Pontíficia Universidade Católica do Rio de Janeiro
Departamento de Engenharia Elétrica

Quantile Regression

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

April 23, 2016

*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.

1 Introduction

Wind Firm Energy Certificate (FEC) estimation imposes several challenges. First, it is a quantile function of an aleatory quantity, named here on wind capacity factor (WP). Due to its non-dispachable profile, accurate scenario generation model could reproduce a fairly dependence structure in order to the estimation of FEC. Second, as it is a quantile functions, the more close to the extremes of the interval, the more sensitive to sampling error.

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. In all approaches we use the time series lags as the regression covariates. To study our methods performance, we use the mean power monthly data of Icaraizinho, a wind farm located in the Brazilian northeast.

The Icaraizinho dataset consists of monthly observations from 1981 to 2011 of mean power measured in Megawatts. As is common in renewable energy generation, there is a strong seasonality component. Figures 1.1 and 1.2 illustrate this seasonality, where we can observe low amounts of power generation for the months between February and May, and the yearly peek between August and November. Figure 1.3 shows four scatter plots with relations between y_t and some of its lags. We choose to present here the four lags that were selected for the quantile regression in the experiment of section 2.1, which are the 1^{st}, 4^{th}, 11^{th} and 12^{th}.

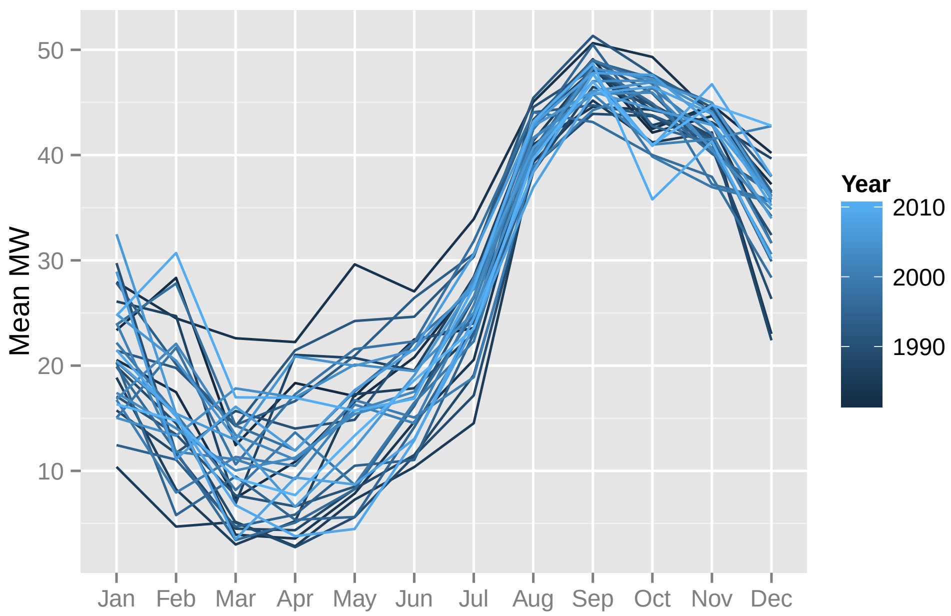


Figure 1.1: Icaraizinho yearly data. Each serie consists of monthly observations for each year.

Here we denote as parametric linear model the well-known quantile regression model [3]. 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, and is defined as

$$\begin{aligned} \mathcal{Q}_y(\alpha|x_1, x_2, ..., x_n) = \beta_0(\alpha) + \beta_1(\alpha)x_1 + \beta_2(\alpha)x_2 + ... \\ + \beta_n(\alpha)x_n + F_\varepsilon^{-1}(\alpha), \end{aligned} \tag{1.1}$$

where F_ε denotes the error density function. A Quantile Regression for the α-quantile is the solution of the following optimization problem:

$$\min_f \sum_{i=1}^{n} \alpha|y_t - f(t)|^+ + (1-\alpha)|y_t - f(t)|^-, \tag{1.2}$$

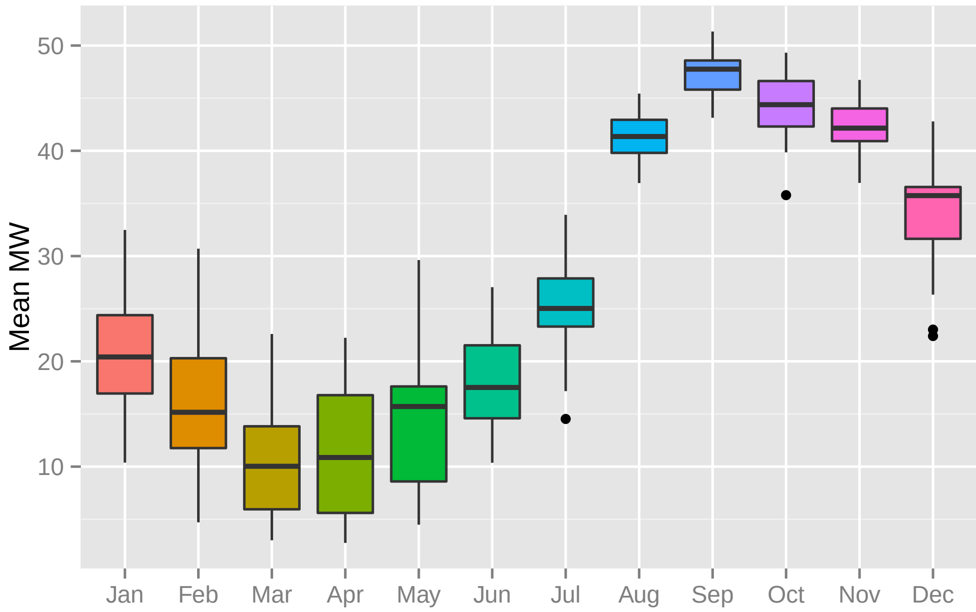


Figure 1.2: Boxplot for each month for the Icaraizinho dataset

where $f(t)$ is the estimated quantile value at a given time t and $|x|^+ = \max\{0, x\}$ and $|x|^- = -\min\{0, x\}$. To model this problem as a Linear Programming problem, thus being able to use a modern solver to fit our model, we can create variables δ_t^+ e δ_t^- to represent $|y - f(t)|^+$ and $|y - f(t)|^-$. So we have:

$$\begin{aligned} \min_f \ & \sum_{i=1}^{n} \left(\lambda\delta_i^+ + (1-\lambda)\delta_i^-\right) \\ \text{s.t.} \ & \delta_t^+ - \delta_t^- = y_t - f(x_t), && \forall t \in \{1, \dots, n\}, \\ & \delta_t^+, \delta_t^- \geq 0, && \forall t \in \{1, \dots, n\}. \end{aligned} \tag{1.3}$$

Section 2 is about linear models, so we investigate the quantile estimation when $f(t)$ is a linear function of the series past values, up to a maximum number of lags p:

$$Q_y(\alpha) = \beta_0(\alpha) + \beta_1(\alpha)y_{t-1} + \beta_2(\alpha)y_{t-2} + \cdots + \beta_p(\alpha)y_{t-p}. \tag{1.4}$$

In that section we investigate two ways of estimating coeficients, one based on Mixed Integer Programming ideas and the other based on the LASSO [5] penalty. Both of them are strategies to make regularization.

In section 3 we introduce a Nonparametric Quantile Autoregressive model with a ℓ_1-penalty term, in order to properly simulate FEC densities for several α-quantiles. In this nonparametric approach we don't assume any form for $f(t)$, 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.

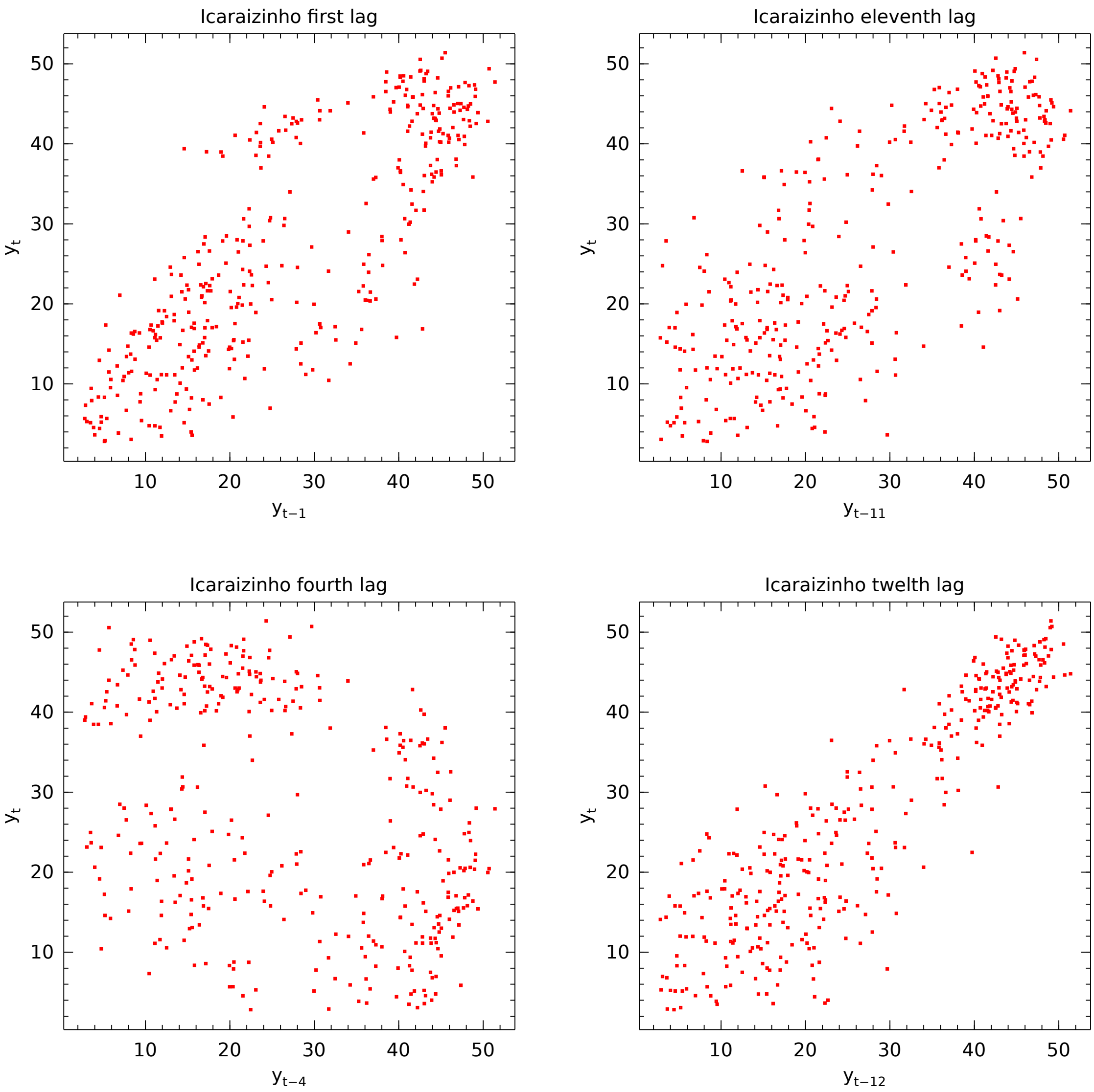


Figure 1.3: Relationship between y_t and some chosen lags.

2 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 Auto Regression. We won't be choosing the full model because this can lead to a bigger variance in our estimators, and this is often linked with bad performance in forecasting applications, that is our objective here. So our strategy will be using some sort of regularization method in order to improve performance. We investigate two ways of doing this. 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. This problem 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.

When we choose $f(t)$ to be a linear function like on equation 1.1, that we reproduce below:

$$Q_y(\alpha) = \beta_0(\alpha) + \beta_1(\alpha)y_{t-1} + \beta_2(\alpha)y_{t-2} + \cdots + \beta_p(\alpha)y_{t-p}.$$

substituting the linear model for $f(t)$ on problem 1.3 we get the following LP problem:

$$\begin{aligned} \min_{\beta_0,\beta} & \sum_{i=1}^{n} \left(\alpha\delta_i^+ + (1-\alpha)\delta_i^-\right) \\ \text{sujeito à} \ & \delta_i^+ - \delta_i^- = y_i - \beta_0 - \beta^T x_i, \qquad \forall i \in \{1,\dots,n\}, \\ & \delta_i^+, \delta_i^- \geq 0, \qquad \forall \{1,\dots,n\}. \end{aligned} \tag{2.1}$$

Now we are going to investigate two ways doing regularization for the above problem.

2.1 Best subset selection with Mixed Integer Optimization

In this part, we investigate the usage of Mixed Integer Optimization to select which variables are included in the model, up to a limit of inclusions imposed *a priori*. The optimization problem is described below:

$$\min_{\beta_0,\beta,z} \quad \textstyle\sum_{i=1}^{n} \left(\alpha\delta_i^+ + (1-\alpha)\delta_i^-\right) \tag{2.2}$$

$$\text{s.t} \quad \delta_i^+ - \delta_i^- = y_i - \beta_0 - \textstyle\sum_{p=1}^{P} \beta_p x_{i,p}, \qquad \forall i \in \{1,\dots,n\}, \tag{2.3}$$

$$\delta_i^+, \delta_i^- \geq 0, \qquad \forall i \in \{1,\dots,n\}, \tag{2.4}$$

$$-M_U z_p \leq \beta_p \leq M_U z_p, \qquad \forall p \in \{1,\dots,P\}, \tag{2.5}$$

$$\textstyle\sum_{p=1}^{P} z_p \leq K, \tag{2.6}$$

$$z_p \in \{0,1\}, \qquad \forall p \in \{1,\dots,P\}. \tag{2.7}$$

The objective function and constraints (2.4) and (2.5) are those from the standard linear quantile regression. The other constraints implement the process of regularization, forcing a maximum of K variables to be included. By (2.5), variable z_p is a binary that assumes 1 when the coefficient β_p is included. M_U is chosen in order to guarantee that $M_U \geq \|\hat{\beta}\|_\infty$. The solution given by β_0 and *beta* will be the best linear quantile regression with K nonzero coefficients.

We ran this optimization for each value of $K \in \{1,\dots,12\}$ and quantiles $\alpha \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$. We could see that for all quantiles the 12^{th} lag was the one included when $K = 1$. When $K = 2$, the 1^{st} lag was always included, sometimes with β_{12}, some others with β_4 and once with β_{11}. These 4 lags that were present until now are the only ones selected when $K = 3$. For $K = 4$, those same four lags were selected for three quantiles (0.05, 0.1 and 0.5), but for the others (0.9 and 0.95) we have β_6, β_7 and β_9 also as selected. From now on, the inclusion of more lags represent a lower increase in the fit of the quantile regression. The estimated coefficient values for all K's are available in the appendices section.

2.2 Best subset selection with a ℓ_1 penalty

Another way of doing regularization is including the ℓ_1-norm of the coefficients on the objective function. The advantage of this method is that coefficients are shrunk towards zero, and only some of them will have nonzero coefficients. By lowering the penalty we impose on the ℓ_1-norm, more variables are being added to the model. This is the same strategy of the LASSO, and its usage for the quantile regression is discussed in [4]. The proposed optimization problem to be solved is:

$$\min_{\beta_0,\beta} \sum_{i=1}^{n} \alpha|y_t - f(t)|^+ + (1-\alpha)|y_t - f(t)|^- + \lambda\|\beta\|_1 \tag{2.8}$$

$$f(t) = \beta_0 - \sum_{p=1}^{P} \beta_p x_{i,p},$$

where the regressors $x_{i,p}$ used are its lags. In order to represent the above problem to be solved with linear programming solver, we restructure the problem as below:

$$\begin{aligned}
\min_{\beta_0,\beta} \quad & \textstyle\sum_{i=1}^{n} \left(\alpha\delta_i^+ + (1-\alpha)\delta_i^-\right) + \lambda \sum_{p=1}^{P} \xi_p && (2.9)\\
\text{subject to} \quad & \delta_i^+ - \delta_i^- = y_i - \beta_0 - \textstyle\sum_{p=1}^{P} \beta_p x_{i,p}, \qquad \forall i \in \{1,\dots,n\}, && (2.10)\\
& \delta_i^+, \delta_i^- \geq 0, \qquad \forall i \in \{1,\dots,n\}, && (2.11)\\
& \xi_p \geq \beta_p, \qquad \forall p \in \{1,\dots,P\}, && (2.12)\\
& \xi_p \geq -\beta_p, \qquad \forall p \in \{1,\dots,P\}, && (2.13)
\end{aligned}$$

Once again, this model is built upon the standard linear programming model for the quantile regression (equation 2.1). On the above formulation, the ℓ_1 norm of equation (2.8) is substituted by the sum of ξ_p, which represents the absolute value of β_p. The link between variables ξ_p and β_p is made by constraints (2.12) and (2.13). Note that the linear coefficient β_0 is not included in the penalization, as the sum of penalties on the objective function 2.9.

For low values of λ the penalty is small and thus we have models with all nonzero coefficients. On the other hand, when λ is bigger all coefficients are zero and we have a constant model. For instance, we don't penalize the linear coefficient β_0. For the same quantiles values α we experimented on section 2.1 ($\alpha \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$).

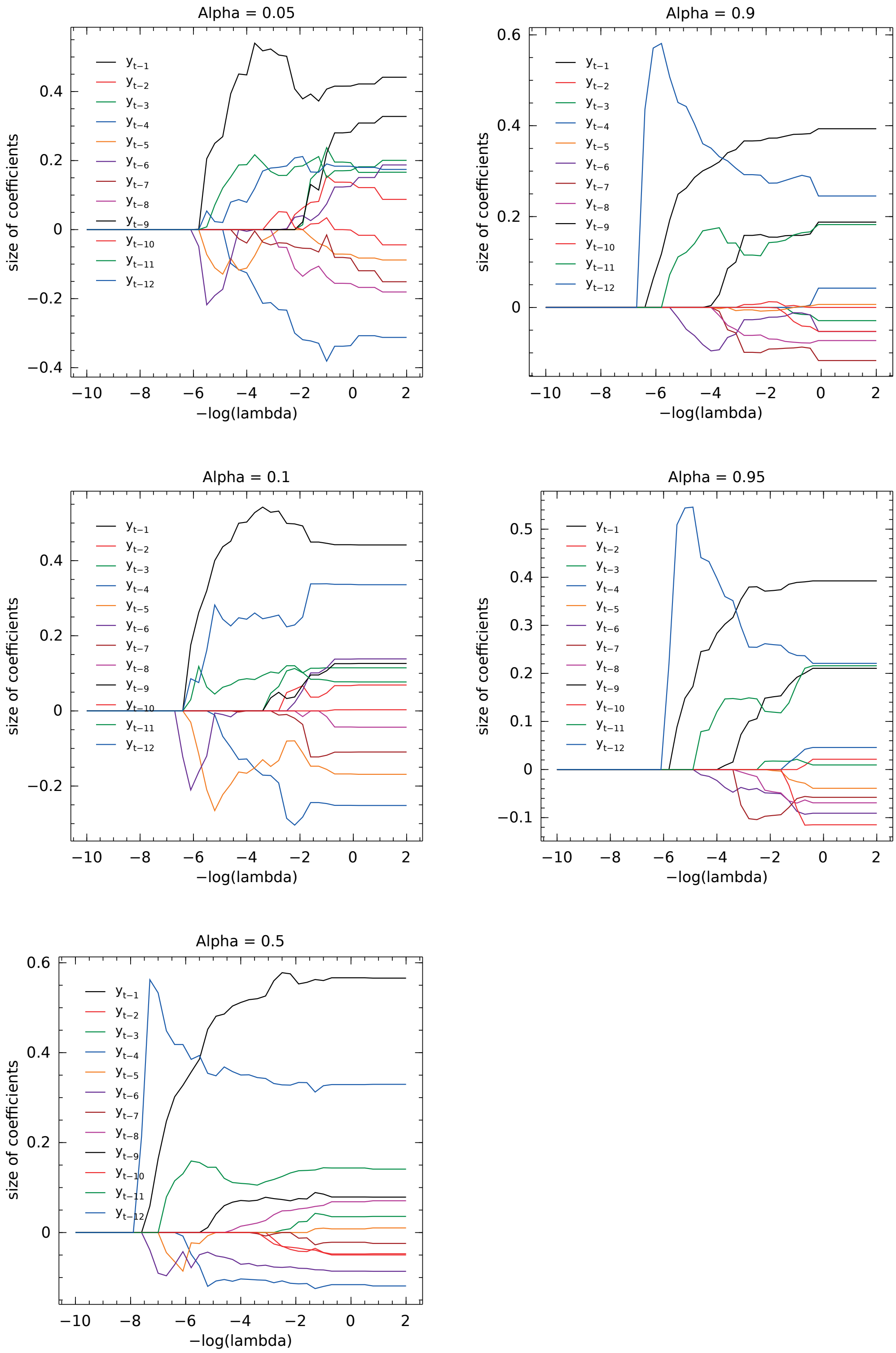


Figure 2.1: Coefficients path for a few different values of α-quantiles. λ is presented in a negative log scale, to make visualization easier.

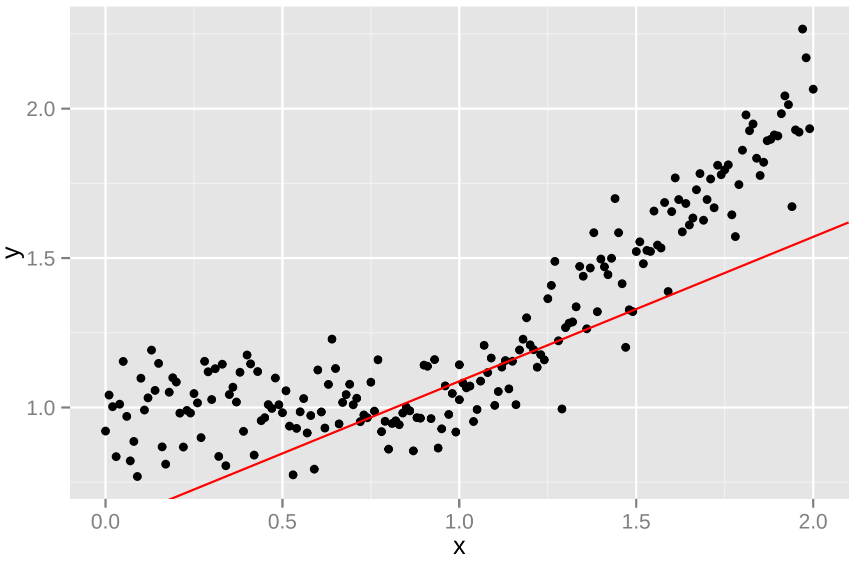


Figure 3.1: Example of data where nonlinearity is present and a linear quantile estimator is employed

3 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 (we illustrate this in figure 3.1). To prevent this issue from occurring we propose a modification which we let the prediction $\mathcal{Q}_{y_t|y_{t-1}}(\alpha)$ 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].

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 $I = \{2, \ldots, n-p-1\}$. 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:

$$\begin{aligned}\mathcal{Q}^{\alpha}_{y_t|y_{t-1}}(i) = \arg\min_{q_i} \sum_{i \in I} \left(|y_i - q_i|^{+}\alpha + |y_i - q_i|^{-}(1-\alpha) \right) \\ + \lambda \sum_{i \in I} |D^2 q_i|,\end{aligned} \tag{3.1}$$

where $D^2 q_t$ is the second derivative of the q_t function, calculated as follows:

$$D^2 q_i = \left(\frac{q_{i+1} - q_i}{x_{i+1} - x_i} \right) - \left(\frac{q_i - q_{i-1}}{x_i - x_{i-1}} \right).$$

The first part on the objective function is the usual quantile regression condition for $\{q_i\}$. The second part is the ℓ_1-filter. The purpose of a filter is to control the amount of variation for our estimator q_i. When no penalty is employed we would always get $q_i = y_i$. 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:

$$\min_{q_i} \quad \sum_{i=1}^{n} \left(\alpha\delta_i^+ + (1-\alpha)\delta_i^-\right) + \lambda \sum_{i=1}^{n} \xi_i \tag{3.2}$$

$$s.t. \qquad \delta_i^+ - \delta_i^- = y_i - q_i, \qquad \forall i \in \{3,\ldots,n-1\}, \tag{3.3}$$

$$D_i = \left(\frac{q_{i+1}-q_i}{x_{i+1}-x_i}\right) - \left(\frac{q_i-q_{i-1}}{x_i-x_{i-1}}\right) \qquad \forall i \in \{3,\ldots,n-1\}, \tag{3.4}$$

$$\xi_i \geq D_i, \qquad \forall i \in \{3,\ldots,n-1\}, \tag{3.5}$$

$$\xi_i \geq -D_i, \qquad \forall i \in \{3,\ldots,n-1\}, \tag{3.6}$$

$$\delta_i^+, \delta_i^-, \xi_i \geq 0, \qquad \forall i \in \{3,\ldots,n-1\}. \tag{3.7}$$

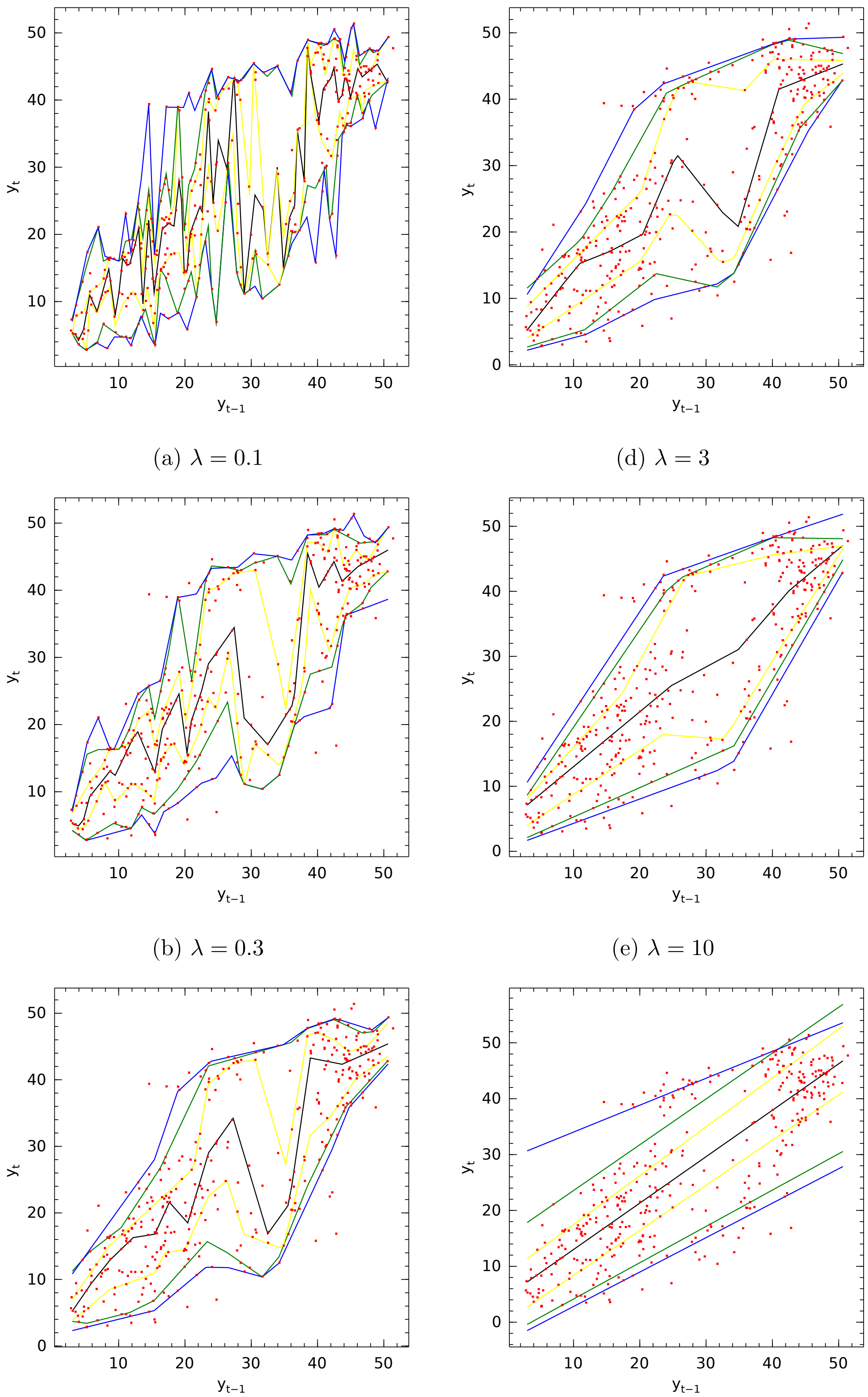
The output of our optimization problem is a sequence of ordered points $\{(x_i, q_i)\}_{i\in I}$. The next step is to interpolate these points in order to provide an estimation for any other value of x. To address this issue, we propose using a B-splines interpolation, that will be developed in another study.

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.

Figure 3.2 shows the quantile estimation for a few different values of λ.

When estimating quantiles for a few different values of α, however, sometimes we find them overlapping each other, which we call crossing quantiles. This effect can be seen in figure 3.2f, where the 95%-quantile crosses over the 90%-quantile. To prevent this, one can include a non-crossing constraint:

$$q_i^{\alpha} \leq q_i^{\alpha'}, \quad \forall i \in I, \alpha < \alpha'. \tag{3.8}$$



(a) $\lambda = 0.1$ (d) $\lambda = 3$

(b) $\lambda = 0.3$ (e) $\lambda = 10$

(c) $\lambda = 1$ (f) $\lambda = 200$

Figure 3.2: 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.

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.

[4] Youjuan Li and Ji Zhu. L1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 2012.

[5] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

4 Appendices

	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
β_0	-15.33	9.38	1.48	1.34	8.72	-1.68	4.94	0.65	-0.27	-0.16	-3.96	-2.55
β_1	-0.00	0.79	0.66	0.58	0.46	0.40	0.48	0.46	0.46	0.47	0.42	0.44
β_2	-0.00	-0.00	-0.00	-0.00	-0.00	0.33	-0.00	-0.00	-0.00	-0.00	0.14	0.09
β_3	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.20	0.20	0.19	0.20	0.17
β_4	-0.00	-0.47	-0.28	-0.27	-0.29	-0.35	-0.31	-0.40	-0.35	-0.35	-0.34	-0.31
β_5	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.05	-0.07	-0.09
β_6	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.11	0.08	0.11	0.17	0.12	0.19
β_7	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.16	-0.15	-0.08	-0.15
β_8	-0.00	-0.00	-0.00	-0.00	-0.15	-0.00	-0.31	-0.26	-0.17	-0.17	-0.16	-0.18
β_9	-0.00	-0.00	-0.00	-0.00	-0.00	0.14	0.16	0.20	0.26	0.23	0.28	0.33
β_{10}	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.04
β_{11}	-0.00	-0.00	0.26	0.17	0.21	0.08	0.16	0.19	0.17	0.18	0.17	0.20
β_{12}	1.17	-0.00	-0.00	0.18	0.15	0.19	0.22	0.20	0.20	0.18	0.18	0.17

Table 4.1: $\alpha = 0.05$. Coefficients β_i for each value of K, where K is the number of nonzero coefficients, excluding the intercept β_0, which is always included.

	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
β_0	-10.68	10.07	3.56	1.24	0.76	3.01	3.33	3.02	1.05	2.26	1.55	1.57
β_1	-0.00	0.81	0.63	0.61	0.55	0.49	0.49	0.50	0.48	0.44	0.44	0.44
β_2	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.04	-0.00	-0.00	0.04	0.07	0.07
β_3	-0.00	-0.00	-0.00	-0.00	0.15	0.20	0.16	0.15	0.13	0.11	0.12	0.12
β_4	-0.00	-0.43	-0.33	-0.28	-0.37	-0.33	-0.34	-0.30	-0.24	-0.24	-0.26	-0.25
β_5	-0.00	-0.00	-0.00	-0.00	-0.00	-0.08	-0.07	-0.12	-0.14	-0.15	-0.17	-0.17
β_6	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.11	0.10	0.10	0.14	0.14
β_7	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.07	-0.11	-0.13	-0.11	-0.11
β_8	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.04	-0.04
β_9	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.09	0.10	0.13	0.13
β_{10}	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00
β_{11}	-0.00	-0.00	-0.00	0.14	0.17	0.17	0.16	0.15	0.11	0.09	0.08	0.08
β_{12}	1.09	-0.00	0.35	0.27	0.25	0.22	0.22	0.26	0.33	0.34	0.33	0.33

Table 4.2: $\alpha = 0.1$. Coefficients β_i for each value of K, where K is the number of nonzero coefficients, excluding the intercept β_0, which is always included.

	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
β_0	2.72	-3.38	8.64	4.88	0.62	2.98	2.70	2.62	2.27	1.87	2.43	2.53
β_1	-0.00	0.59	0.52	0.51	0.57	0.54	0.56	0.56	0.58	0.58	0.57	0.57
β_2	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.03	-0.06	-0.05	-0.05
β_3	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.04	0.03	0.04
β_4	-0.00	-0.00	-0.25	-0.18	-0.14	-0.11	-0.11	-0.12	-0.11	-0.11	-0.11	-0.12
β_5	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.01
β_6	-0.00	-0.00	-0.00	-0.00	-0.00	-0.06	-0.09	-0.08	-0.08	-0.08	-0.09	-0.09
β_7	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.02	-0.02
β_8	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.06	0.06	0.05	0.06	0.08	0.07
β_9	-0.00	-0.00	-0.00	-0.00	0.08	0.09	0.06	0.09	0.07	0.07	0.08	0.08
β_{10}	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.05	-0.04	-0.05	-0.05	-0.05
β_{11}	-0.00	0.54	-0.00	0.15	0.14	0.11	0.10	0.11	0.14	0.14	0.15	0.14
β_{12}	0.92	-0.00	0.42	0.34	0.32	0.33	0.32	0.34	0.33	0.34	0.32	0.33

Table 4.3: $\alpha = 0.5$. Coefficients β_i for each value of K, where K is the number of nonzero coefficients, excluding the intercept β_0, which is always included.

	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
β_0	12.14	10.06	6.60	11.05	13.22	12.04	13.34	13.28	12.58	13.69	13.47	13.71
β_1	-0.00	0.24	0.39	0.39	0.40	0.38	0.38	0.38	0.38	0.40	0.40	0.40
β_2	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.02
β_3	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01	-0.04	-0.03	-0.02
β_4	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.03	-0.00	0.05	0.05	0.04
β_5	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.01
β_6	-0.00	-0.00	-0.00	-0.14	-0.00	-0.00	-0.03	-0.05	-0.01	-0.07	-0.07	-0.07
β_7	-0.00	-0.00	-0.00	-0.00	-0.19	-0.10	-0.10	-0.11	-0.09	-0.11	-0.11	-0.10
β_8	-0.00	-0.00	-0.00	-0.00	-0.00	-0.08	-0.07	-0.08	-0.08	-0.07	-0.07	-0.08
β_9	-0.00	-0.00	-0.00	0.14	0.16	0.15	0.16	0.18	0.16	0.19	0.19	0.19
β_{10}	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.04	-0.06	-0.06	-0.06
β_{11}	-0.00	-0.00	0.20	-0.00	0.11	0.15	0.12	0.16	0.16	0.18	0.18	0.19
β_{12}	0.80	0.63	0.39	0.42	0.26	0.29	0.28	0.23	0.29	0.24	0.24	0.25

Table 4.4: $\alpha = 0.9$. Coefficients β_i for each value of K, where K is the number of nonzero coefficients, excluding the intercept β_0, which is always included.

	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
β_0	16.73	11.74	11.51	13.77	13.45	13.48	14.36	14.84	12.36	14.04	13.09	14.00
β_1	-0.00	0.26	0.32	0.35	0.38	0.38	0.40	0.43	0.40	0.40	0.39	0.39
β_2	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.02	0.02
β_3	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.01
β_4	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.04	0.06	0.06	0.05
β_5	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.04	-0.03	-0.04
β_6	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.05	-0.10	-0.07	-0.09	-0.08	-0.09
β_7	-0.00	-0.00	-0.00	-0.15	-0.14	-0.12	-0.09	-0.05	-0.06	-0.06	-0.06	-0.06
β_8	-0.00	-0.00	-0.00	-0.00	-0.00	-0.04	-0.05	-0.07	-0.05	-0.08	-0.07	-0.07
β_9	-0.00	-0.00	-0.00	0.16	0.11	0.14	0.16	0.19	0.19	0.22	0.22	0.21
β_{10}	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.15	-0.14	-0.11	-0.12	-0.11
β_{11}	-0.00	-0.00	0.17	-0.00	0.14	0.13	0.12	0.25	0.23	0.18	0.21	0.22
β_{12}	0.71	0.59	0.37	0.41	0.28	0.28	0.25	0.21	0.27	0.25	0.24	0.22

Table 4.5: $\alpha = 0.9$. Coefficients β_i for each value of K, where K is the number of nonzero coefficients, excluding the intercept β_0, which is always included.