Pontíficia Universidade Católica do Rio de Janeiro
Departamento de Engenharia Elétrica

Quantile Regression

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

May 13, 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) [7] 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 models 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. The full Icaraizinho serie can be found on the appendices from this article. 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 time span between February and May, and a yearly peek between August and November. Figure 1.3 shows four scatter plots relating y_t with 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}. They are most likely the four main lags to use for these analysis.

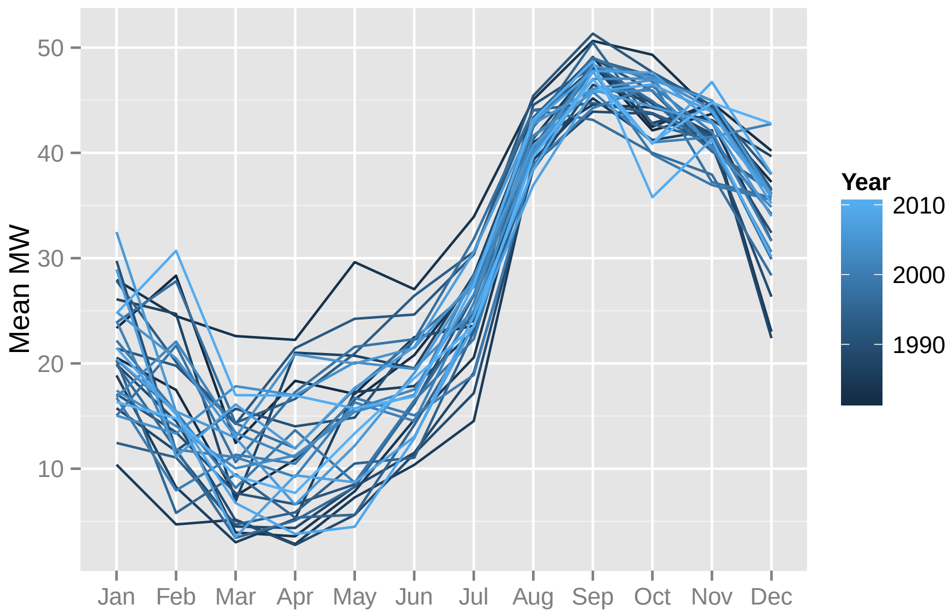


Figure 1.1: Icaraizinho yearly data. Each serie consists of monthly observations for each year.

Here we denote as parametric linear model the well-known quantile regression model [4]. 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. A Quantile Regression for the α-quantile is the solution of the following optimization problem:

$$\min_f \sum_{t=1}^{n} \alpha|y_t - f(x_t)|^+ + (1-\alpha)|y_t - f(x_t)|^-, \tag{1.1}$$

where $f(x_t)$ 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

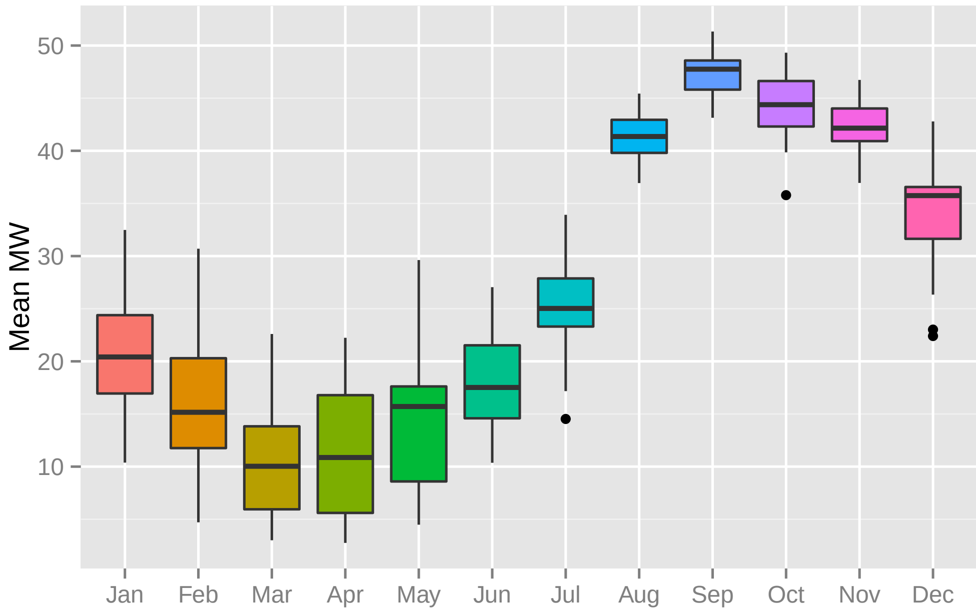


Figure 1.2: Boxplot for each month for the Icaraizinho dataset

modern solver to fit our model, we can create variables ε_t^+ e ε_t^- to represent $|y - f(x_t)|^+$ and $|y - f(x_t)|^-$, respectively. So we have:

$$\begin{aligned} \min_{f,\varepsilon_t^+,\varepsilon_t^-} & \sum_{t=1}^{n} \left(\lambda \varepsilon_t^+ + (1-\lambda)\varepsilon_t^-\right) \\ \text{s.t. } & \varepsilon_t^+ - \varepsilon_t^- = y_t - f(x_t), & \forall t \in \{1,\dots,n\}, \\ & \varepsilon_t^+, \varepsilon_t^- \geq 0, & \forall t \in \{1,\dots,n\}. \end{aligned} \tag{1.2}$$

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

$$f(x_t, \alpha; \beta) = \beta_0(\alpha) + \beta_1(\alpha) y_{t-1} + \beta_2(\alpha) y_{t-2} + \cdots + \beta_p(\alpha) y_{t-p}. \tag{1.3}$$

In that section we investigate two ways of estimating coeficients, one based on Mixed Integer Programming ideas and the other based on the LASSO [8] 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 $f(x_t)$, 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.

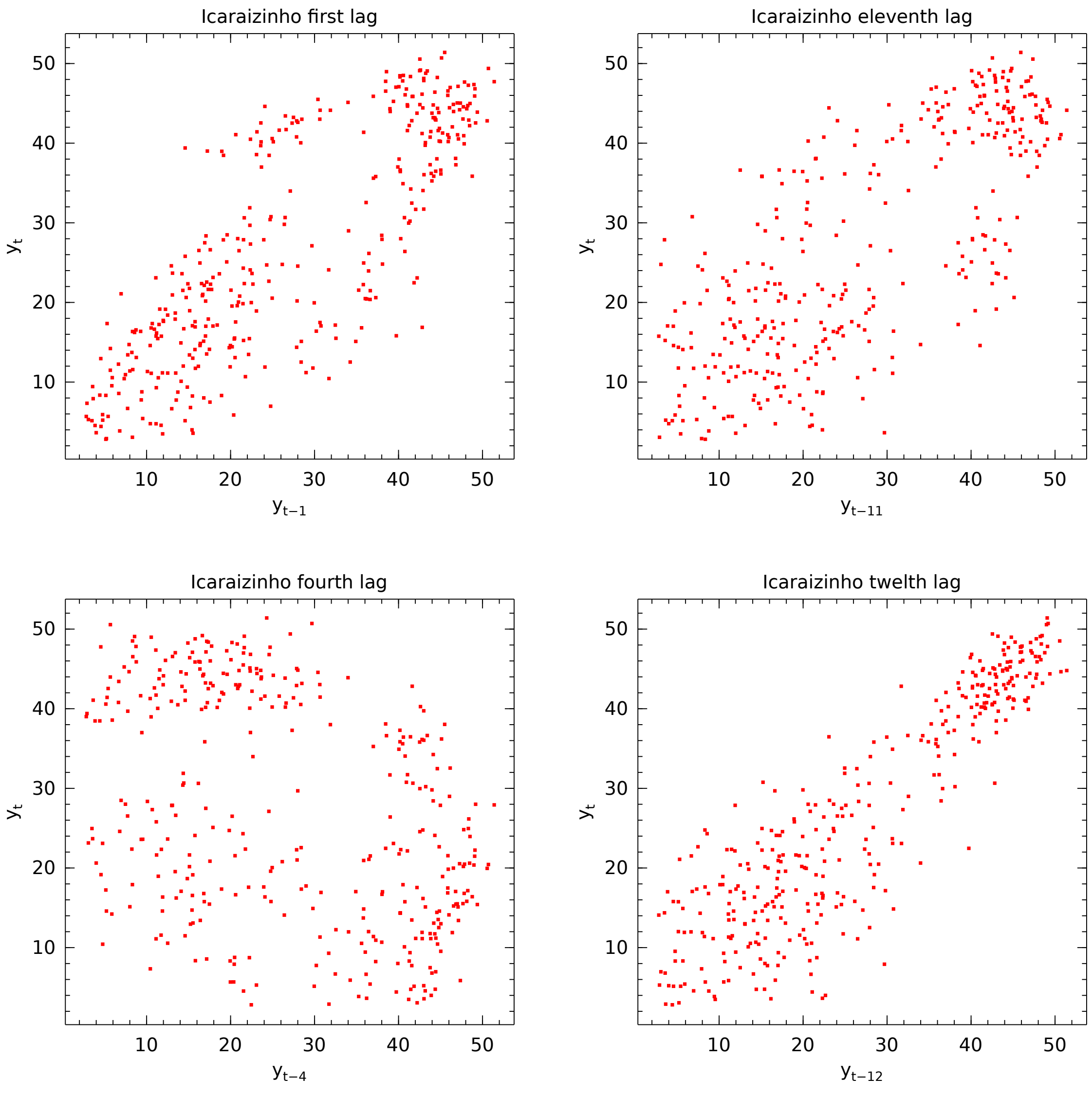


Figure 1.3: Relationship between y_t and some chosen lags.

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 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 [2]. The second way is including a ℓ_1 penalty on the linear quantile regression, as in [3], 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 $f(x_t)$ to be a linear function, as on equation 1.1 (that we reproduce below for convenience):

$$\min_{f} \sum_{t=1}^{n} \alpha|y_t - f(x_t)|^+ + (1-\alpha)|y_t - f(x_t)|^-, \tag{2.1}$$

we can substitute it on problem 1.2, getting the following LP problem:

$$\begin{aligned} \min_{\beta_0,\beta,\varepsilon_t^+,\varepsilon_t^-} & \sum_{t=1}^{n} \left(\alpha\varepsilon_t^+ + (1-\alpha)\varepsilon_t^-\right) \\ \text{s.t. } & \varepsilon_t^+ - \varepsilon_t^- = y_t - \beta_0 - \beta^T x_t, \qquad \forall t \in \{1,\ldots,n\}, \\ & \varepsilon_t^+, \varepsilon_t^- \geq 0, \qquad \forall t \in \{1,\ldots,n\}. \end{aligned} \tag{2.2}$$

In this work, we didn't explore the addition of terms other than the terms y_t past lags. For example, we could include functions of y_{t-p}, such as $log(y_{t-p})$ or $exp(y_{t-p})$. We leave such inclusion for further works.

2.1 Best subset selection with Mixed Integer Programming

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

$$\min_{\beta_0,\beta,z,\varepsilon_t^+,\varepsilon_t^-} \quad \textstyle\sum_{t=1}^{n} \left(\alpha\varepsilon_t^+ + (1-\alpha)\varepsilon_t^-\right) \tag{2.3}$$

$$\text{s.t} \quad \varepsilon_t^+ - \varepsilon_t^- = y_t - \beta_0 - \textstyle\sum_{p=1}^{P} \beta_p x_{t,p}, \qquad \forall t \in \{1,\ldots,n\}, \tag{2.4}$$

$$\varepsilon_t^+, \varepsilon_t^- \geq 0, \qquad \forall t \in \{1,\ldots,n\}, \tag{2.5}$$

$$-M_U z_p \leq \beta_p \leq M_U z_p, \qquad \forall p \in \{1,\ldots,P\}, \tag{2.6}$$

$$\textstyle\sum_{p=1}^{P} z_p \leq K, \tag{2.7}$$

$$z_p \in \{0,1\}, \qquad \forall p \in \{1,\ldots,P\}. \tag{2.8}$$

The objective function and constraints (2.5) and (2.6) 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 (2.6), 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 β will be the best linear quantile regression with K nonzero coefficients.

We ran this optimization for each value of $K \in \{1,\ldots,12\}$ 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

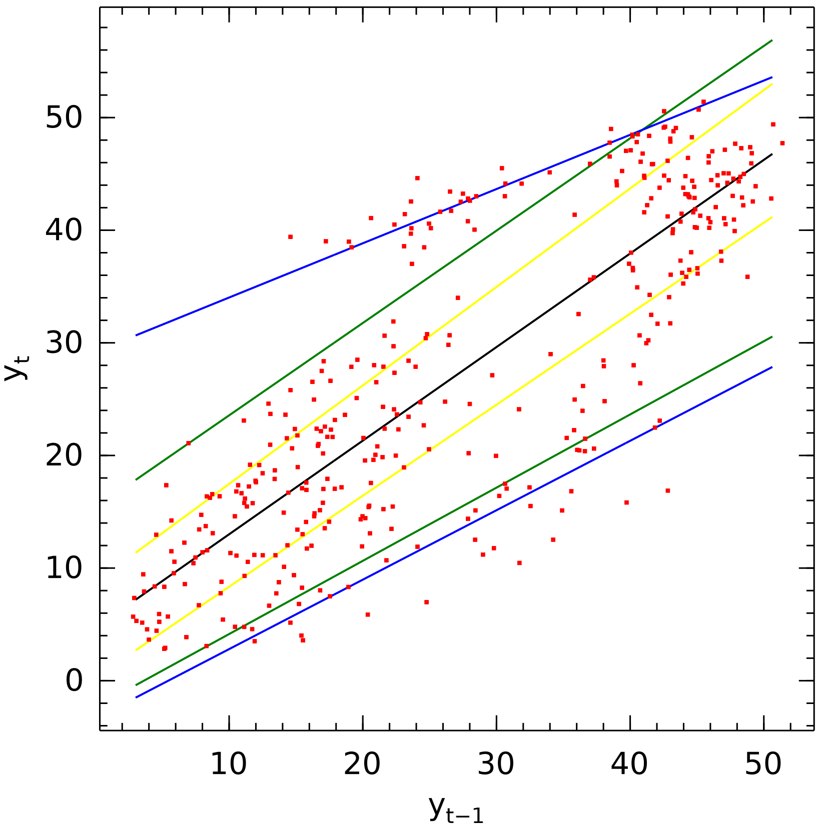


Figure 2.1: Linear Quantile Regression when only y_{t-1} is used as explanatory variable

of the quantile regression. The estimated coefficient values for all K's are available in the appendices section. Figure 2.1 shows a linear estimator for the quantiles $(0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95)$.

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 [5]. The proposed optimization problem to be solved is:

$$\min_{\beta_0,\beta} \sum_{t=1}^{n} \alpha|y_t - f(x_t)|^+ + (1-\alpha)|y_t - f(x_t)|^- + \lambda\|\beta\|_1 \tag{2.9}$$

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

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

$$\beta_\lambda^{*LASSO} = \arg\min_{\beta_0,\beta} \quad \textstyle\sum_{i=1}^{n} \left(\alpha\varepsilon_t^+ + (1-\alpha)\varepsilon_t^-\right) + \lambda \sum_{p=1}^{P} \xi_p \tag{2.10}$$

$$\text{s.t.} \quad \varepsilon_t^+ - \varepsilon_t^- = y_t - \beta_0 - \textstyle\sum_{p=1}^{P} \beta_p x_{t,p}, \qquad \forall t \in \{1,\dots,n\}, \tag{2.11}$$

$$\varepsilon_t^+, \varepsilon_t^- \geq 0, \qquad \forall t \in \{1,\dots,n\}, \tag{2.12}$$

$$\xi_p \geq \beta_p, \qquad \forall p \in \{1,\dots,P\}, \tag{2.13}$$

$$\xi_p \geq -\beta_p, \qquad \forall p \in \{1,\dots,P\}, \tag{2.14}$$

Once again, this model is built upon the standard linear programming model for the quantile regression (equation 2.2). On the above formulation, the ℓ_1 norm of equation (2.9) 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 (2.13) and (2.14). Note that the linear coefficient β_0 is not included in the penalization, as the sum of penalties on the objective function 2.10.

For such estimation to produce good results, however, each variable must have the same relative weight in comparison with one another. So, before solving the optimization problem, we normalize all variables to have mean $\mu = 0$ and variance $\sigma^2 = 1$. For the vector of observations for each covariate (that in our problem represents is a vector of observations of lags y_{t-p}), we apply the transformation $\tilde{y}_{t-p,i} = (y_{t-p,i} - \bar{y}_{t-p})/\sigma_{t-p}$, where $\bar{y}_{t-p}$ is the p-lag mean and σ_{t-p} the p-lag standard deviation. We use the $\tilde{y}_{t-p,i}$ series to estimate the coefficients. Once done that, we multiply each coefficient for its standard deviation to get the correct coefficient: $\beta_i = \tilde{\beta}_i \dot{\sigma}_{t-p}$.

For low values of λ, the penalty is small and thus we have a model where all coefficients have a nonzero value. On the other hand, while λ is increased the coefficients shrink towards zero; in the limit 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 ($\alpha \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$).

It is important to mention that even though we have coefficients that are estimated by this method, we don't use them directly. Instead, the nonzero coefficients will be the only covariates used as explanatory variables of a regular quantile autoregression, solved by the linear programming problem 2.2. In summary, the optimization in equation 2.9 acts as a variable selection for the subsequent estimation, which is normally called the post-lasso estimation [1].

We are interested, finally, in finding the post-lasso coefficients β^*_λ, which is the solution of the optimization problem given below:

$$
\begin{aligned}
\beta^*_\lambda = \underset{\beta_0, \beta, \varepsilon_t^+, \varepsilon_t^-}{\arg\min} & \sum_{t=1}^{n} \left(\alpha \varepsilon_t^+ + (1-\alpha) \varepsilon_t^- \right) \\
\text{s.t. } & \varepsilon_t^+ - \varepsilon_t^- = y_t - \beta_0 - \sum_{p \in L_\lambda} \beta_p x_{t,p}, \qquad \forall t \in \{1, \dots, n\}, \\
& \varepsilon_t^+, \varepsilon_t^- \geq 0, \qquad \forall t \in \{1, \dots, n\}.
\end{aligned} \tag{2.15}
$$

Note that only a subset of the P covariates will have nonzero values, which are given by the set

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

Hence, we have that

$$\beta^{*LASSO}_{\lambda,p} = 0 \iff \beta^*_{\lambda,p} = 0.$$

2.3 Model selection

On sections 2.1 and 2.2, we presented ways of doing regularization. But regularization can be done with different levels of parsimony. For example, we can select a different number K of variables to be included in the best subset selection via MIP or choose different values of λ for the ℓ_1 penalty. Each of these choices leads to a different model, so one needs to know how to select the best one among the options we have. One way of achieving this is by using an information criteria to guide our decision.

An information criteria summarizes two aspects. One of them refers to how well the model fits the in-sample observations. The other part penalizes the quantity of covariates used in the model. By penalizing how big our model is, we prevent overfitting from happening. So, in order to enter the model, the covariate must supply enough goodness of fit. In [6], it is presented a variation of the Schwarz criteria for M-estimators that includes quantile regression. The Schwarz Information Criteria (SIC) adapted to the quantile autoregression case is presented below:

$$SIC(j) = n \log(\hat{\sigma}_j) + \frac{1}{2} p_j \log n, \tag{2.16}$$

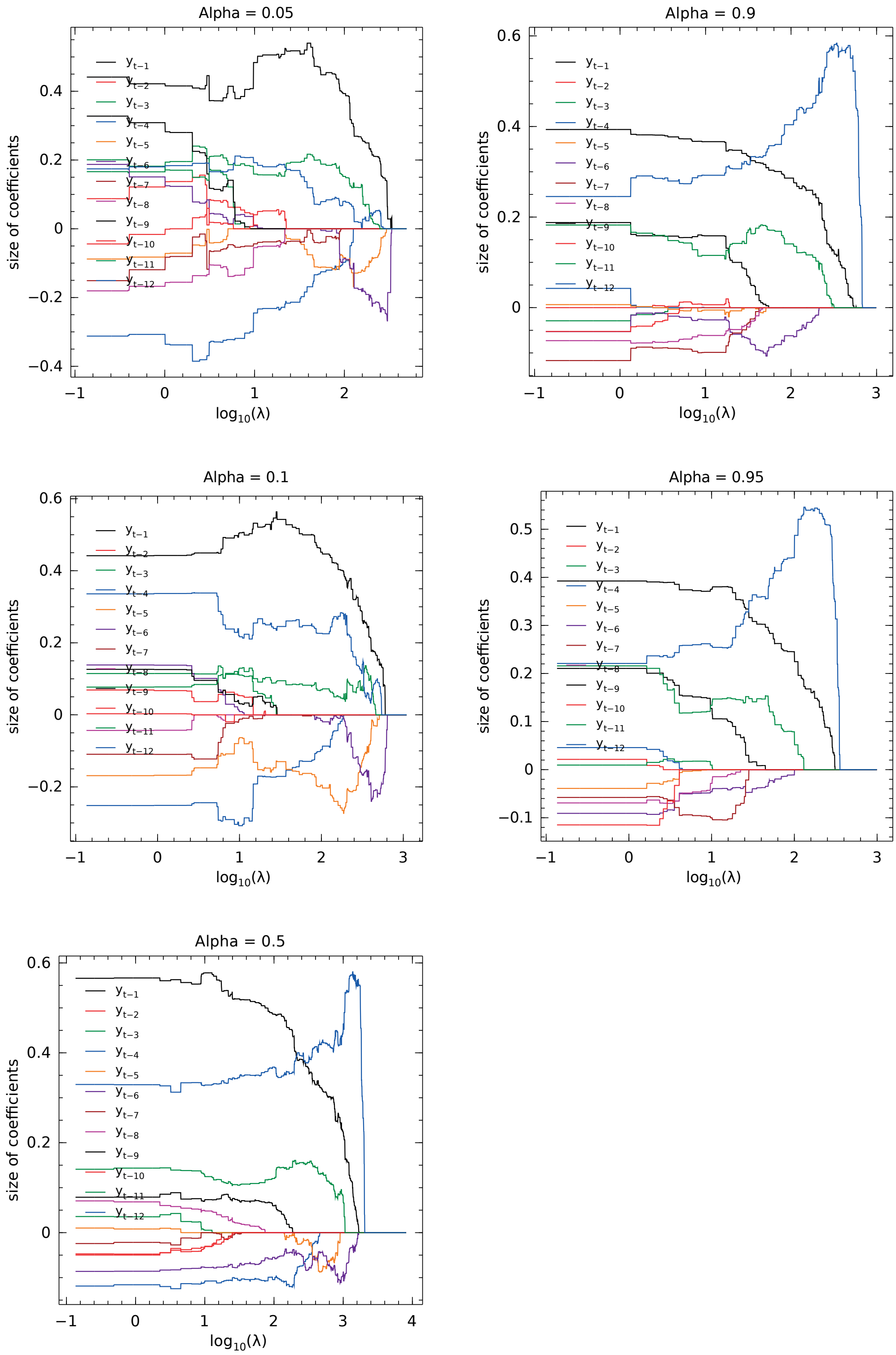


Figure 2.2: Coefficients path for a few different values of α-quantiles. λ is presented in a $\log_{10}$ scale, to make visualization easier.

where $\hat{\sigma}_j = \frac{1}{n}\sum_{t=1}^{n} \rho_\alpha(y_t - x_t^T \hat{\beta}_n(\alpha))$, $\rho_\alpha(\cdot)$ is the penalization function and p_j the j^{th} model's dimension. This procedure leads to a consistent model selection if the model is well specified.

Optimizing a LP problem is many times faster than a similar-sized MIP problem. One of our goals is to test whether a solution of a model with a ℓ_1-norm can approximate well a solution given by the MIP problem. We propose an experiment that is described as follows. First, we calculate the quantity $k(\lambda)$ of nonzero coefficients, for each given lambda:

$$k_\lambda = \|\beta_\lambda^*\|_0. \tag{2.17}$$

Then, for each number K of total nonzero coefficients (from 1 until 13, where 1 means that only the intercept is included), there will be a penalty λ_K^* which minimizes the SIC:

$$\lambda_k^* = \arg\min_{\lambda} \left\{ SIC\left(k_\lambda\right) \mid k_\lambda = K \right\}. \tag{2.18}$$

Thus, we can compare the SIC of the best lasso fit where exactly K variables are selected with the SIC selected by the MIP problem, also with K variables selected.

To help us view the difference of results between both methods, we define a distance metric d between the subset of coefficients chosen by each one of them. Let

$$d(\beta_{MIP(K)}^*, \beta_{\lambda_K^*}^*) = \frac{1}{2K}\sum_{p=1}^{P}\left| I(\beta_{MIP(K),p}^*) - I(\beta_{\lambda_K^*,p}^*)\right|, \tag{2.19}$$

where I is an indicator function such that $I(x) = 0$ if $x = 0$ and $I(x) = 1$ otherwise.

Figure 2.3 shows the results of these experiments for quantiles $\alpha \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$. The results point us that for small values of K the distance between coefficients is bigger and where we observe the biggest differences between the SIC values. The minimum SIC value for the MIP problem is usually found between 4 and 6 variables in the model.

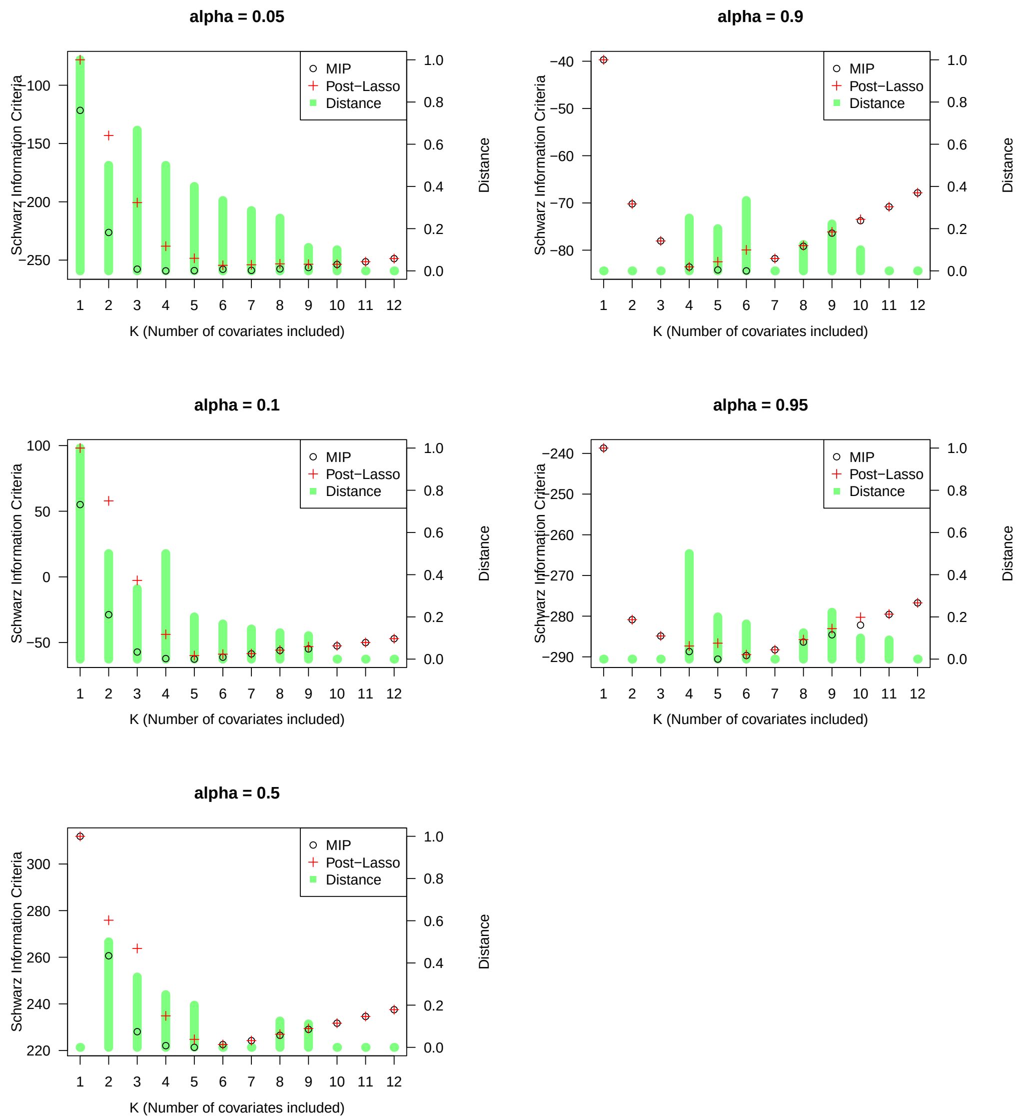


Figure 2.3: Comparison of SIC between a solution with Lasso as a variable selector and the best subset selection with MIP. The bars represent the distance d as defined by equation 2.19. (*) When the distance is zero, it means that the same variables are selected from both methods for a given k. Thus, in these cases we have the same SIC for both of them.

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 (for example, scatter plot of y_t versus y_{t-1} that is seen on the upper left of figure 1.3). 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 [3].

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

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

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

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

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

$$\min_{q_t} \quad \textstyle\sum_{t=1}^n \left(\alpha\delta_t^+ + (1-\alpha)\delta_t^-\right) + \lambda \sum_{t=1}^n \xi_t \tag{3.2}$$

$$s.t. \quad \delta_t^+ - \delta_t^- = y_t - q_t, \qquad \forall t \in \{3, \ldots, n-1\}, \tag{3.3}$$

$$D_t = \left(\tfrac{q_{t+1}-q_t}{x_{t+1}-x_t}\right) - \left(\tfrac{q_t - q_{t-1}}{x_t - x_{t-1}}\right) \qquad \forall t \in \{3, \ldots, n-1\}, \tag{3.4}$$

$$\xi_t \geq D_t, \qquad \forall t \in \{3, \ldots, n-1\}, \tag{3.5}$$

$$\xi_t \geq -D_t, \qquad \forall t \in \{3, \ldots, n-1\}, \tag{3.6}$$

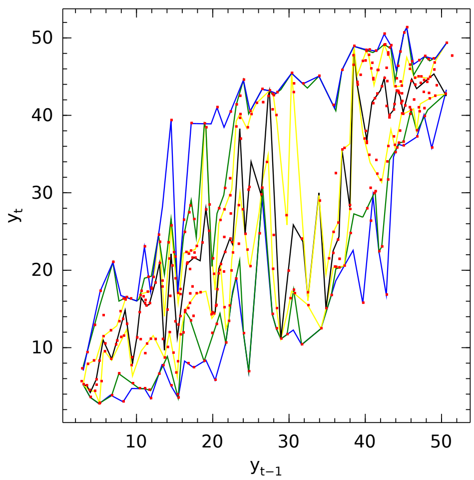
$$\delta_t^+, \delta_t^-, \xi_t \geq 0, \qquad \forall t \in \{3, \ldots, n-1\}. \tag{3.7}$$

The output of our optimization problem is a sequence of ordered points $\{(x_t, q_t)\}_{t\in T}$. 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.

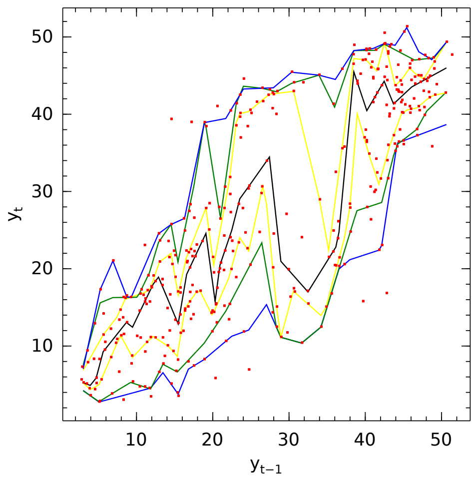
The quantile estimation is done for different values of λ. By using different levels of penalization on the second difference, the estimation can be more or less adaptive to the fluctuation. It is important to notice that the usage of the ℓ_1-norm as penalty leads to a piecewise linear solution q_t. Figure 3.1 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 3.1f, where the 95%-quantile crosses over the 90%-quantile. To prevent this, we can include a non-crossing constraint:

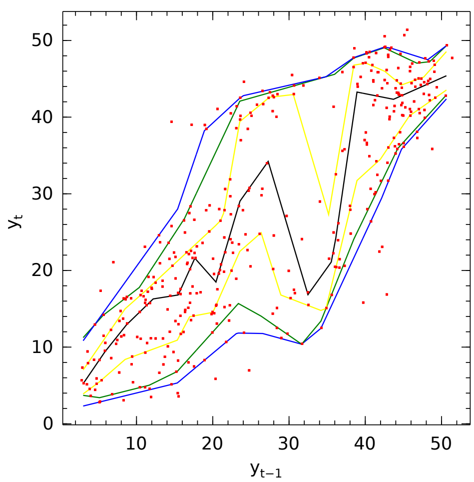
$$q_i^{\alpha} \leq q_i^{\alpha'}, \quad \forall i \in I, \alpha < \alpha'. \tag{3.8}$$



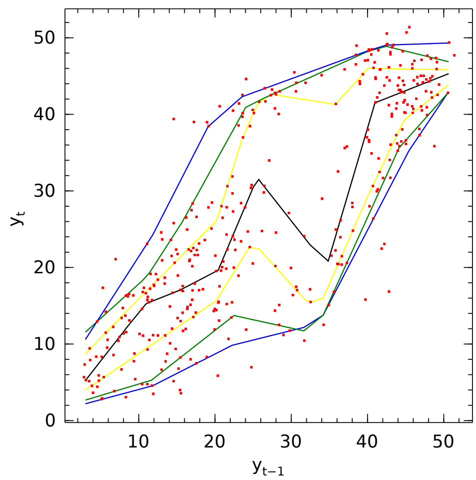
(a) $\lambda = 0.1$



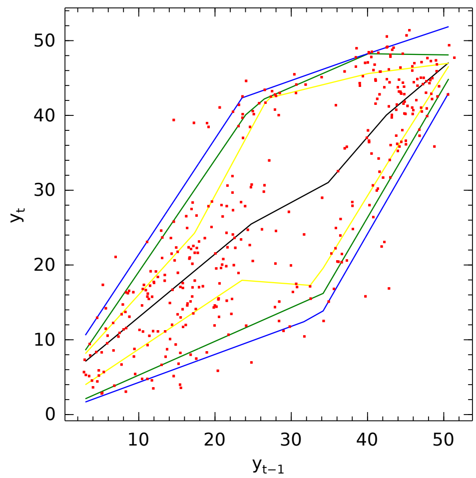
(b) $\lambda = 0.3$



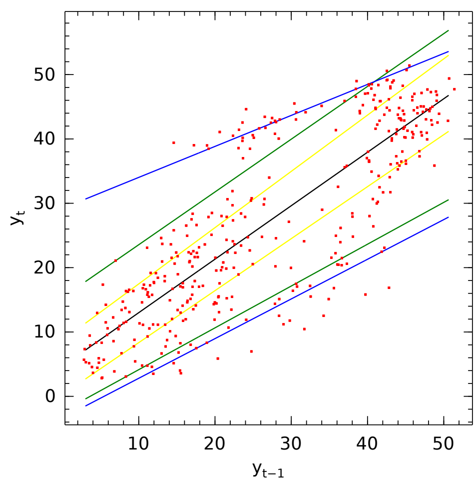
(c) $\lambda = 1$



(d) $\lambda = 3$



(e) $\lambda = 10$



(f) $\lambda = 150$

Figure 3.1: 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.

This means that when α' is a higher quantile than α, then the values from the α'-quantile must be bigger than those of the α-quantile for each and every point.

As a result of this nonparametric estimation, we are able to establish a relation between y_t and y_{t-p} in a way that the model adjusts itself automatically to the present nonlinearities. For this, we only have to supply a numeric value for λ. This approach, however, have yet some issues do be discussed.

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.

References

[1] Alexandre Belloni and Victor Chernozhukov. Least squares after model selection in high-dimensional sparse models. 2009.

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

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

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

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

[6] Jose AF Machado. Robust model selection and m-estimation. *Econometric Theory*, 9:478–493, 1993.

[7] Fernando Porrua, Bernardo Bezerra, Luiz Augusto Barroso, Priscila Lino, Francisco Ralston, and Mario Pereira. Wind power insertion through energy auctions in brazil. In *Power and Energy Society General Meeting, 2010 IEEE*, pages 1–8. IEEE, 2010.

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

4 Appendices

4.1 Proof of quantiles as an optimization problem

Let $Z^\alpha = \arg\min_Q E[\alpha \max\{0, X-Q\} + (1-\alpha)\max\{0, Q-X\}]$. We can rewrite the function as

$$\begin{aligned}
Y &= \alpha \int_Q^\infty (X-Q)dF_x + (1-\alpha)\int_{-\infty}^Q (Q-X)dF_X \\
&= \alpha \int_Q^\infty XdF_x - \alpha Q \int_Q^\infty QdF_x + Q\int_{-\infty}^Q dF_x - \int_{-\infty}^Q XdF_x - \alpha Q \int_{-\infty}^Q dF_x + \alpha \int_{-\infty}^Q XdF_x \\
&= \alpha \int_Q^\infty XdF_x - \alpha Q + QF_X(Q) - \int_{-\infty}^Q XdF_x - \alpha Q F_X(Q) + \alpha \int_{-\infty}^Q XdF_x \\
&= \alpha \int_Q^\infty XdF_x - \alpha Q + QF_X(Q) - \int_{-\infty}^Q XdF_x + \alpha \int_{-\infty}^Q XdF_x
\end{aligned}$$

By the first order condition for optimality, we need that $\frac{dZ(Q^*)}{dQ} = 0$. So, we have:

$$-\alpha Q^* f(Q^*) - \alpha + F_X(Q^*) + Q^* f(Q^*) - Q^* f(Q^*) + \alpha Q^* f(Q^*) = 0$$
$$F_X(Q^*) = \alpha.$$

Thus, we have that Z^α is the $\alpha - \mathit{quantile}$ of random variable X.

4.2 Tables

	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

	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

	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

	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

	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

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1981	23.36	28.34	12.44	18.35	17.10	22.49	23.57	40.10	48.40	42.13	43.70	37.23
1982	20.54	17.48	7.42	10.87	16.57	20.79	27.95	42.55	49.12	42.48	44.78	40.20
1983	27.94	24.50	22.60	22.24	29.62	27.05	33.92	45.06	50.64	49.32	43.83	36.14
1984	20.37	15.35	3.94	3.57	7.85	14.65	20.56	41.01	44.58	44.31	42.94	31.65
1985	10.38	4.71	5.15	2.84	7.27	10.36	14.53	39.33	45.18	41.21	42.15	23.02
1986	18.86	8.25	3.00	5.23	17.29	17.85	23.08	41.36	48.30	42.83	44.36	36.41
1987	26.09	24.71	6.90	21.02	20.73	19.53	28.42	42.94	48.06	44.26	43.11	39.67
1988	15.75	11.66	4.51	4.36	8.29	11.50	19.10	38.40	46.47	44.80	41.79	22.40
1989	19.92	14.52	5.08	2.75	5.62	11.42	17.17	38.94	43.92	43.70	40.69	26.34
1990	29.74	11.70	15.69	14.02	14.85	22.28	24.02	44.55	48.18	44.66	41.51	32.41
1991	17.09	13.46	7.68	6.63	8.51	16.17	26.46	43.36	49.00	45.86	40.14	36.57
1992	21.41	19.78	14.25	21.45	24.24	24.64	30.34	45.43	51.33	47.66	44.50	37.97
1993	27.86	20.13	14.36	16.63	20.94	26.43	30.60	44.07	44.73	43.78	41.40	34.18
1994	12.45	11.06	4.70	5.85	10.49	11.04	23.03	38.50	48.92	47.30	44.97	36.55
1995	20.31	5.80	9.47	5.36	5.62	14.15	23.54	42.48	50.49	42.74	41.15	29.90
1996	19.89	11.85	3.43	5.08	8.26	16.29	24.89	40.52	48.44	44.92	40.15	36.37
1997	23.89	27.80	14.30	11.95	17.55	22.22	31.82	44.07	43.14	40.00	37.94	28.36
1998	15.04	21.70	10.61	17.28	21.57	22.31	27.26	42.45	49.04	46.76	37.22	35.74
1999	22.18	15.39	8.18	13.66	8.67	16.49	22.30	40.43	47.75	39.85	36.95	35.54
2000	16.75	7.95	11.33	10.47	16.73	15.07	18.90	38.91	44.26	46.34	41.98	31.62
2001	24.03	11.82	11.09	9.23	16.30	14.53	25.73	41.57	45.79	40.99	41.52	42.76
2002	16.81	22.08	13.40	11.07	15.71	17.52	26.55	41.64	45.80	45.94	40.64	30.58
2003	17.42	14.05	10.03	11.26	15.39	17.01	28.29	39.98	47.02	47.07	40.47	34.85
2004	15.04	13.34	17.84	16.97	20.10	19.48	25.03	40.11	48.25	47.21	44.13	35.79
2005	24.89	20.47	13.01	20.88	19.98	21.48	27.81	42.74	46.09	46.93	44.98	36.08
2006	32.48	15.44	12.93	6.59	12.19	19.08	27.79	40.72	46.01	44.38	42.85	33.99
2007	28.93	11.13	16.10	11.91	17.68	21.57	30.56	42.95	47.80	47.61	42.97	35.98
2008	20.42	15.46	3.51	9.37	8.71	13.02	23.61	36.93	45.82	46.49	43.91	35.19
2009	21.48	15.16	6.74	3.80	4.48	12.88	24.53	38.40	47.70	40.87	46.73	38.03
2010	24.75	30.70	16.99	16.95	15.72	16.86	27.43	43.18	48.71	35.79	41.30	30.15
2011	16.33	14.79	9.30	7.70	13.35	18.60	23.53	39.62	46.97	40.99	44.75	42.79

Table 4.1: Monthly Mean Power data of Icaraizinho wind farm, located on the Brazilian northeast. It is available for download here: https://raw.githubusercontent.com/mcruas/data/master/icaraizinho.csv.