to \infty$, our estimator approaches the linear quantile regression.

The full model can be rewritten as a LP problem as bellow:

$$\min_{q_{\alpha t}, \delta_t^+, \delta_t^-, \xi_t} \quad \sum_{\alpha\in A}\sum_{t\in T'} \left(\alpha\delta^+_{t\alpha} + (1-\alpha)\delta^-_{t\alpha}\right) \tag{2.10}$$

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

$$s.t. \quad \delta^+_t - \delta^-_{t\alpha} = y_t - q_{t\alpha}, \qquad \forall t \in T', \forall \alpha \in A, \tag{2.11}$$

$$D^1_{t\alpha} = \tfrac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t}, \qquad \forall t \in T', \forall \alpha \in A, \tag{2.12}$$

$$D^2_{t\alpha} = \tfrac{\left(\frac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t}\right) - \left(\frac{q_{\alpha t} - q_{\alpha t-1}}{x_t - x_{t-1}}\right)}{x_{t+1} - 2x_t + x_{t-1}}. \qquad \forall t \in T', \forall \alpha \in A, \tag{2.13}$$

$$\gamma_{t\alpha} \geq D^1_{t\alpha}, \qquad \forall t \in T', \forall \alpha \in A, \tag{2.14}$$

$$\gamma_{t\alpha} \geq -D^1_{t\alpha}, \qquad \forall t \in T', \forall \alpha \in A, \tag{2.15}$$

$$\xi_{t\alpha} \geq D^2_{t\alpha}, \qquad \forall t \in T', \forall \alpha \in A, \tag{2.16}$$

$$\xi_{t\alpha} \geq -D^2_{t\alpha}, \qquad \forall t \in T', \forall \alpha \in A, \tag{2.17}$$

$$\delta^+_{t\alpha}, \delta^-_{t\alpha}, \gamma_{t\alpha}, \xi_{t\alpha} \geq 0, \qquad \forall t \in T', \forall \alpha \in A, \tag{2.18}$$

$$q_{t\alpha} \leq q_{t\alpha'}, \qquad \forall t \in T', \forall (\alpha, \alpha') \in A \times A, \alpha(2.19)$$

The output of our optimization problem is a sequence of ordered points $\{(x_t, q_{t\alpha})\}_{t\in T}$, for all $\alpha \in A$. The next step is to interpolate these points in order to provide an estimation for any other value of x_t. To address this issue, we propose using a linear interpolation, that will be developed in another study. Note that $q_{t\alpha}$ is a variable that represents only one point of the α-quantile function $q_\alpha(x_t)$.

The quantile estimation is done for different values of λ. By using different levels of penalization on the second difference, the estimation can be more or less adaptive to the fluctuation. It is important to notice that the usage of the ℓ_1-norm as penalty leads to a piecewise linear solution $q_{t\alpha}$. Figure 2.1 shows the quantile estimation for a few different values of λ.

The first issue is how to select an appropriate value for λ. A simple way is to do it by inspection, which means to test many different values and pick the one that suits best our needs by looking at them. The other alternative is to use a metric to which we can select the best tune. We can achieve this by using a cross-validation method, for example.

The other issue occurs when we try to add more than one lag to the analysis at the same time. This happens because the problem solution is a set of points that we need to interpolate. This multivariate interpolation, however, is not easily solved, in the sense that we can either choose using a very naive estimator such as the K-nearest neighbors or just find another method that is not yet adopted for a wide range of applications.

(a) $\lambda = 0.1$

(b) $\lambda = 0.3$

(c) $\lambda = 1$

(d) $\lambda = 3$

(e) $\lambda = 10$

(f) $\lambda = 150$

Figure 2.1: Quantile estimations for a few different values of λ. The quantiles represented here are $\alpha = (5\%, 10\%, 25\%, 50\%, 75\%, 90\%, 95\%)$. When $\lambda = 0.1$, on the upper left, we clarly see a overfitting on the estimations. The other extreme case is also shown, when $\lambda = 200$ the nonparametric estimator converges to the linear model.

Figure 2.2: Estimated α-quantiles. On the left using a linear model and using a nonparametric approach (using $\lambda = 3$) on the right.

2.3 A comparison between both approaches

The last two sections introduced two different strategies to arrive in a Quantile Function $Q_{y_t|X}$. But what are the differences between using one method or the other?

To provide a comparison between both approaches, we estimate a quantile function to predict the one-step ahead quantile function. We use as explanatory variable only the last observation y_{t-1} - so $x_t = y_{t-1}$ - and estimate $\hat{q}_\alpha(y_{t-1})$, for every $\alpha \in \{0.05, 0.1, \ldots, 0.9, 0.95\}$. The result of both methods is shown on Figure 2.2.

While the linear model produces α-quantile functions which are linear by imposition, on the nonparametric model the α-quantiles are flexible enough to form a hull on the data and adapt to its nonlinearities. The difference between the estimated quantile functions $\hat{Q}_{y_t|y_{t-1}}$ on both methods are shown on Figure 2.2.

It is also important to test how the choice of the set A affects the estimated quantile function. We experimented with two different sizes of A. In one of them, a dense grid of probabilities is used: $A = \{0.005, 0.01, \ldots, 0.99, 0.995\}$, consisting of 199 elements. On the other only 19 elements are used to produce the quantile function ($A = \{0.05, 0.1, \ldots, 0.9, 0.95\}$).

Figure 2.3: 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.

Figure 2.4: Sensitivity to different choices of set A. On the left, we have the estimated quantiles for the linear model, while on the right for the nonparametric model. On both, the red line shows the quantile function estimated with $A = \{0.005, 0.01, \ldots, 0.99, 0.995\}$, consisting of 199 elements. The blue line is the estimated quantile function when $A = \{0.05, 0.1, \ldots, 0.9, 0.95\}$, consisting of only 19 elements.

3 Simulation

In this section, we investigate how to simulate future paths of the time series y_t. Let n be the total number of observations of y_t. We produce S different paths with size K for each. We have n observations of y_t and we want to produce . Given a vector of explanatory variables x_t, let $q_\alpha(x_t)$ be given by the α-quantile estimated as described on section 2.

The variables chosen to compose x_t can be either exogenous variables, autoregressive components of y_t or both. As the distribution of ε_t is unknown, we have to use a nonparametric approach in order to estimate its one-step ahead density.

The coefficients $\beta_{0\alpha}$ and β_α are the solution of the minimization problem given in the problem defined in (2.5)-(2.8), reproduced here for convenience:

$$\min_{q_\alpha, \varepsilon^+_{t,\alpha}, \varepsilon^-_{t,\alpha}} \quad \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon^+_{t,\alpha} + (1-\alpha)\varepsilon^-_{t,\alpha} \right) \tag{3.1}$$

$$\text{s.t.} \quad \varepsilon^+_{t,\alpha} - \varepsilon^-_{t,\alpha} = y_t - q_\alpha(x_t), \qquad \forall t \in T_\tau, \forall \alpha \in A, \tag{3.2}$$

$$\varepsilon^+_{t,\alpha}, \varepsilon^-_{t,\alpha} \geq 0, \qquad \forall t \in T_\tau, \forall \alpha \in A, \tag{3.3}$$

$$q_\alpha(x_t) \leq q_{\alpha'}(x_t), \qquad \forall t \in T_\tau, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \tag{3.4}$$

To produce S different paths of $\{\hat{y}_t\}_{t=n+1}^{n+K}$, we use the following procedure:

Procedure for simulating S scenarios of y_t

---

1. At first, let $\tau = n + 1$.

2. In any given period τ, for every $\alpha \in A$, we use the problem defined on (2.5)-(2.8) to estimate quantiles $q_\alpha(x_\tau)$. Note that x_τ is supposed to be known at time τ. In the presence of exogenous variables that are unknown, it is advisable to incorporate its uncertainty by considering different scenarios. In each scenario, though, x_τ must be considered fully known.

3. Let $\hat{Q}_{y_\tau | X = x_\tau}(\alpha, x_\tau)$ be the estimated quantile function of y_τ. To estimate $\hat{Q}_{y_\tau}$, we first define a discrete quantile function $\tilde{Q}_{y_\tau}$. By mapping every $\alpha \in A$ with its estimated quantile $\hat{q}_\alpha$, we define function $\tilde{Q}_{y_\tau}$. When we interpolate

   This process is described in more details on section 2.

   qualquer coisa

4. Once we have a distribution for y_{n+1}, we can generate S different simulated values, drawn from the distribution function $\hat{F}_{y_{n+1}} = \hat{Q}^{-1}_{y_\tau}$, derived from the quantile function found by doing steps 2 and 3. Let X be a random variable with uniform distribution over the interval $[0, 1]$. By using results from the Probability Integral Transform, we know that the random variable $F^{-1}_{y_{n+1}}(X)$ has the same distribution as y_{n+1}. So, by drawing a sample of size S from X and applying the quantile function $Q_{y_{n+1}}(\alpha)$, we have our sample of size K for y_{n+1}.

5. Each one of the S different values for y_{n+1} will be the starting point of a different path. Now, for each $\tau \in [n+2, n+K]$ and $s \in S$, we have to estimate quantiles $q_{\alpha\tau,s}$ and find a quantile function for $\hat{Q}_{y_{\tau,s}}$ just like it was done on steps 2 and 3. Note that when $\tau > n + 2$, every estimate will be scenario dependent, hence there will be S distribution functions estimated for each period τ. From now on, in each path just one new value will be drawn randomly from the one-step ahead distribution function - as opposed to what was carried on step 3, when S values were simulated. As there will be S distribution functions - one for each path, in each period τ it will be produced exact S values for y_τ, one for its own path. Repeating this step until all values of τ and s are simulated will give us the full simulations that we are looking for.

3.1 Simulating for a seasonal time series

Construir seção

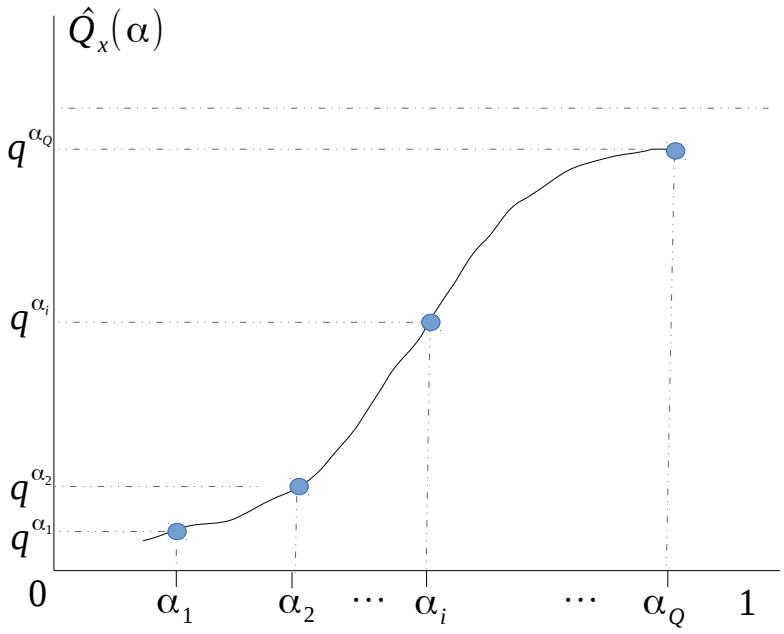

Figure 3.1: Fitting a distribution function from quantile estimations

References

[1] Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *arXiv preprint arXiv:1507.03133*, 2015.

[2] Seung-Jean Kim, Kwangmoo Koh, Stephen Boyd, and Dimitry Gorinevsky. ℓ_1 trend filtering. *SIAM review*, 51(2):339–360, 2009.

[3] Roger Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.