Pontíficia Universidade Católica do Rio de Janeiro
Departamento de Engenharia Elétrica

Quantile Regression

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

April 11, 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.

1 Regularization

When dealing with many candidates to use as covariates, one has to deal with the problem of selecting a subset of variables to use in constructing the model. This means that the vector of coefficients $\beta_\alpha = [\beta_{1\alpha} \cdots \beta_{P\alpha}]$ should not have all nonzero values. There are many ways of selecting a subset of variables among. A classic approaches for this problem is the Stepwise algorithm [2], which includes variables in sequence.

The approach we use in of doing regularization and selecting the best model for estimating the quantile function. At first, we use a Mixed Integer Linear Programming optimization problem (MILP) to find the best subset among all choices of covariates. The second way is by using a LASSO-type technique, which consists in penalizing the ℓ_1-norm of regressors, thus shrinking the size of estimated coefficients towards zero.

1.1 Best subset selection with MILP

In this part, we investigate the usage of MILP to select which variables are included in the model, up to a limit of inclusions imposed *a priori*. We establish that only K coefficients $\beta_{p\alpha}$ may have nonzero values, for each quantile α. This assumption is modeled with binary variables $z_{p\alpha}$, which indicates whether $\beta_{p\alpha}$ is included or not. The optimization problem that incorporates this idea is described below:

$$\min_{\beta_{0\alpha},\beta_\alpha,z_{p\alpha}\varepsilon^+_{t\alpha},\varepsilon^-_{t\alpha}} \quad \sum_{\alpha\in A}\sum_{t\in T}\left(\alpha\varepsilon^+_{t\alpha} + (1-\alpha)\varepsilon^-_{t\alpha}\right) \tag{1.1}$$

$$\text{s.t} \quad \varepsilon^+_{t\alpha} - \varepsilon^-_{t\alpha} = y_t - \beta_{0\alpha} - \sum_{p=1}^{P}\beta_{p\alpha}x_{t,p}, \qquad \forall t\in T, \forall\alpha\in A, \tag{1.2}$$

$$\varepsilon^+_{t\alpha}, \varepsilon^-_{t\alpha} \geq 0, \qquad \forall t\in T, \forall\alpha\in A, \tag{1.3}$$

$$-Mz_{p\alpha} \leq \beta_{p\alpha} \leq Mz_{p\alpha}, \qquad \forall p\in\{1,\dots,P\}, \tag{1.4}$$

$$\sum_{p=1}^{P} z_{p\alpha} \leq K, \qquad \forall\alpha\in A, \tag{1.5}$$

$$z_{p\alpha}\in\{0,1\}, \qquad \forall p\in\{1,\dots,P\}, \forall\alpha\in A, \tag{1.6}$$

$$\beta_{0\alpha} + \beta^T_\alpha x_t \leq \beta_{0\alpha'} + \beta^T_{\alpha'}x_t, \qquad \forall t\in T, \forall(\alpha,\alpha')\in A\times A, \alpha<\alpha', \tag{1.7}$$

The objective function and constraints (1.2), (1.3) and (1.7) are those from the standard linear quantile regression. By constraint (1.4), variable $z_{p\alpha}$ is a binary that assumes 1 when the coefficient $\beta_{p\alpha}$ is included, while (1.5) guarantees that at most K of them are nonzero. The value of M is chosen in order to guarantee that $M \geq \|\hat{\beta_\alpha}\|_\infty$. The solution given by $\beta^*_{0\alpha}$ and $\beta^*_\alpha = [\beta^*_{1\alpha}\cdots\beta^*_{P\alpha}]$ will be the best linear α-quantile regression with K nonzero coefficients.

We ran this optimization on the Icaraizinho dataset for each value of $K\in\{0,1,\dots,12\}$ and quantiles $\alpha\in\{0.05, 0.1, 0.5, 0.9, 0.95\}$. The full results table can be accessed on section 3.2. For all tested α-quantiles the 12^{th} lag was the one included when $K=1$. When $K=2$, the 1^{st} lag was included for all values of α, 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.

Defining groups for variables

Consider the optimization problem defined on (1.1)-(1.7). Equation (1.4) permits a different subset of variables for each α-quantile, as long as it is a set of K variables. For two similar probabilities α and α', however, it is not plausible that their chosen model be too different (for example, in one $\beta_{1\alpha}$ and $\beta_{4\alpha}$ are selected while $\beta_{2\alpha}$ and $\beta_{5\alpha}$) are selected by the other).

To address this issue, we propose to divide all $\alpha\in A$ into groups. The collection G of all groups g form a partition of A, and each α will belong to exactly one group g. The subset of selected covariates

must be the same for all α in the same group g. To model these properties as constraints, we use the following equations and inequalities, that take the place of inequality 1.4 on the optimization problem:

$$z_{p\alpha} := 2 - (1 - z_{pg}) - I_{g\alpha} \tag{1.8}$$

$$\sum_{g \in G} I_{g\alpha} = 1, \qquad \forall \alpha \in A, \tag{1.9}$$

$$-M z_{p\alpha} \leq \beta_{p\alpha} \leq M z_{p\alpha}, \quad \forall p \in P, \quad \forall \alpha \in A, \quad \forall g \in G, \tag{1.10}$$

$$I_{g\alpha}, z_{pg} \in \{0, 1\}, \qquad \forall p \in P, \quad \forall g \in G, \tag{1.11}$$

where G is a set of group index and z_{pg} is a binary variable that equals 1 iff covariate p is included on group g and $I_{g\alpha}$ equals 1 iff α belongs to group g. The logic behind constraint 1.10 is that

$$\text{If } z_{pg} = 0 \text{ and } I_{g\alpha} = 1 \text{ then } \beta_{p\alpha=0}.$$

This means that if covariate p belongs to group g, this covariate is not among group's g subset of variables, than its coefficient must be equal to 0, for that α. Note that variable $z_{p\alpha}$ behaves differently that when we are not considering groups. This means that if probability α belongs to group g but variable p is not selected to be among the ones of group g, than $\beta_{p\alpha}$ is zero. Equation (1.8) defines $z_{p\alpha}$ to simplify the problem.

1.2 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 by changing a continuous parameter λ, which penalizes the size of the ℓ_1-norm. When the value of λ gets bigger, fewer variables are selected to be used. This is the same strategy of the LASSO methodology, and its usage for the quantile regression is discussed in [3]. The proposed optimization problem to be solved is:

$$\min_{\beta_{0\alpha}, \beta_\alpha} \sum_{t \in T} \alpha |y_t - q_\alpha(x_t)|^+ + \sum_{t \in T} (1 - \alpha) |y_t - q_\alpha(x_t)|^- + \lambda \|\beta_\alpha\|_1, \tag{1.12}$$

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

For such estimation to be coherent, however, each covariate must have the same relative weight in comparison with one another. So, before solving the optimization problem, we perform a linear transformation such that all variables have mean $\mu = 0$ and variance $\sigma^2 = 1$. We apply the transformation $\tilde{x}_{t,p} = (x_{t,p} - \bar{x}_{t,p})/\hat{\sigma}_{x_{t,p}}$, where $\bar{x}_{t,p}$ and $\hat{\sigma}_{x_{t,p}}$ are respectively the sample's unconditional mean and standard deviation. The $\tilde{y}_{t-p,i}$ series will be used to estimate the coefficients, as this series has the desired properties.

After the process of normalization, we can rewrite problem 1.12 as a LP problem, as shown below:

$$\tilde{\beta}_\lambda^{*LASSO} = \underset{\beta_0, \beta, \varepsilon^+_{t\alpha}, \varepsilon^-_{t\alpha}}{\arg\min} \quad \textstyle\sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon^+_{t\alpha} + (1 - \alpha) \varepsilon^-_{t\alpha} \right) + \lambda \sum_{p=1}^{P} \xi_{p\alpha} \tag{1.13}$$

$$\text{s.t.} \qquad \varepsilon^+_{t\alpha} - \varepsilon^-_{t\alpha} = y_t - \beta_{0\alpha} - \textstyle\sum_{p=1}^{P} \beta_{p\alpha} \tilde{x}_{t,p}, \qquad \forall t \in T, \forall \alpha \in A, \tag{1.14}$$

$$\varepsilon^+_{t\alpha}, \varepsilon^-_{t\alpha} \geq 0, \qquad \forall t \in T, \forall \alpha \in A, \tag{1.15}$$

$$\xi_{p\alpha} \geq \beta_{p\alpha}, \qquad \forall p \in P, \forall \alpha \in A, \tag{1.16}$$

$$\xi_{p\alpha} \geq -\beta_{p\alpha}, \qquad \forall p \in P, \forall \alpha \in A. \tag{1.17}$$

This model is built upon the standard linear programming model for the quantile regression (equation **??**). On the above formulation, the ℓ_1 norm of equation (1.12) is substituted by the sum of ξ_p, which represents the absolute value of $\beta_{p\alpha}$. The link between variables ξ_p and $\beta_{p\alpha}$ is made by constraints (1.16) and (1.17). Note that the linear coefficient $\beta_{0\alpha}$ is not included in the penalization, as the sum of penalties on the objective function 1.13. Each component of the output $\tilde{\beta}_\lambda^{*LASSO}$ must be corrected by multiplying each coefficient for its standard deviation: $\beta_{p\alpha,\lambda}^{*LASSO} = \tilde{\beta}_{p\alpha,\lambda}^{*LASSO} \hat{\sigma}_{x_{t,p}}$.

For low values of λ, the penalty over the size of coefficients is small. Because of that, our output of problem (1.13)-(1.17) model where all coefficients have a nonzero value. On the other hand, when 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 1.1 ($\alpha \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$).

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 **??**. In summary, the optimization in equation 1.12 acts as a variable selection for the subsequent estimation, which is normally called the post-LASSO estimation [1].

In this estimation made *a posteriori*, 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

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

The post-lasso coefficients β^*_λ are the solution from the optimization problem given below:

$$\begin{aligned}(\hat{\sigma}^*_\lambda, \beta^*_\lambda) \overset{(obj,var)}{\longleftarrow} \min_{\beta_0,\beta,\varepsilon^+_t,\varepsilon^-_t} & \sum_{t\in T} \left(\alpha\varepsilon^+_t + (1-\alpha)\varepsilon^-_t\right) \\ \text{s.t. } & \varepsilon^+_t - \varepsilon^-_t = y_t - \beta_0 - \sum_{p\in L_\lambda} \beta_p x_{t,p}, \quad \forall t \in T, \\ & \varepsilon^+_t, \varepsilon^-_t \geq 0, \qquad\qquad\qquad\quad \forall t \in T.\end{aligned} \tag{1.18}$$

The variable $\hat{\sigma}^*_\lambda$ receives the value of the objective function on its optimal solution.

1.3 Model selection

On sections 1.1 and 1.2, we presented two ways of doing regularization. Nonetheless, regularization can be done with different levels of parsimony. For example, one can select a different number K of variables to be included in the best subset selection via MILP or choose different values of λ for the ℓ_1 penalty. Each of these choices may lead to a different model, and th .....

Solving a LP problem is much faster than a similar-sized MILP 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 MILP problem. We propose an experiment that is described as follows. First, we calculate the quantity $\|\beta^*_\lambda\|_0$ of nonzero coefficients, for each given lambda, for the LASSO estimations. Then, for each number K of total nonzero coefficients, there will be a penalty λ^*_K which minimizes the errors from the quantile regression's objective function (given on equation (1.18)):

$$\lambda^*_K = \arg\min_\lambda \{\, \hat{\sigma}^*_\lambda \quad \mid \|\beta^*_\lambda\|_0 = K\}\,. \tag{1.19}$$

We, then, define the set L^m_K, which contains all nonzero indexes, for a given K, of method m. Thus, we can compare the best lasso fit where exactly K variables are selected with the best fit given by the MILP problem, also with K variables selected.

As the MILP solution is the exact solution for the problem, while the LASSO solution is an approximation, we use the former as a *benchmarking* for the quality of the latter solution. To help us view the difference of results between both methods, we define a similarity metric d between the subset of coefficients chosen by each one of them. It is desirable that the LASSO solution be as related with the MILP solution as possible. The similarity is calculated as the solution of the following optimization

(K rang
ing from
0 until
12, wher
0 means
that only
the in-
tercept i
included

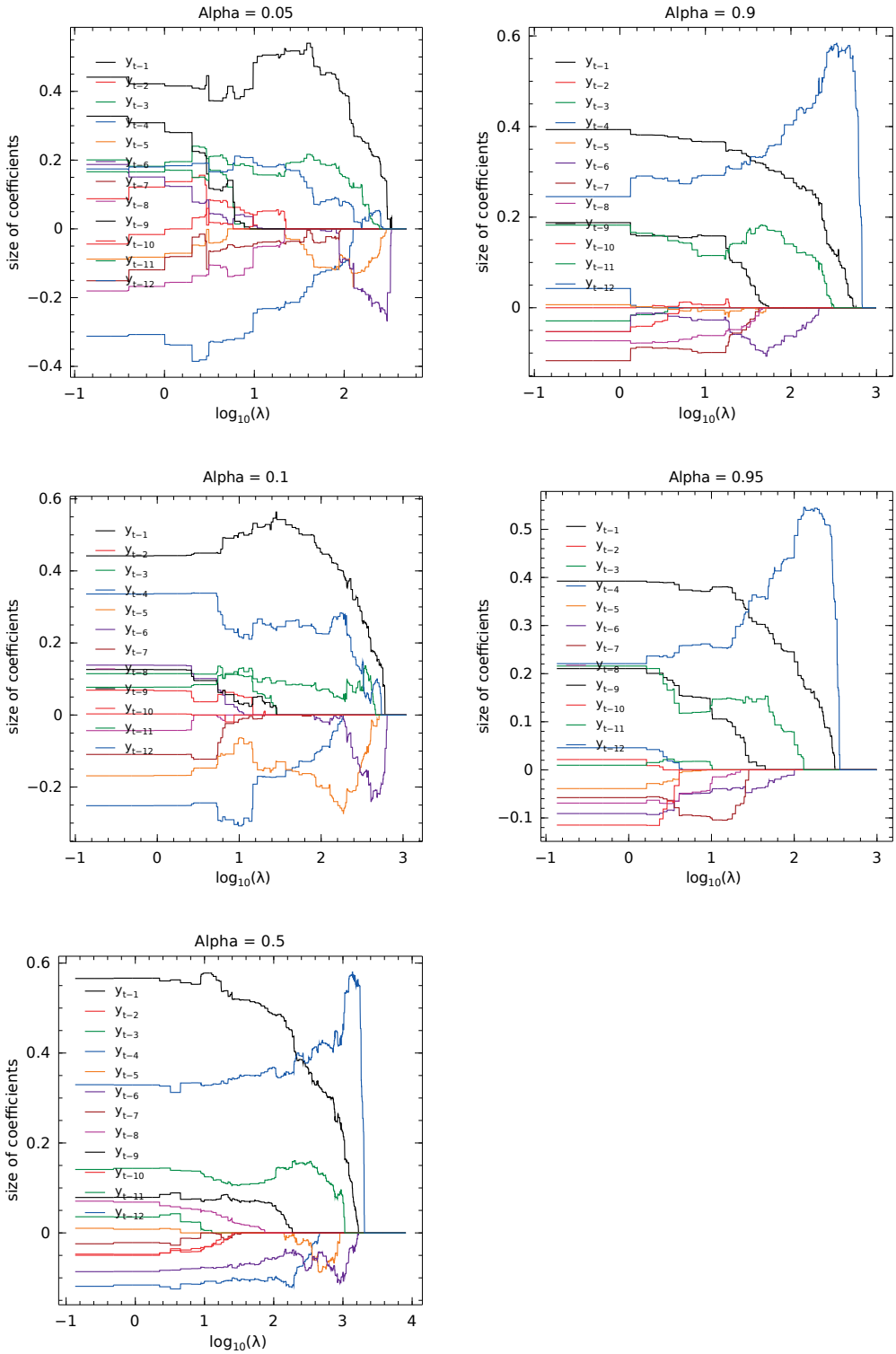

Figure 1.1: Coefficients path for a few different values of α-quantiles. λ is presented in a $\log_{10}$ scale, to make visualization easier.

problem

$$d(\beta^*_{MILP(K)}, \beta^*_{\lambda^*_K}) = 1 - \max_{0 \leq \delta_{ij} \leq 1} \quad \sum \sum_j \delta_{ij} |\rho_{ij}| \tag{1.20}$$

$$\text{s.t.} \quad \sum_j \delta_{ij} = 1 \quad \forall i \in L_K^{MILP}, \tag{1.21}$$

$$\sum_i \delta_{ij} = 1 \quad \forall j \in L_K^{LASSO}, \tag{1.22}$$

$$\delta_{i,j} = 0, \quad \forall i \in \overline{L}_K^{MILP}, \forall j \in \{1, \ldots, P\}, \tag{1.23}$$

$$\delta_{i,j} = 0, \quad \forall j \in \overline{L}_K^{LASSO}, \forall i \in \{1, \ldots, P\}, \tag{1.24}$$

where ρ_{ij} is the correlation between variables i and j and $\delta_{ij} = 1$ means that the selected variable i from the set L_k^{MILP} is associated with the variable j from set L_k^{LASSO}, and the constraints guarantee that each variable is related with only one other variable. The set $\overline{L}_K^m$ represents the variable indexes $\{1, \ldots, P\} \backslash L_K^m$ for the method m which are not present in L_K^m. When $d = 0$, both solutions are equal, and the LASSO method was able to select the best subset among the available possibilities.

As seen before, we have a best solution for each desired K. The question that arises now is how to select the ideal number of variables to use. 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 for a covariate to be included in the model, it must supply enough goodness of fit. In [4], 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:

$$SIC(m) = n \log(\hat{\sigma}^*) + \frac{1}{2} K \log n, \tag{1.25}$$

where K is the model's dimension. This procedure leads to a consistent model selection if the model is well specified.

Figure 1.2 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. In this experiment, the minimum SIC value for the MILP problem is usually found between 4 and 6 variables in the model.

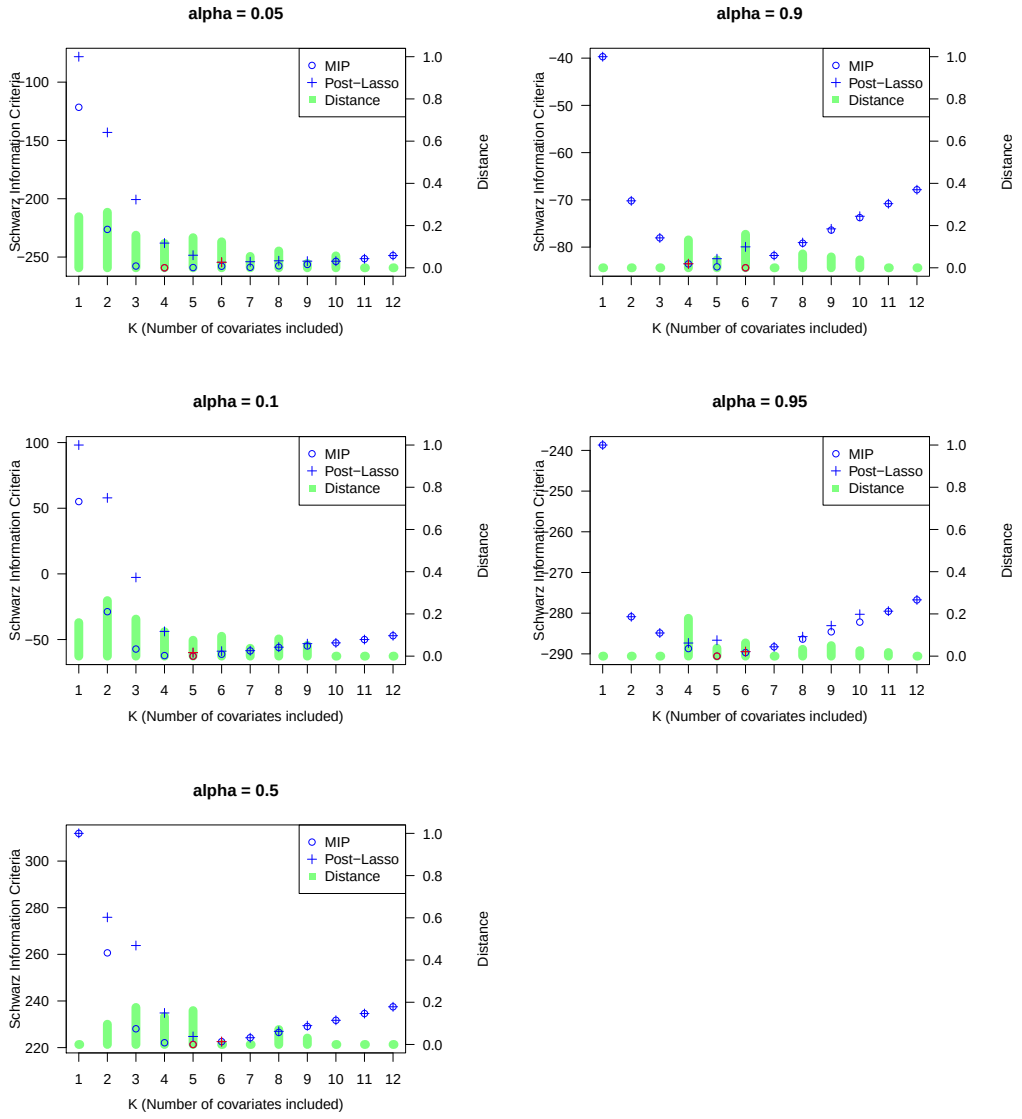

Figure 1.2: **REVER: Comparison of SIC values between a solution with LASSO as a variable selector and the best subset selection with MILP**. The Information Criteria is displayed on the y-axis, while the number of variables included is shown on the x-axis. Both solutions of the MILP and the best LASSO for a given K are The bars represent the distance d as defined on problem (1.20)-(1.24).
(*) 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.

2 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 **??**.

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 (**??**)-(**??**), reproduced here for convenience:

$$\min_{q_\alpha, \varepsilon^+_{t,\alpha}, \varepsilon^-_{t,\alpha}} \quad \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon^+_{t,\alpha} + (1-\alpha)\varepsilon^-_{t,\alpha} \right) \tag{2.1}$$

$$\text{s.t.} \quad \varepsilon^+_{t,\alpha} - \varepsilon^-_{t,\alpha} = y_t - q_\alpha(x_t), \qquad \forall t \in T_\tau, \forall \alpha \in A, \tag{2.2}$$

$$\varepsilon^+_{t,\alpha}, \varepsilon^-_{t,\alpha} \geq 0, \qquad \forall t \in T_\tau, \forall \alpha \in A, \tag{2.3}$$

$$q_\alpha(x_t) \leq q_{\alpha'}(x_t), \qquad \forall t \in T_\tau, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \tag{2.4}$$

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 (**??**)-(**??**) 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 **??**.

   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.

We applied this procedure on the ENA data for the brazilian southeast region. The quantiles of scenarios are shown on figure

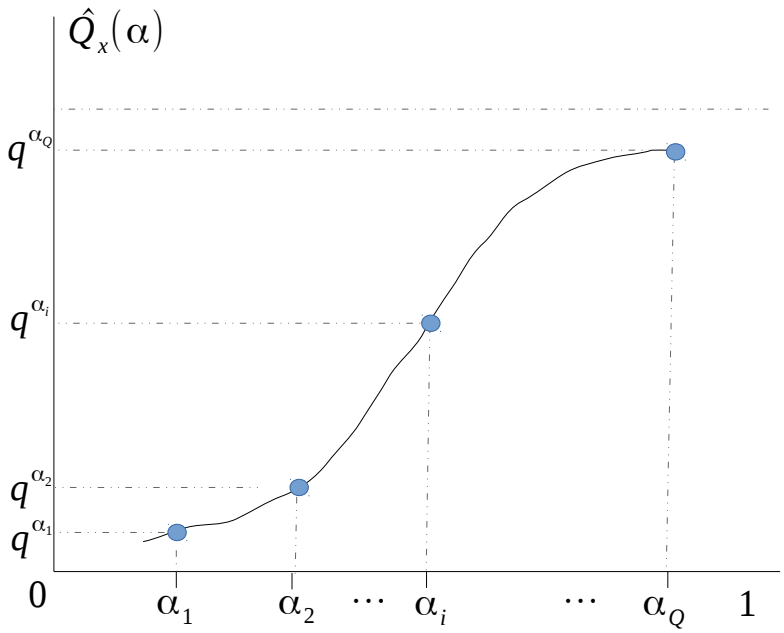

Figure 2.1: Fitting a distribution function from quantile estimations

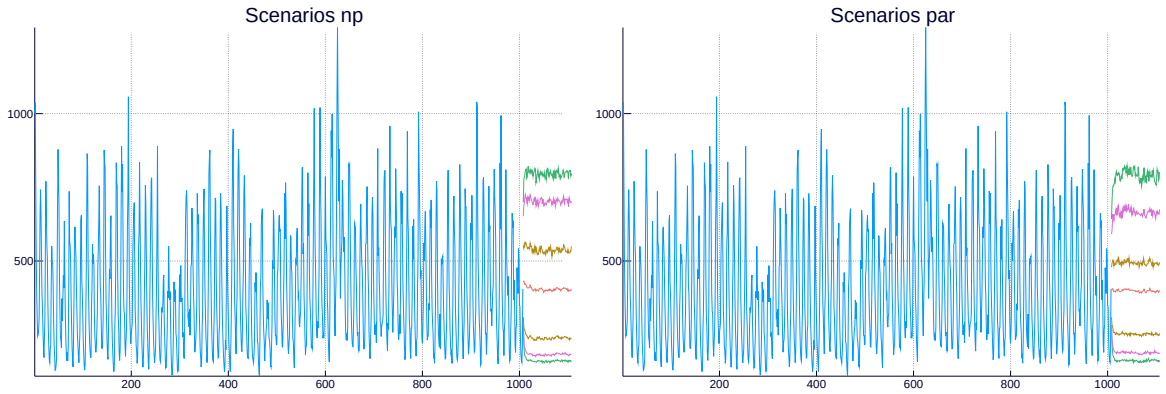

Figure 2.2: Scenario simulation for the nonparametric (on the left) and parametric (on the right) for the ENA Southeast dataset

3 Appendices

3.1 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

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

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 $\alpha - quantile$ of random variable X.

3.2 MIP coefficients tables

The following tables inform the size of Coefficients when using the regularization method based on MIP described on session 1.1. When using this method, we choose a parameter K which defines the total number of nonzero coefficients (without accounting the intercept β_0, which is always included). In each column we find the estimated values of coefficients for each different choice of K. As coefficients are quantile dependent, we provide tables for $\alpha \in (0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95)$.

	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 3.1: Coefficients for quantile $\alpha = 0.05$

	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 3.2: Coefficients for quantile $\alpha = 0.1$

.

	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 3.3: Coefficients for quantile $\alpha = 0.5$

.

	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 3.4: Coefficients for quantile $\alpha = 0.9$

.

	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 3.5: Coefficients for quantile $\alpha = 0.95$

.

	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

References

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

[2] MA Efroymson. Multiple regression analysis. *Mathematical methods for digital computers*, 1:191–203, 1960.

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

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