Optimization of Electricity Reserved Capacity

Federica Asnicar

Abstract

This thesis aims to developing a model that describes the dependency of the Actual Upward Regulating power, referring to the Danish Electric Energy Market, on the Wind Forecasts, in order to predict the needs of ancillary services at least one day in advance.

The model is based on Nonparametric Regression Quantiles Methods and the softwares used in the thesis are R (to implement the nonparametric regression quantile model) and Matlab (to plot the results in three dimensions).

The use of nonparametric regression quantile methods allows us to better model the data, introducing spline smoothers, in order to get a more adaptable model.

Preface

This thesis was prepared at Informatics Mathematical Modelling, the Technical University of Denmark in partial fulfillment of the requirements for acquiring the Mc degree in mathematic. The work started January 2, 2006.

The thesis deals with different aspects of mathematical modeling of the Electricity Power Market in Denmark using data and partial knowledge about the structure of the systems.

Lyngby, August 10 - 2006

Federica Asnicar

Acknowledgements

I wish to thank my supervisors Henrik Madsen and Henrik Aalborg Nielsen for their support, guidance and participation in discussions during the project.

I wish to express my warmest thanks to my italian supervisor dott. Giacomo Aletti, Università Statale di Milano, for his valuable comments and advice.

I also wish to thank Energinet.dk and Peter Børre Eriksen for being allowed to use the data.

To my parents.

Contents

List of Figures

CHAPTER 1

The Scandinavian Electric Power Exchange Market

1.1 Introduction

Energinet.dk is the transmission system operator in Denmark. Energinet.dk, and its two wholly owned subsidiaries − Eltransmission.dk A/S and Gastransmission.dk A/S − operates and maintains the electricity and gas transmission grids.

The main aim of this thesis arises from the necessity for Energinet.dk of efficiently incorporating the wind power into the overall energy system.

In 2004 the total costs of ancillary services and regulating reserves were DKK 392 millions. Hence the necessity of developing a upward regulating power model (see section 1.4).

Energinet.dk has worked with IMM (the Department of Informatics and Mathematical Modelling at the Technical University of Denmark (DTU)), on a number of different projects concerning electricity, wind power and the electricity market in general.

In this thesis the studies on developing a upwars regulating power model are referred to the Quantile Reggression Theory, first developed by R.Koenker and G.Bassett in 1978 [16].

Energinet.dk has provided data and other useful information, on which this thesis has mainly been based.

The informations about the Scandinavian Electricity Market, contained in this chapter, mostly come from the Energinet.dk website [14] and the NordPool website [3].

1.2   Nord Pool

Nord Pool ASA - The Nordic Power Exchange - is the world's only multinational exchange for trading electric power.

Nord Pool was established in 1993, and is owned by the two national grid companies, Statnett SF in Norway (50%) and Affärsverket Svenska Kraftnät in Sweden (50%).

Nord Pool has two marketplaces: Elspot and Eltermin.

Elspot is a market where physical kilowatt hours are traded in the same manner as shares are traded for example on a stock exchange. Elspot is the spot market for power.

Eltermin is a marketplace for financial futures trading, where price hedging contracts are traded.

Market participants who desire to purchase power via Nord Pool's Elspot marketplace must submit their bids to Nord Pool no later than 12 noon the day before they desire delivery. Correspondingly, the market participants who desire to sell power via Elspot must send their bids to Nord Pool the day before they desire to make delivery.

The sale of electrical power takes place in various arenas. A producer may sell directly to consumers, or power companies may buy electrical power from a producer or power exchange (Nord Pool) and then resell the power. Brokers help establish contracts between buyers and sellers, while traders buy and resell power.

A power company can decide whether it wants to:

- buy all its power on the market

- buy some power on the market and generate the rest itself

- generate precisely the power required to meet the customers' expected consumption during this hour

- generate more power than their customers' expected consumption and sell thus power on the market

The bids to buy and sell for each hour of the following day are compiled by Nord Pool into an overall curve for the demand and an overall curve for the supply. The system price is read from where the two curves intersect (see Figure 1.1).

Nord Pool sets a system price for each hour during the following day. The price can vary from hour to hour, but it is fixed for one hour at a time.

When Nord Pool has made its calculations, it releases the prices for the following day and tells the market participants how much power they have purchased or sold for the various hours of the following day, and the transmission system operators are notified of the contracted purchase and sales volumes. This can be entered then in the balance accounting for various market participants.

Situations may of course arise where too little power is supplied in an area at the system price, and bottlenecks in the grid may make it impossible to supply an adequate volume of power to this area. In this case it will become a high price zone: an area where the price is higher than the system price.

Correspondingly, situations may arise where too much power is supplied in an area at the system price, and the grid cannot transport enough power out of this area. In such cases this zone becomes a low price area: an area where the price is less than the system price.

Norway can be divided into several price zones (three zones this winter), while Finland, Sweden and Western Denmark (Jylland/Fyn) cannot be divided into more than one price zone internally.
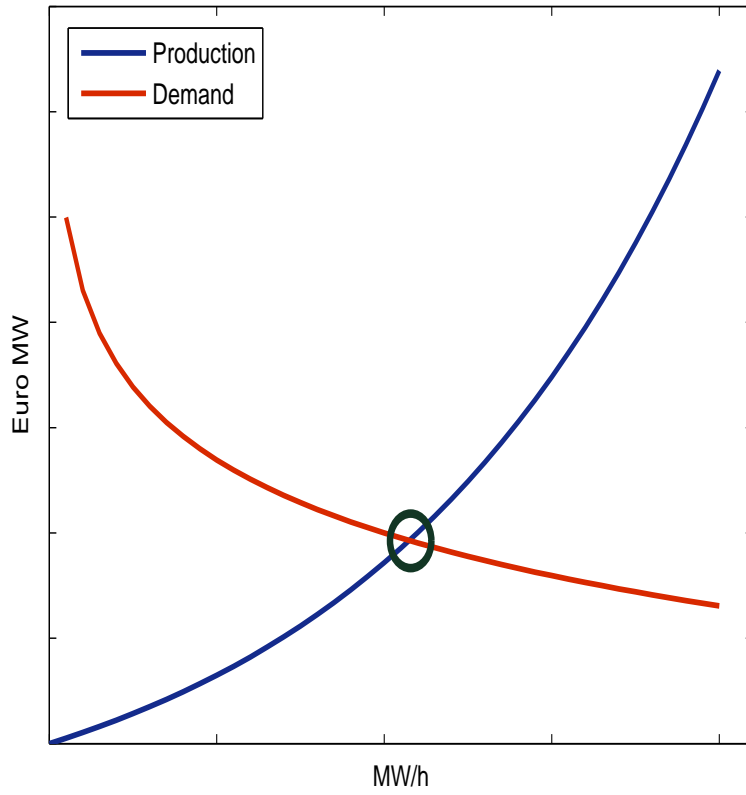
Figure 1.1: Intersection between Demand curve and Production curve.

1.3 Denmark

The **Kingdom of Denmark** (Danish: Kongeriget Danmark) is the geographically smallest and southernmost Nordic country in addition to being the oldest. It locates north of Germany (its only land neighbourgh), southwest of Sweden, and south of Norway, in Scandinavia.

- Denmark Area: 43,094 square kilometres - of which:
  - Land: 42,394 km^2
  - Water: 700 km^2
- Population: 5,432,335 (July 2005 est.)
- Electric consumption: 35 TWh (2003)
- Electric generation: 44 TWh (2003) - of which:
  - Other thermal power (gas/coal): 87 per cent
  - Other renewable power (wind): 13 per cent

The Danish climate is temperate, humid and overcast: mild, windy winters and cool summers.

Denmark's energy production is mainly based on imported coal, oil and natural gas from the Danish sector of the North Sea, and wind energy. In addition there are straw and other biological fuels, solar energy and geothermal energy, which together constitute a small proportion. This is, however, increasing with technological developments.

1.3.1 Energinet.dk

Electricity is mainly produced in regional power stations by burning coal supplemented with natural gas, oil, biological fuel and waste products. Natural gas plays an increasingly important role, while the use of oil is decreasing. The regional power stations supply is distributed throughout the country by one electricity company: Energinet.dk.

Energinet.dk is formally established as a merger between Eltra, Elkraft System, Elkraft Transmission and Gastra. Two subsidiaries, Eltransmission.dk A/S

Figure 1.2: Denmark
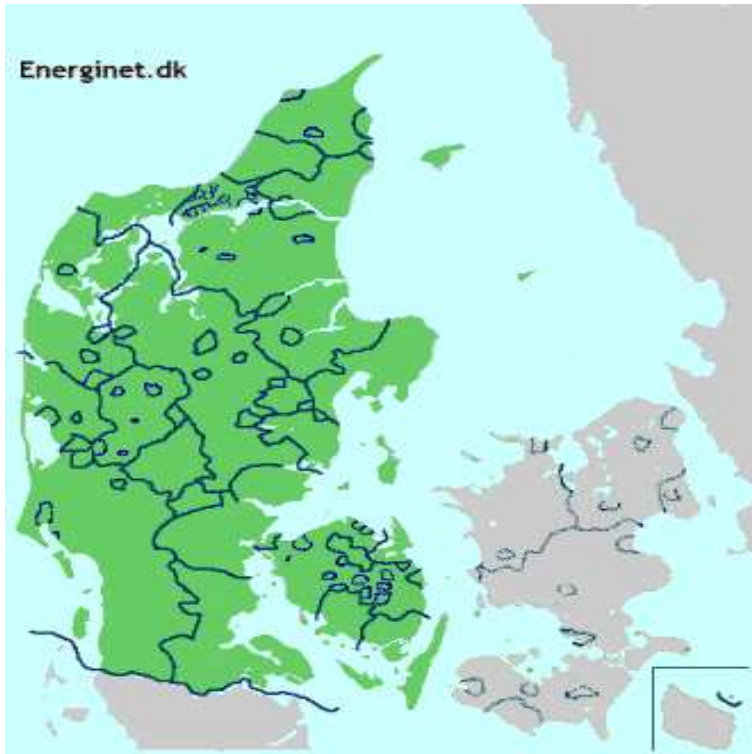
Figure 1.3: Energinet.dk transmission lines.

and Gastransmission.dk A/S were established at the same time. The merger was effected with retrospective effect as from 1 January 2005.

Energinet.dk is a result of a broad political compromise, which was made in March 2004. The bill on Energinet Danmark was passed on 14 December 2004.

As system operators, Energinet.dk and its two subsidiaries have four main responsibilities:

- Ensuring the physical of the system in the short and long term.

- Administering market access and planning market function.

- Planning, developing and operating the overall transmission grid and the international connection.

- Ensuring that the electricity and natural gas systems live up to the energy policies pursued in Denmark.

A special challenge for Energinet.dk is the efficient incorporation of wind power into the overall energy system.

Being responsible for ensuring that the electricity and natural gas systems live up to the energy policies pursued in Denmark involves, among other things, the implementation of sophisticated systems for system operation, monitoring and control, allowing the handling of considerable volumes of wind power.

In October 2004 Eltra was placed on "positive outlook" by Standard and Poor's, one of the world's preeminent providers of credit ratings and globally recognized financial-market indices.

However, Standard and Poor's also highlights a weakness: a risk of losses on counterparties in the real-time market.

Energinet.dk must ensure a balance between consumption and production at the moment of operation. The balancing is achieved via the regulating power market in which generators offer to change their production relative to the plan against payment.
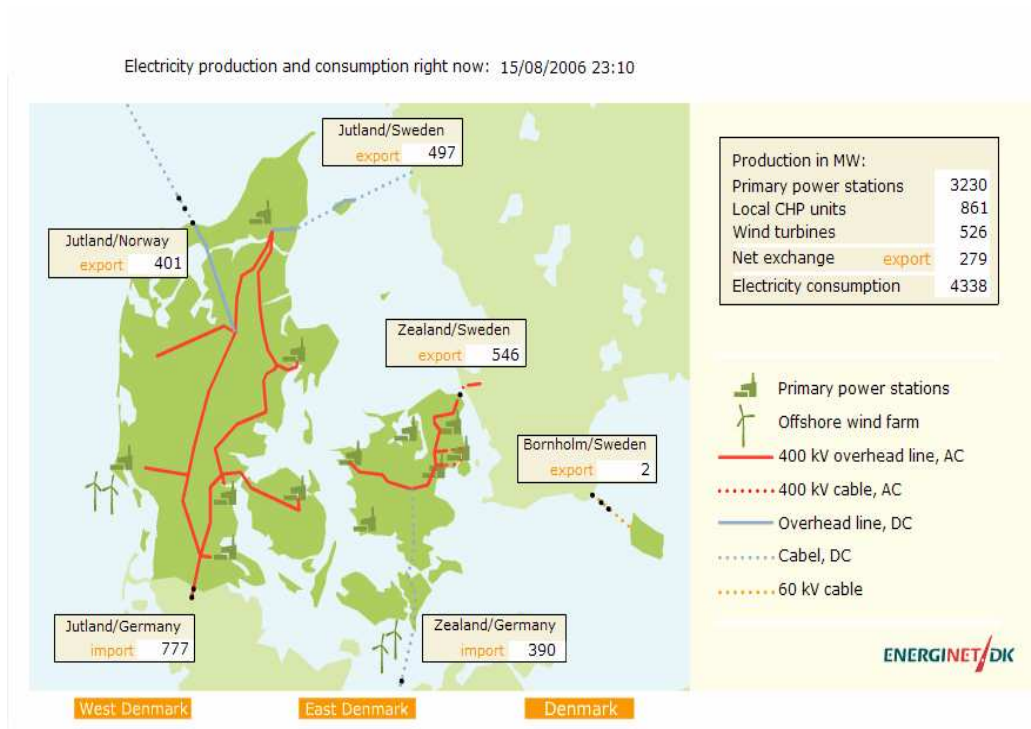
Figure 1.4: The electricity grid.

1.4 Regulating Power

A transmission system operator, like Energinet.dk, is also responsible for keeping the power system in balance, and thus it is responsible for the overall physical management and control of the national power system. Technically this means that the frequency is maintained at 50 Hz.

The frequency of the power system can be considered a measure of the balance or imbalance between production and consumption: for the NordPool market the frequency has to be between 49.9 Hz and 50.01 Hz.

If the consumption exceeds the production, then the frequency of the system will drop to less than 50 Hz($\Rightarrow$ **upward regulating power**). If the production is too great and it exceeds the consumption, then the frequency of the system will increase to over 50 Hz ($\Rightarrow$ **downward regulating power**).

The volume of power that the transmission system operator trades in this manner is called regulating power. The transmission system operator chooses who will change their production or consumption based on a price offer that the producers and consumers have given for this. The producer or consumer who has given the lowest price for the change that is required will be chosen.

Energinet.dk (Eltra in Western Denmark) collaborates with the other Nordic system operators on the delivery of regulating power. However, due to congestion on the transmission lines to Norway and Sweden, there are times where is physically impossible to obtain regulating power in the Nordic region. The large share of wind power in Western Denmark requires flexible solutions to ensure the necessary balance between production and consumption.

During the operational day, Eltra maintains the system balance by purchasing regulating power from producer in Western Denmark, via the system operators in Norway and Sweden or under bilateral agreement players south of the Danish-German border. Eltra sells the regulating power to players which are seeing deviation between their planned and they actual production/consumption.

The players' imbalance are settled under a dual-price system, giving rise to a surplus:

- Imbalances which go in the same direction as the system imbalance and which thereby contribute to an increased imbalance are settled at the price of manual regulating power.

- Imbalances which go against the system imbalance and which thereby help
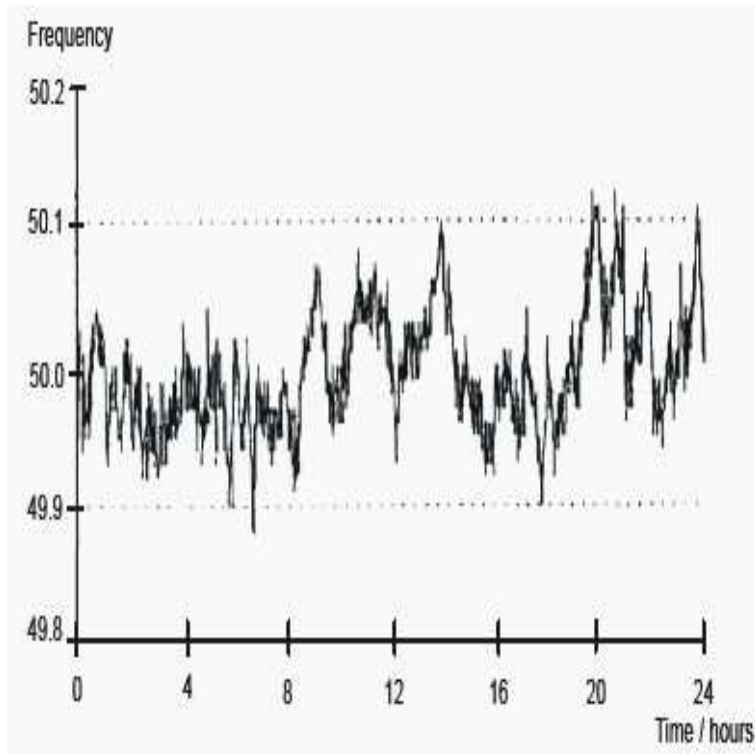
Figure 1.5: Market balance for 1 day

the system are settled at spot price.

The surplus is used to cover other costs in the system accounts. The surplus from the dual-price system has primarily fallen because there has been a significant narrowing of the difference between the prices for upward downward regulating power. The narrowing is primarily due to the fact that in 2004 Eltra improved the scope for buyng regulating power in the other Nordic countries. This has resulted, especially, in increases in the price of downward regulating power. Eltra has thus been able to sell surplus power at a higher price abroad.

The Quantile Regression Theory

2.1 Introduction

In this chapter, we describe the theory about quantiles and Quantile Regression.

Koenker and Bassett [16] proposed an original approach in regression analysis in order to permit estimating quantile functions of a conditional distribution.

After a brief introduction concerning quantiles and conditionals quantiles, in section 2.3 we describe Quantile Regression as linked to Optimization, as seen in [17], mostly, and [19], [10], [16].

In section 2.4 we describe the method of Quantile Regression [17] and in section 2.5 we introduce the Nonparametric Regression Models and the smoothing splines [13], with some guidelines in order to choose the smoothing parameter.

In Section 2.6 we summarize the foremost features of the Local Regression Theory.

In the last section, Section 2.7, we briefly introduce the splines theory [9].

2.2   Quantiles and Conditional Quantiles

The τth quantile of F_Y, given a random variable Y with the distribution function F_Y, denoted as $q_Y(\tau)$, is the solution to $F_Y(q) = \tau$, i.e.

$$q_Y(\tau) = F_Y^{-1}(\tau) = inf\{y : F_Y(y) \geq \tau\}. \tag{2.1}$$

where $\tau \in (0, 1)$.

Example 2.1 *The most important quantiles have a "name":*

- $q_{0.25} \longrightarrow$ *First quartile*

- $q_{0.5} \longrightarrow$ *Second Quartile*

- $q_{0.75} \longrightarrow$ *Third Quartile*

- $q_{0.01} \longrightarrow$ *First Percentile*

- $q_{0.02} \longrightarrow$ *Second Percentile*

- $\ldots$

- $q_{0.99} \longrightarrow$ *99th Percentile*

and, more in general, τth quantile or the $\tau\%$ quantile.

Example 2.2 *In Figure (2.1) it is possible to see the 90% quantile of a Gaussian distribution.*

Now, let X be a random variable and let $F_{Y|X}$ be the conditional distribution of Y conditioned by X.

It is possible to define the τth quantile of $F_{Y|X}(y)$ similarly to (2.1) in the following way:

$$q_{Y|X}(\tau) = F_{Y|X}^{-1}(\tau) \tag{2.2}$$

Example 2.3 *Again, important conditional quantiles are*
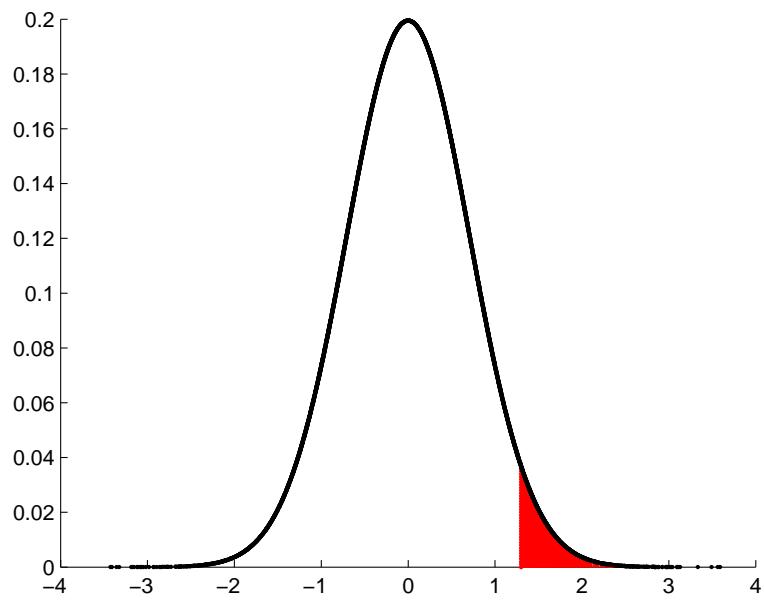
Figure 2.1: 90th percentile for a N(0,1): the colored area corresponds to the 10% of the total area.

- $q_{Y|X}(0.5)$, *the conditional median, represents the center (point of symmetry) of $F_{Y|X}$.*

- $q_{Y|X}(\tau)$, *with $\tau \to 0$, describes the left tail of $F_{Y|X}$.*

- $q_{Y|X}(\tau)$, *with $\tau \to 1$, describes the right tail of $F_{Y|X}$.*

2.3 Quantiles Regression and Optimization

Now, let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables. We introduce the following

Definition 2.1 A *decision function* δ is a measurable function defined on $\mathbb{R}^n$ into $\mathbb{R}$. The value $\delta(y_1, \ldots, y_n)$ of δ at $(y_1, \ldots, y_n)$ is called a *decision*.

Definition 2.2 For estimating a parameter τ, a *loss function* is a nonnegative function which expresses the (financial) loss incurred when τ is estimated by $\delta(y_1, \ldots, y_n)$:

$$L[\tau; \delta(y_1, \ldots, y_n)] = |\tau - \delta(y_1, \ldots, y_n)|, \qquad (2.3)$$

or, more generally,

$$L[\tau; \delta(y_1, \ldots, y_n)] = v(\tau)|\tau - \delta(y_1, \ldots, y_n)|^k, \qquad (2.4)$$

where $k > 0$;

or $L[\cdot; \delta(y_1, \ldots, y_n)]$ is taken to be a convex function of τ.

The most convenient form of a loss function is the *squared loss function*, i.e.

$$L[\tau; \delta(y_1, \ldots, y_n)] = [\tau - \delta(y_1, \ldots, y_n)]^2. \qquad (2.5)$$

In our discussion, it's important to define the following

Definition 2.3 The risk function corresponding to the loss function $L(\cdot; \cdot)$ is denoted by $R(\cdot; \cdot)$ and is defined by

$$R(\tau, \delta) = E_\tau L[\tau; \delta(y_1, \ldots, y_n)] \tag{2.6}$$

i.e. the risk corresponding to a given decision function is the average loss incurred if that decision function is used.

We suppose now that τ is a random variable itself with probability distribution function λ, to be called a *prior* probability distribution function. Then we set

$$R(\delta) = E_\lambda R(\tau, \delta) = \int R(\tau, \delta) \lambda(\tau) d\tau. \tag{2.7}$$

$R(\tau)$ is the *average*, with respect to λ, *risk* over the entire parameter space Ω when the estimator δ is used.

If the observed value of Y_i, is y_j, $j = 1, \ldots, n$, we determine the conditional probabiliy distribution function of τ, given $Y_1 = y_1, \ldots, Y_n = y_n$. This is called the *posterior* probability distribution function of τ $(F_\tau(y))$.

We look for an estimator which minimizes the maximum (over τ) risk. Our attention is restricted to the class of all estimators for which $R(\tau; \delta)$ is finite for all $\tau \in \Omega$.

We consider the following loss function

$$L[\tau; \delta] = \delta(\tau - I(\delta < 0)), \tag{2.8}$$

where $\delta = \delta(y_1, \ldots, y_n)$, for some $\tau \in (0, 1)$.

So, we want to minimize the risk function, i.e. the expected loss function:

$$R(\tau; \delta) = EL[\tau; \delta] = E[\delta(\tau - I(\delta < 0))]. \tag{2.9}$$

Let it be $\delta = Y - \hat{y}$. Then
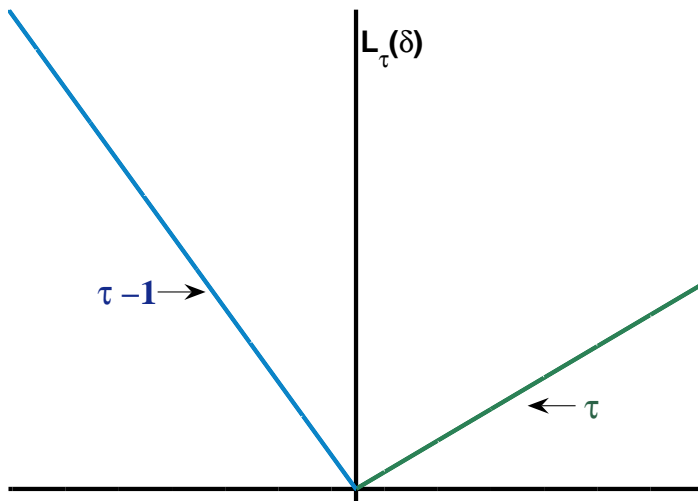
Figure 2.2: $L_\tau(\delta)$ for $\tau = 0.3$

$$R(\tau; \delta) = EL_\tau(Y - \hat{y})$$
$$= \int (Y - \hat{y})(\tau - I_{\{(Y-\hat{y})<0\}}) dF_\tau(y)$$
$$= (\tau - 1) \int_{-\infty}^{\hat{y}} (y - \hat{y}) dF_\tau(y) + \tau \int_{\hat{y}}^{+\infty} (y - \hat{y}) dF_\tau(y)$$

$$(2.10)$$

Now, deriving (2.10) with respect to $\hat{y}$, we obtaine:

$$\frac{\partial R(\tau; \delta)}{\partial \hat{y}} = -(\tau - 1) \int_{-\infty}^{\hat{y}} dF_\tau(y) - \tau \int_{\hat{y}}^{+\infty} dF_\tau(y)$$
$$= -\tau \int_{-\infty}^{+\infty} dF_\tau(y) + \int_{-\infty}^{\hat{y}} dF_\tau(y)$$
$$= -\tau + F_\tau(\hat{y})$$

$$(2.11)$$

Thus

$$\tau = F_\tau(\hat{y}). \tag{2.12}$$

and (2.1).

As we have seen, an optimization problem naturally arises from solving a quantile regression problem.

2.4   The Method of Quantile Regression

Given the data x_t for $t = 1, ..., n$, we consider the following linear model:

$$y_t = x_t'\beta + \epsilon_t \tag{2.13}$$

In order to approximate a particular conditional quantile of y_t, β must be estimated properly.

β_τ, i.e. the τth conditional quantile regression estimator of β, can be obtained by minimizing, with respect to β, the following problem:

$$R_n(\beta) = \frac{1}{n}\left[\tau \sum_{t \in T_1} |y_t - x_t'\beta| \quad + \quad (1-\tau) \sum_{t \in T_2} |y_t - x_t'\beta|\right] \tag{2.14}$$

where $T_1 = \{t : y_t \geq x_t'\beta\}$ and $T_2 = \{t : y_t < x_t'\beta\}$.

We set $\tau = \frac{1}{2}$ in (2.14) :

$$\begin{aligned}
R_n(\beta) &= \frac{1}{n}\left[\frac{1}{2}\sum_{t \in T_1}|y_t - x_t'\beta| + \frac{1}{2}\sum_{t \in T_2}|y_t - x_t'\beta|\right] \\
&= \frac{1}{2n}\sum_{t=1}^{n}|y_t - x_t'\beta|
\end{aligned}$$

$$\tag{2.15}$$

thus

$$R_n^{LAD}(\beta) = 2R_n(\beta) = \frac{1}{n}\sum_{t=1}^{n}|y_t - x_t'\beta| \tag{2.16}$$

where LAD method, Least Absolute Deviation method, is a alternative method to the OLS method, Ordinary Least Squares method.

In OLS the goal is minimizing the Sum of Squared Residuals

$$R_{SSR}(\beta) = \sum_{t=1}^{n}(y_t - x_t'\beta)^2 \tag{2.17}$$

as in LAD the function to be minimized is the Sum of Absolute Residuals

$$R_{SAR}(\beta) = \sum_{t=1}^{n}|y_t - x_t'\beta| \tag{2.18}$$

The regression estimated by the LAD method ($\tau = 0.5$) is in effect a particular case of conditional quantile regression ($0 < \tau < 1$) and is usually referred to as a *median regression.*

Example 2.4 *In figure (2.3) it is possible to see the OLS function and the LAD function for a sample of 1000 data from the following model:*

$$y_t = \beta x_t + \epsilon_t \tag{2.19}$$

where $\epsilon_t \sim N(0,1)$

The (2.14) can then be written in a compact form:

$$R_n(\beta) = \frac{1}{n} \sum_{t=1}^{n} (\tau - \mathbb{I}_{\{y_t - x_t'\beta < 0\}})(y_t - x_t'\beta) \tag{2.20}$$

where $\mathbb{I}_{\{y_t - x_t'\beta < 0\}}$ is the indicator function of the event $\{y_t - x_t'\beta < 0\}$.

The first order condition of minimizing (2.20) is

$$\frac{1}{n} \sum_{t=1}^{n} x_t(\tau - 1_{y_t - x_t'\beta < 0}) = 0 \tag{2.21}$$

except at $y_t = x_t'\beta$ where the derivative is not defined.

The solution of (2.21), solved for β, is $\hat{\beta}_\tau$, the τth quantile regression estimator of β.

2.5   Nonparametric Regression Models

A nonlinear regression model is given by the equation

$$Y = f(\beta, X) + \epsilon \tag{2.22}$$
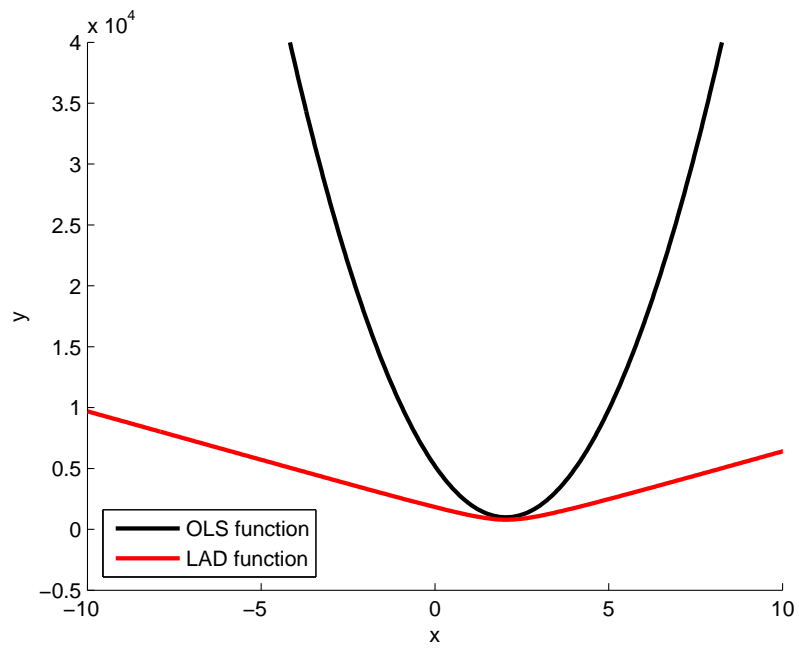
Figure 2.3: OLS and LAD functions for an example of a linear model

where $\beta = (\beta_1, ..., \beta_p)$ is a vector of parameters to be estimated, and $X = (X_1, ..., X_k)$ is a vector of predictors; the errors $\epsilon = (\epsilon_1, ..., \epsilon_s)$ are assumed to be normally and independently distributed with mean 0 and constant variance σ^2. The function $f()$ usually is specified in advance, as it is in a linear regression model.

The *nonparametric regression model*, as the nonlinear regression model, is written in a similar way, but the function f is left unspecified:

$$Y = f(X) + \epsilon = f(X_1, X_2, ..., X_k) + \epsilon \tag{2.23}$$

The object of nonparametric regression is to estimate the regression function $f()$ directly, rather than to estimate the parameters. Most methods of nonparametric regression implicitly assume that $f()$ is a smooth, continuous function.

As in nonlinear regression, it is usually assumed that $\epsilon \sim N(0, \sigma^2)$.

Nonparametric simple regression is often called *scatterplot smoothing* because an important application is to tracing a smooth curve through a scatterplot of Y against X.

One of the most important and most used model is the additive regression model,

$$Y = \alpha + f_1(X_1) + f_2(X_2) + \ldots + f_k(x_k) + \epsilon \tag{2.24}$$

where the partial-regression functions $f_j()$ are assumed to be smooth, and are to be estimated from the data.

Each of the functions can be approximated by linear combinations of known basis functions of the corresponding explanatory variable, i.e.

$$f_j(X_j) = \sum_{k=1}^{n_j} \phi_{jk}(x_j)\theta_{jk} \tag{2.25}$$

where $\phi_{jk}(x_j)$ are the basis functions and θ_{jk} are unknown coefficients.

The resulting model, i.e. the model consisting of (2.24) and (2.25), is a linear regression model. However, the resulting estimates of the functions generally

have larger bias than those based on non-parametric approximations.

A restriction must be imposed on (2.24) (and (2.25)) in order to obtain unique estimates and the resulting basis functions must be derived.

Example 2.5 *If it must be that $f_j(0) = 0$, it follows from (2.25) that*

$$\theta_{j1} = -\sum_{k=2}^{n_j} \theta_{jk} \frac{\phi_{jk}(0)}{\phi_{j1}(0)} \tag{2.26}$$

Setting τ_{j1} as in (2.26) into (2.24) results in the expression

$$f_j(X_j) = \sum_{k=2}^{n_j} (\phi_{jk}(X_j) - \frac{\phi_{jk}(0)}{\phi_{j1}(0)} \phi_{j1}(X_j)) \theta_{jk} \tag{2.27}$$

where the term in front of the coefficients θ_{jk} defines the $n_j - 1$ new basis functions.

If some of the functions in (2.25) are known to be periodic, this restriction can be imposed on the basis functions.

Using cubic spine basis functions, the functions in (2.25) have continuous derivatives up to order two. This property should also be imposed when constructing the periodic basis.

This model is substantially more restrictive than the general nonparametric regression model, but less restrictive than the linear regression model, which assumes that all of the partial-regression functions are linear.

Also *semiparametric models* are kinds of additive regression model, in which some of the predictors enter linearly.

Example 2.6

$$Y = \alpha + \beta_1 X_1 + f_2(X_2) + ... + f_k(x_k) + \epsilon \tag{2.28}$$

Other useful kinds of additive model are models in which some predictors appear as higher-dimensional terms in the model, as we can see in Example 2.7.

Example 2.7

$$Y = \alpha + f_{12}(X_1, X_2) + f_3(X_3) + ... + f_k(X_k) + \epsilon \tag{2.29}$$

2.5.1   Approximation by Smoothing Splines

We consider the Sobolev space W_p^k of real functions on $[0,1]$ with $k-1$ absolutely continuous derivatives and kth derivative existing almost everywhere as a function in $L_p[0,1]$. Let it be k, $k \geq 2$, an integer and p, $p \in [1,\infty)$.

Given the function

$$|f|_p = (|f(x)|^p)^{\frac{1}{p}}, \tag{2.30}$$

our aim is finding the smoothest interpolating function of the points $(x_i, y_i), i = 1,...n$ in the sense of solving

$$\inf\{\|g^{(k)}\|_p : g \in W_p^k, g(x_i) = y_i, i = 1,...n\}. \tag{2.31}$$

For $p = 1$, (2.31) has no solution for $f \in W_k^1$.

If we let W_1^k include functions whose kth derivatives are measures, the expanded problem has a solution: a spline s of degree $k-1$ with measures $s^{(k)}$ concentrated on n or fewer points.

Any solution, $\hat{f}$, must interpolate itself at the observed $\{x_i\}$ and must minimize the *roughness penalty*, subject to a given fidelity constraint [13].

Given that,in order to determine the form of the solution to the *smoothing problem* it suffices to consider the *interpolation problem*.

We have the interesting result in the case $k = 2$.

Theorem 2.4 *Let it be*

$$U^2 = \{g : g(x) = a_0 + a_1 x + \int_0^1 (x-y)_+ d\mu(y), V(\mu) < \infty, a_1 \in \mathbb{R}, i = 0, 1\} \tag{2.32}$$

The function $g \in U^2$ minimising

$$\sum \rho_i\{y_i - g(x_i)\} + \lambda V(g') \tag{2.33}$$

is a linear spine with knots at the points $x_1(i = 1, ...n)$.

Demonstration.

Let f be any interpolator of the points $\{(x_1, y_i) : i = 1, ..., n\}$ with an absolutely continuous first derivative.

The total variation of an absolutely continuous function is the integral of the absolute value of its derivative. Thus, it is possible to choose $u_i \in (x_i, x_{i+1})$ such that

$$f'(u_i) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \tag{2.34}$$

where $i = 1, \ldots, n - 1$.

Then,

$$V(f') \geq \sum_{i=1}^{n-1} | \int_{u_i}^{u_{i+1}} f''(x)dx | \geq \sum_{i=1}^{n-1} | f'(u_{i+1}) - f'(u_i) | = V(\hat{f}'), \tag{2.35}$$

where $\hat{f}$ is the piecewise linear interpolator with knots at the x_i.

For any continuous piecewise linear g exists a sequence of functions $\{g_n\}$ with absolutely continuous first derivative such that $lim V(g_n') = V(g')$, and it is possible to demonstrate that $\hat{f}$ minimises $V(g)$ for all such g.

Consider the simple-regression problem consisting in finding the function $f(x)$ with two continuous derivatives that minimizes the penalized sum of squares

$$SS(\lambda) = \sum_{t=1}^{n} [Y_t - f(X_t)]^2 + \lambda \int_{X_{min}}^{X_{max}} [f''(X)]^2 dX \tag{2.36}$$

where λ is a *smoothing parameter*.

The parameter λ controls the smoothness of the objective function.

When λ is sufficiently large, the solution will be the bivariate linear quantile regression fit ($\hat{f}(x)$ will be selected so that $\hat{f}''(x)$ is everywhere 0).

When λ is sufficiently small, all the n observations will be interpolated if all the X-values are distinct, otherwise the τth quantiles at each distinct design point are interpolated. This is similar to a local-regression estimate with $span = \frac{1}{n}$ (see Section 2.6).

The first term in (2.36) is the residual sum of squares.

The second term in (2.36) is a *roughness penalty*, which is large when the integrated second derivative of the regression function $f''(X)$ is large; i. e., when $f(X)$ is *rough* (with rapidly changing slope).

The function $\hat{f}(X)$ that minimizes (2.36) is a natural cubic spline with knots at the distinct observed values of X as seen in [13].

2.5.2   Selecting the Smoothing Parameter

The adjustable smoothing parameter may be selected by visual trial and error, picking a value that balances smoothness against fidelity to the data. More formal methods of selecting smoothing parameters typically try to minimize the mean-squared error of the fit, either by employing a formula approximating the mean-square error (e.g., so-called *plug in* estimates), or by some form of crossvalidation.

For quantile smoothing splines, the problem of computing a family of solutions for various λ, i. e. the *smoothing parameter*, is greatly eased by the fact that the problem is a parametric linear program in the parameter λ, because the objective function is linear in λ. As λ changes, the orientation of the linear function changes.

The suitable solutions, $\hat{g}_{\tau,\lambda}$, are those functions, piecewise constant in λ, such that $\hat{g}_{\tau,\lambda}$ solves (2.36) $\forall \lambda \in [\lambda_{i-1}, \lambda_i]$, with $0 = \lambda_0 < \lambda_1 < ... < \lambda_j$.

The number of distinct solutions, J, in λ, may be thought in the order of $O_p(nlogn)$ [Portnoy 1991, for the problem of the number of distinct solutions for the linear quantile regression problem]. An important consequence of this is that initially the much smaller linear quantile regression problem corresponding to $\lambda = \infty$ can be solved and gradually the roughness penalty can be relaxed in order to avoid, as much as possible, a direct solution of a large problem.

As it has be said, the solutions of (2.36) depend upon the penalty parameter λ and involve the number of interpolated points. If no two observations share the same design point, the number p_λ of interpolated x_t's must be at least 2 and at most n. Actually, the number of actives knots can be p_λ, i. e. depending on λ.

As seen in [13], the criterion

$$SIC(p_\lambda) = log[n^{-1} \sum_{t=1}^{n} \rho_\tau\{y_t - \hat{g}(x_t)] + \frac{1}{2}n^{-1}p_\lambda logn \qquad (2.37)$$

which may be interpreted as the Schwarz (1978)criterion for the quantile smoothing spline problem, seems to perform well in some limited applications. Machado (1993) considers similar criteria for parametric quantile regression and more general M-estimators of regression.

A valid method to choose the smoothing parameter is **crossvalidation**: the data are divided into subsets (possibly comprising the individual observations), the model is successively fit omitting each subset in turn, and then the fitted model is used to *predict* the response for the left-out subset. Trying this procedure for different values of the smoothing parameter will suggest a value that minimizes the **crossvalidation** estimate of the mean-squared error. Because **crossvalidation** is very computationally intensive, approximations are often employed.

2.6   Local Regression

One of the most common smoothing methods is Local Regression.

Local Regression estimation is based on the principle that a smooth function can be well approximated by a low degree polynomial in the neighborhood of any point. It was introduced in the statistical literature in the late 1970's.

Example 2.8 *A local linear approximation of the function f is*

$$f(x_i) \approx \alpha_0 + \alpha_1(x_i - x)$$

for $x - h \leq x_i \leq x + h$.

Moreover, a local quadratic approximation of the function f is

$$f(x_i) \approx \alpha_0 + \alpha_1(x_i - x) + \frac{\alpha_2^2}{2}(x_1 - x)^2$$

for $x - h \leq x_i \leq x + h$.

The interval $x - h \leq x_i \leq x + h$ is called *smoothing window*, and h is a fixed parameter known as the *bandwidth*. The *span* is the parameter which controls the degree of smoothing.

The local approximation can be fitted by locally *weighted least squares*. In the case of local linear regression, coefficient estimates are chosen to minimize

$$\sum_{i=1}^{n} W(\frac{x_i - x}{h})(Y_i - (\alpha_0 + \alpha_1(x_i - x)))^2. \tag{2.38}$$

where $W(\cdot)$ is a weight function.

The weight function is chosen so that most weight is given to those observations close to the fitting point.

The following function is a common choice for the weight function,

$$W(x) = \begin{cases} (1 - x^2)^2, & -1 \leq x \leq 1 \\ 0, & x < -1 \vee x > 1 \end{cases}$$

Each local least squares problem defines the local linear regression estimate $\hat{f}$ at one point x; if x is changed, the smoothing weights $W(\cdot)$ change, and so the estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$ change.

2.7   Splines

Let $x_0, ..., x_n$ be $n + 1$ distinct nodes of $[a, b]$, with $a = x_0 < x_1 < ... < x_n = b$.

The function $s_k(x)$ on the interval $[a, b]$ is a spline of degree k relative to the nodes x_j if

$$s_k \in \mathbb{P}_k([x_j, x_{j+1}]), \forall j = 1, ..., n - 1 \tag{2.39}$$

$$s_k \in C^{k-1}([a, b]) \tag{2.40}$$
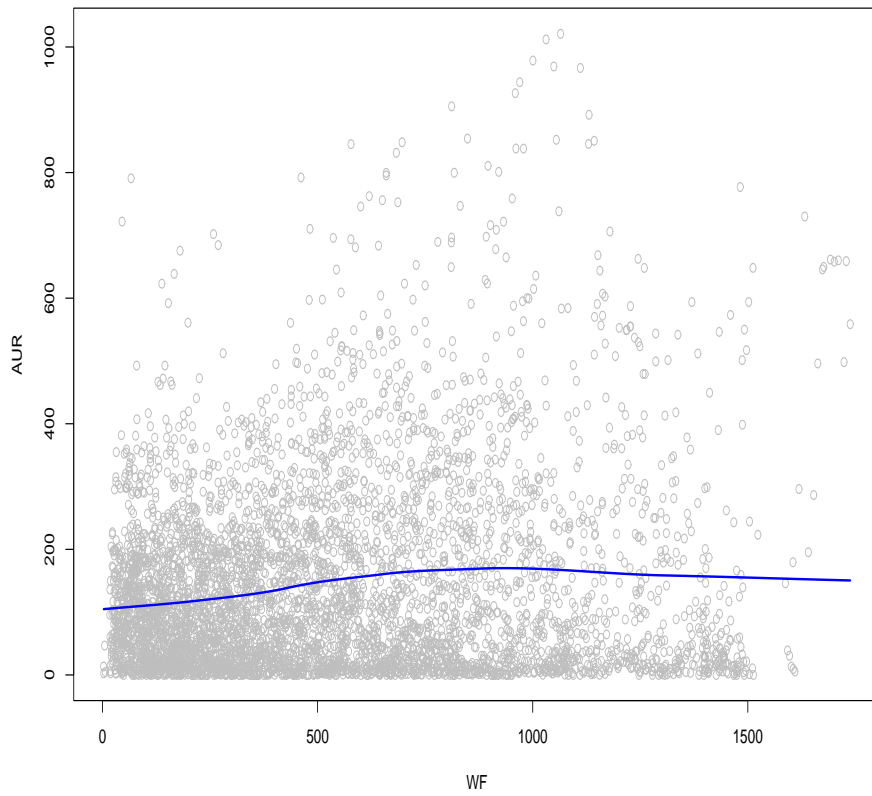
Figure 2.4: The figure shows the local regression model for AUR depending on WF.

Consider the space of splines s_k on [a, b] relative to n + 1 distinct nodes and denote it by S_k, then dim $S_k = n + k$.

Any polynomial of degree k on $[a, b]$ is a spline.

In the practice a spline is represented by a different polynomial on each subinterval and for this reason there could be a discontinuity in its $k - th$ derivative at the internal nodes $x_1, ..., x_n - 1$. The nodes for which this actually happens are called *active nodes*.

Easily, it can be seen that a spline may be represented in the following way:

$$s_{k,j}(x) = \sum_{i=1}^{k} s_{i,j}(x - x_j)^i, \tag{2.41}$$

if $x \in [x_j, x_{j+1}]$

A spline can be conveniently represented using $k + n$ spline basis functions.

Interpolatory cubic splines are particularly significant since they are the splines of minimum degree that yield C^2 approximations and they are sufficiently smooth in the presence of small curvatures.

Let $x_0, ..., x_n$ be $n + 1$ ordered nodes of $[a, b]$, with $a = x_0 < x_1 < ... < x_n = b$ and f_i, $i = 0, ..., n$ the corresponding evaluations. Since the spline wanted to interpolate those values is of degree 3, its second-order derivative must be continuous.

Often the spline is prolonged outside the end points of the interval [a, b] and a and b are treated as internal points. This strategy produces a spline with a "smooth" behavior.

In order to generate s_3 it may be provided a basis $\{\phi_i\}$ for the space S_3 of cubic splines, whose dimension is equal to $n + 3$. The n+3 basis functions $\{\phi_i\}$ may have global support in the interval $[a, b]$ or local support.

The functions ϕ_i, for $i, j = 0, ..., n$, are defined through the following interpolation constraints

$$\phi_i(x_j) = \delta_{ij} \tag{2.42}$$

and

$$\phi_i(x_0)' = \phi_i(x_n)' = 0 \tag{2.43}$$

and two splines, ϕ_{n+1} and ϕ_{n+1}, must be added.

For instance, if the spline must satisfy some assigned conditions on the derivative at the end points, the two added splines must be such that

$$\phi_{n+1}(x_j) = 0 \, j = 1, ..., n \phi_{n+1}(x_0)' = 1 \phi_{n+1}(x_n)' = 0 \tag{2.44}$$

$$\phi_{n+2}(x_j) = 0 \, j = 1, ..., n \phi_{n+2}(x_0)' = 0 \phi_{n+2}(x_n)' = 1 \tag{2.45}$$

In this way the spline becomes

$$s_3(x) = \sum_{i=0}^{n} f_i \phi_i(x) + f_0' \phi_{n+1}(x) + f_n' \phi_{n+2}(x) \tag{2.46}$$

where f_0' and f_n' are two given values. The resulting basis $\{\phi_i, \, i = 0, ..., n+2\}$ is called a *cardinal spline basis*.
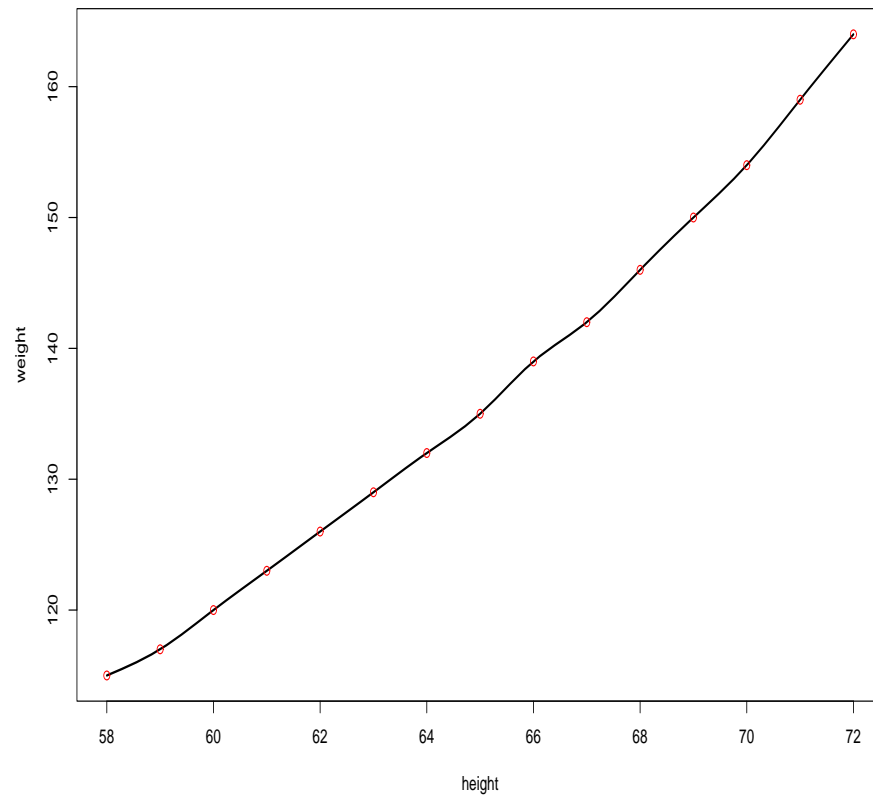
Figure 2.5: The figure shows an interpolation spline between the average heights and weights for American women aged 30-39, i.e. the piecewise polynomial representation of a univariate spline function. The red points are the knots of the spline.

CHAPTER 3

Extreme Value Theory

3.1 Introduction

In this Chapter we introduce the Extreme Value Theory (EVT) in general and we propose a formulation for the EVT.

In section 3.3 we describe the Extreme Value Distribution and in Section 3.5 we briefly summarize the results regarding the Extreme Value Theory for the Nonparametric Extreme Regression Quantiles.

3.2 Introduction to Extreme Value Theory

Extremes are unusual or rare events. In classical data analysis tasks extremes are often labeled as outliers and even ignored. Cutting off extreme data might not be relevant if the aim is to estimate parameters about frequent events. Otherwise, Extreme Value Theory should be applied in order to correctly analyze the data.

Speaking about Extreme Value Theory (EVT) means considering the limiting distributions for the minimum or the maximum of a very large collection of

random observations from the same arbitrary distribution.

Emil Julius Gumbel, a German mathematician, pacifist and anti-Nazi campaigner, developed new distributions in the 1950s. He was a pioneer in the application of extreme value theory. The first application of the extreme value distributions (the Gumbel distribution, the Generalized Extreme Value distribution and the Generalized Pareto Distribution (GPD)) was to answer environmental questions, quickly followed by the finance industry.

It has been showed by Gumbel that for any initial distribution (well-behave distribution, i.e., $F(x)$ is continuous and has an inverse), only a few models are needed, depending on whether you are interested in the maximum or the minimum, and also if the observations are bounded above or below.

We often encounter extreme value distribution s for the minimum in the context of reliability modeling.

The potential of Extreme Value Theory applied to financial problems has only been recognized recently. The end of the last decade has been characterized by significant instabilities in financial markets worldwide. This has led to numerous criticisms about the existing systems for risk management and motivated the search for more appropriate methodologies able to cope with rare events of heavy consequences.

The problem is to model very rare phenomena, which mainly lie outside the range of available observations. EVT provides a confirmed theoretical foundation on which we can build statistical models describing extreme events.

3.3 Extreme Value Distributions

The family of extreme value distributions can be subsumed under a single parametrization known as the Generalized Extreme Value distribution (GEV). The d.f. of the standard GEV is

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-\frac{1}{\xi}}}, & \xi \neq 0 \\ e^{-e^{-x}}, & \xi = 0 \end{cases}$$

where x is such that $1 - \xi x > 0$ and ξ is known as the shape parameter (see [8]).

Three well known distributions are special cases of the ones above, as follows (see [11]).

If $\xi > 0$ we have the Frechet distribution (for the maxima) with shape parameter $\alpha = \frac{1}{\xi}$, given by

$$F(x) = \begin{cases} e^{-(\frac{x-\lambda}{\delta})^\alpha}, & x \geq \lambda \\ 0, & otherwise \end{cases}$$

where λ and β are constants known as the *location* and the *scale* parameter, such that $\delta > 0$ and $\beta > 0$.

If $\xi > 0$ we have the Weibull distribution (for the maxima) with shape parameter $\alpha = -\frac{1}{\xi}$, given by

$$W(x) = \begin{cases} e^{-(\frac{\lambda-x}{\delta})^\alpha}, & x \leq \lambda \\ 1, & otherwise \end{cases}$$

where, again, the location parameter λ and the scale parameter δ are such that $\delta > 0$ and $\beta > 0$.

If $\xi = 0$ we have the Gumbel distribution (for the maxima), given by

$$G(x) = e^{-e^{(\frac{\lambda-x}{\delta})}}, \quad -\infty < x < +\infty; \delta > 0$$

3.4 Extreme Value Theory: formulation

Given a sequence of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution F, we consider

$$M_n = max(X_1, X_2, \ldots, X_n) \tag{3.1}$$

Recall that two distributions, F_1 and F_2 belong to the same family of distribu-

tions if there exist two constants c and d such that

$$F_2(cx + d) = F_1(x) \tag{3.2}$$

Suppose further that there exist two sequences of real numbers $a_n > 0$ and b_n such that, if $n \to +\infty$,

$$Pr\left(\frac{M_n - b_n}{a_n}\right) \to G(x) \tag{3.3}$$

for some non-degenerate distribution function $G(x)$. If this condition holds we say that F is in the maximum domain of attraction of G and we write $F \in MDA(H)$.

Further, G belong to one of the family distributions defined above:

- Frechet

- Weibull

- Gumbel

This was shown by Fisher and Tippett (1928):

$F \in MDA(G) \Rightarrow G$ is of the type G_ξ for some ξ.

Distributions in the maximum domain of attraction of the Gumbel $(MDA(G_0))$ include the normal, **exponential**, gamma and lognormal distributions.

In this work, we are interested in exponential distribution with two parameters.

3.5   Nonparametric Extreme Regression Quantiles

We consider the the nonparametric model

$$y = g_\tau(x) + \epsilon \tag{3.4}$$
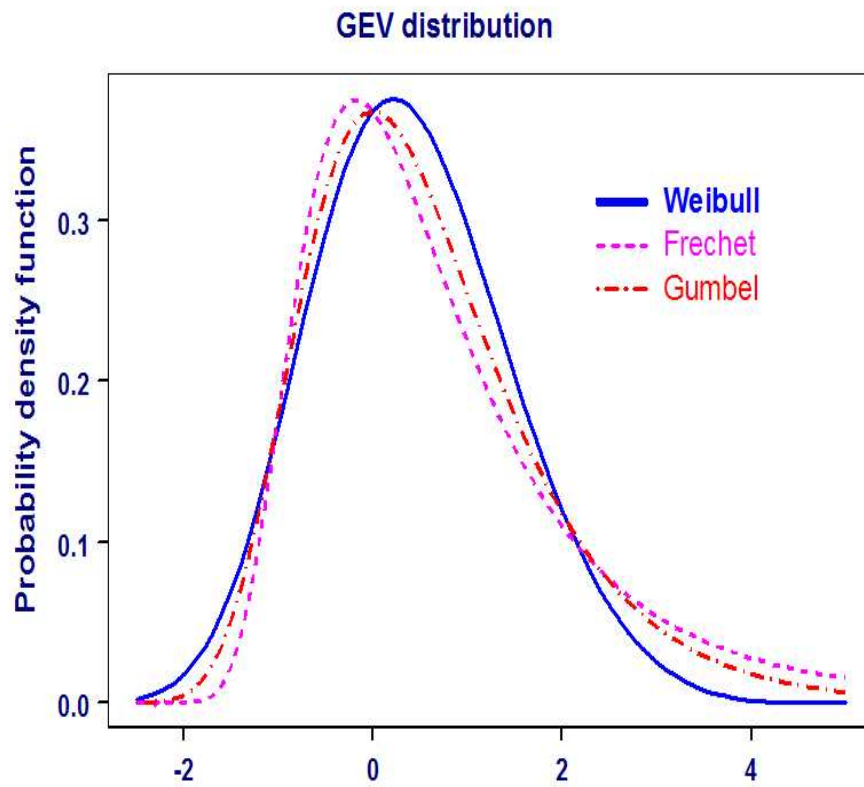
Figure 3.1: The three GEV distributions: the Gumbel (with a light upper tail and positively skewed); the Frechet (with a heavy upper tail and infinite higher order moments) and the Weibull (with a bounded upper tail).

where $g_\tau(x)$ is the condizional τth quantile function of y given x, $g(x) = inf\{z : P(y > \bar{x}|x = \bar{x}) > 0\}$, and $\epsilon \geq 0$.

As seen in Chapter 2, the τth quantile regression estimator of $g_\tau(x)$ is defined as a solution to the optimization problem (see Section 2.3).

Naturally, we aim to view the extreme quantiles as a limit of the non extreme quantiles, and hence

$$g_1(x, \beta) = \lim_{\tau \to 1} g_\tau(x, \beta) = \lim_{\tau \to 1} F_y^{-1}(\tau|x = \bar{x}) \tag{3.5}$$

We consider again

$$y = g(x) + \epsilon \tag{3.6}$$

g is an unknown smooth function. The task is to estimate g and its derivatives for the extreme cases.

Usually the estimator is the locally polynomial estimator (see Section 2.6). In [25] it has been shown that, under certain regularity conditions, the estimator is a consistent pointwise estimator. Moreover, under similar global regularity conditions, the global estimator of g and its derivatives is a consistent estimator in the L_q, $q \in [1, +\infty]$, norm.

Nonparametric Quantile Regression with R

4.1 Introduction

R is a system for statistical computation and graphics. Since 1997 there has been a core group (the "R Core Team") who can modify the R source code archive. The group currently consists of Doug Bates, John Chambers, Peter Dalgaard, Robert Gentleman, Kurt Hornik, Stefano Iacus, Ross Ihaka, Friedrich Leisch, Thomas Lumley, Martin Maechler, Duncan Murdoch, Paul Murrell, Martyn Plummer, Brian Ripley, Duncan Temple Lang, Luke Tierney, and Simon Urbanek.

R consists of a language plus a run-time environment with graphics, a debugger, access to certain system functions, and the ability to run programs stored in script files. R is an integrated suite of software facilities for data manipulation, calculation and graphical display.

The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. Most of the user-visible functions in R are written in R. The R distribution contains functionality for a large number of statistical procedures. Among these are nonlinear

regression models and non parametric regression model.

R is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in R are essentially written for a single piece of data analysis.

The development model is similar to that of the Linux operating system. Like Linux, R is an "open source" system. Source-code is available for inspection or for adaptation to other systems.R provides a language environment that is attractive for the development of new scientific computational tools. The R system may struggle to handle very large data sets. R can process a data set containing one hundred thousand observations and twenty variables.

In this Chapter we briefly introduce the R tools used in this thesy. All the information are available on The Comprehensive R Archive Network (CRAN) at http://cran.r-project.org/.

The information about the packages and the functions implemented in this thesy come by the online documentation.

4.2   The R software in the project

The R packages used in the project are:

- SparseM
- splines
- tripack
- quantreg
- akima
- rgl

The first four packages are needed to implement the function *rqss*. The last two packages are graphic tools required to plot contour curves in 2 and 3 dimensions.

The *splines* packages includes regression spline functions and classes.

The *akima* package implements linear or cubic spline interpolation functions for irregular gridded data.

The *rgl* package includes a 3D visualization device system.

To install a package from R one simply types, once that R is running,

> install.packages("packageName")

and, then, it is possible to make the package accessible to the current R session by the command

> library("packageName")

4.2.1 The SparseM package

A sparce matrices is a matrices with a high percentage of zero entries.

In particular, in using smoothing splines (as in Nonparametric Quantile Regression) design matrices are extremely sparse often with less than 1% of nonzero entries. Developments in sparse linear algebra have produced efficient methods for handling unstructured sparsity in an efficient way.

The package SparseM provides some basic linear algebra functionality for sparse matrices as *coercion*, *basic unary* and *binary operations on matrices* and *linear equation solving*.

The Cholesky factorization and backsolve routines are based on Ng and Peyton (1993).

This packages admitte more than twenty different storage formats for sparse matrices. The primary storage mode for SparseM is the one named *compressed sparse row* (**csr**) format. An n by m matrix A with real elements a_{ij} , stored in **csr** format consists of three arrays:

- **ra**: a real array of *nnz* elements containing the non-zero elements of A, stored in row order. Thus, if $i < j$, all elements of row i precede elements from row j. The order of elements within the rows is immaterial.

- **ja**: an integer array of *nnz* elements containing the column indices of the elements stored in **ra**.

- **ia**: an integer array of n+1 elements containing pointers to the beginning of each row in the arrays **ra** and **ja**. Thus ia[i] indicates the position in the arrays **ra** and **ja** where the ith row begins. The last $(n+1)$st element of **ia** indicates where the $n+1$ row would start, if it existed.

Obviously, the chosen of the smoothing parameter λ for the Nonparametric Regression Quantile influences the sparsity of the matrix.

4.2.2 The quantreg package

The package quantreg is an implementation of statistical methods, named Quantile Regression, for estimating and plotting inferences about conditional quantile functions.

In order to implements Nonparametric Regression Model and Smoothing Spline (see Chapter 2, Section 2.5), it is available the *rqss* function of the quantreg package. This is based upon total variation penalty methods for fitting univariate and bivariate functions.

The most general class of models that the function *rqss* can manage is the following

$$Y = f(\beta, X) + \epsilon \tag{4.1}$$

where f is an unknown function (i.e. the dependence of the Y on $X = X_1, ..., X_k$ is not known) and ϵ represents independent identically distributed errors with mean zero and variance σ^2.

It is possible to estimate f by using an additive models. Models of this type can be expressed as

$$Y = \alpha + f_1(X_1) + f_2(X_2) + \dot{+} f_k(x_k) + \epsilon \tag{4.2}$$

Again, the Additive Quantile Regression Smoothing is explicated in R by the function *rqss*. This function fitting additive quantile regression models with possible univariate and bivariate nonparametric terms estimated by total variation regularization.

In order to use this function, we write a command line in the following way

$$rqss(formula, tau = 0.5, \dots, method = "fn") \tag{4.3}$$

formula is a object with the response on the left of a $\sim$ operators and terms, separated by + operators, on the right. The terms may include *qss* terms that represent additive nonparametric components. These terms can be univariate or bivariate.

$$qss(x, constraint = "N", lambda = 1) \tag{4.4}$$

lambda represents the smoothing parameter governing the tradeoff between fidelity and the penalty component for this term.

tau represents the quantile to be estimated, this must be a number between 0 and 1. By default, tau is set to be equal to 0.5.

There are currently two algorithmic methods to choose between in order to compute the fit. Both are implementations of the Frisch-Newton interior point method described in detail in Portnoy and Koenker(1997). They are implemented using sparse Cholesky decomposition as described in Koenker and Ng (2003) Option "sfnc" is used if the user specifies inequality constraints. Option "sfn" is used if there are no inequality constraints. Linear inequality constraints on the fitted coefficients are specified by a matrix R and a vector r, specified inside the qss terms, representing the constraints in the form $Rb \geq r$.

Qualitative constraints may also be specified for the qss terms ("N","I","D","U","C" "UI","UD","CI","CD" for none, increasing, decreasing, convex, concave, convex and increasing, etc). By default, there are no constraints set in the model.

The function returns a fitted object representing the estimated model specified in the formula.

CHAPTER 5

The Model

5.1 Introduction

After a brief description of the data to be analyzed, we start modeling them with a linear quantile regression model, seeing that it's not enough in order to describe precisely the dependency of Actual Upward Regulating power and Wind Forecasts.

Applying the nonparametric quantile regression theory, we model the data, choosing a good smoothing parameter. We show how much the smoothing parameter choice affects the resulting model.

5.2 Data Description

The data used in this thesys were provided by *Energinet.dk*. Those data, referring to the the period since the 1th of January 2005 until the 31th of January 2005, include

- Wind Forecasts (**WF**) (MWh/h)

The wind power forecasted one day in advanced by simulation programs developed by Energinet.dk.

- Actual Upward Regulating power (**AUR**) (MWh/h)
  The right quantity of ancillary services and regulating power that should have been bought by Energinet.dk.

- Upward Regulating power (**UR**) (MWh/h)
  The quantity of ancillary services and regulating power that have been actually bought by Energinet.dk.

Moreover, we used in this project the data regarding the actual electric energy Consumptions (**C**) (MWh/h) in Denmark, obtained by the NordPool website (http://www.nordpool.com/).

The data, divided per month, were collected in a "one year file" and then elaborated.

The former timetable included five minutes AUR, UR and C data for all the 2005. In order to make the calculus and the analysis smoother, the data were moved from the five minutes (5M) schedule to the one hour schedule (1H) processing them by the *sample mean*.

Given a sample of size n, we consider n random variables $X_1, X_2, \cdot, X_n$, each corresponding to one randomly selected observation. The *sample mean* is defined to be

$$\hat{x} = \frac{1}{n}(X_1 + X_2 + \cdot + X_n) \tag{5.1}$$

In our case $n = 12$ and we consider, to calculate the sample mean, all the 5M observations per 1 hour.

The **AUR**, **UR**, **WF** and **C** data regarding the wind storm on January 2005, the 8th, in Denmark, were deleted (from 12.00 am to 12.00 pm) from the file in order to avoid to bias the results.

As we can easily see from the first plot in figure (5.1), there was a "peak" corresponding to the 8th January data.

In the second plot in figure (5.1) we can see the AUR power data without the 8th January data.
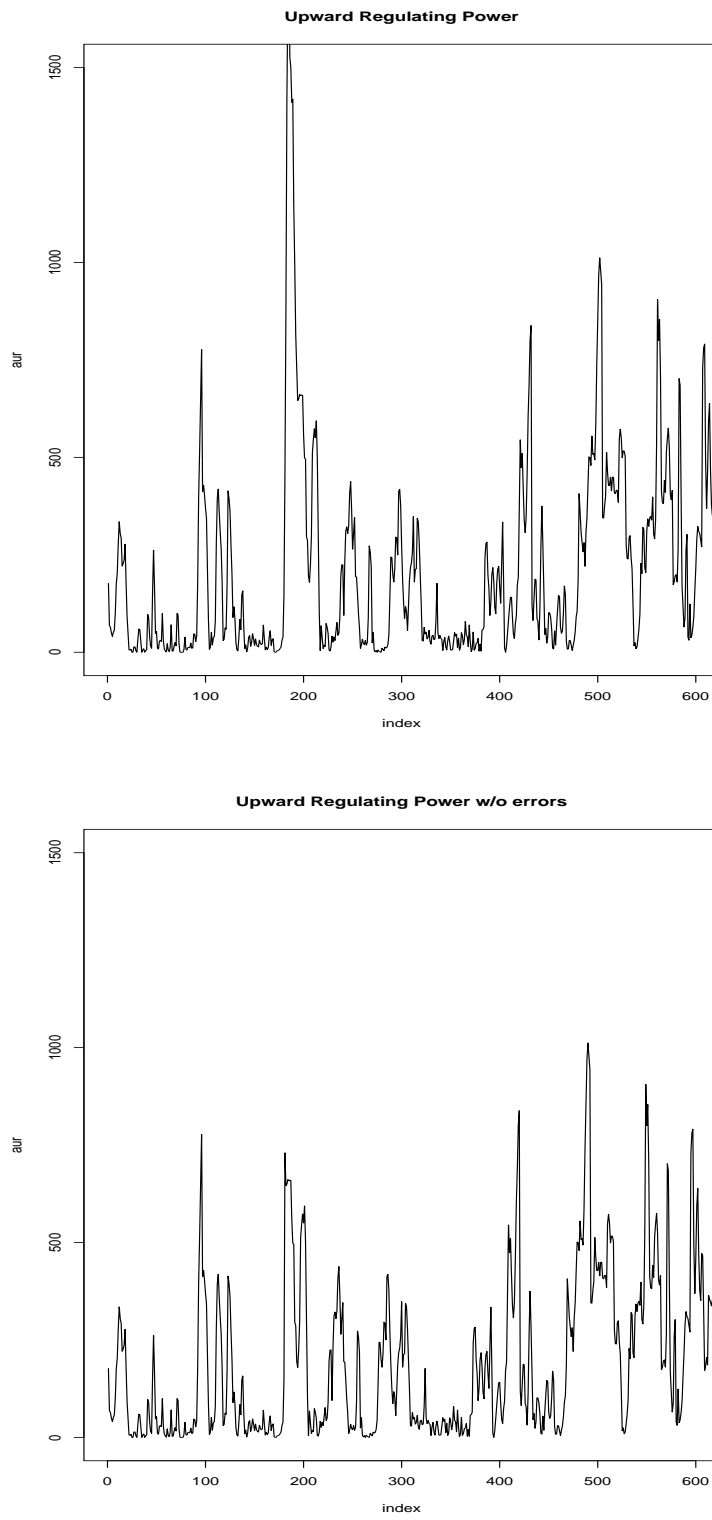
Figure 5.1: Above: AUR data before the corrections. Below: AUR data after the correction.
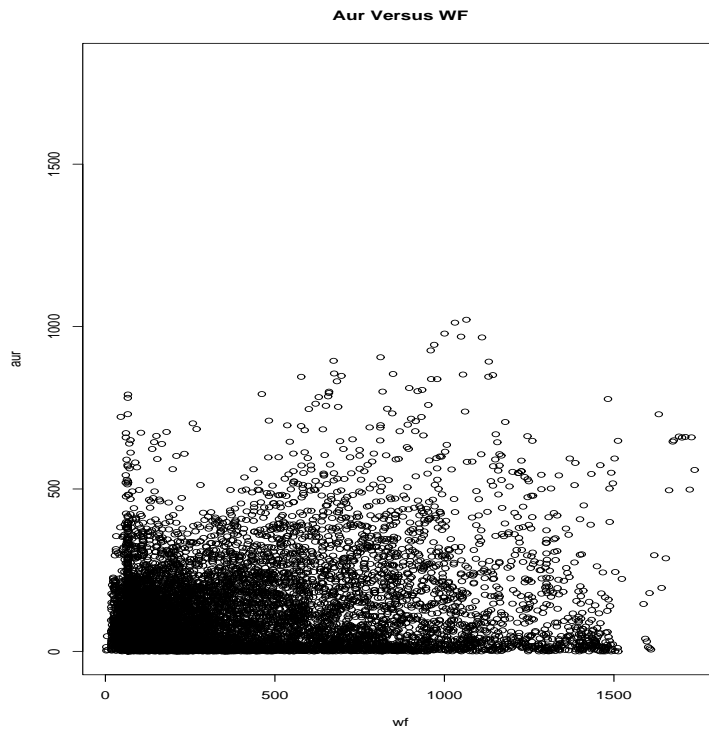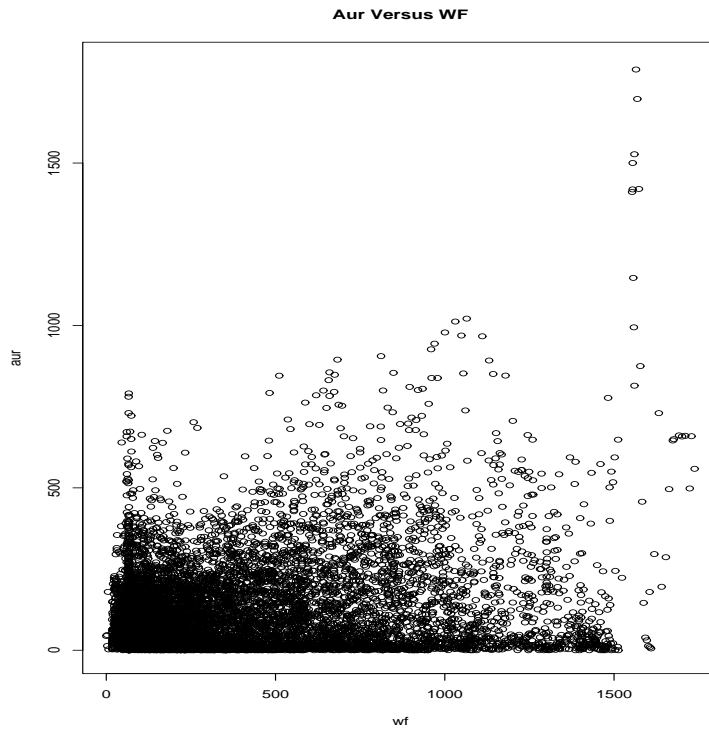
Figure 5.2: Above: scatterplot of AUR data versus WF data before the corrections. Below: scatterplot of AUR data versus WF data after the corrections.

The **AUR**, **UR**, **WF** and **C** data, regarding the hour changing on the 31st of October, were deleted as well in order to avoid duplication of the data.

Then, the data were divided in three main periods:

- January - February - March - April

- May - June - July - August

- September - October - November - December

5.3   Linear Quantile Regression

The first guess, in order to model the data, was to consider a linear quantile regression model to illustrate the dependency between the wind forecasts and the actual upward regulation.

We considered the linear quantile regression model

$$Q_{Y_i}(\tau|x_i) = x_i\beta(\tau) \tag{5.2}$$

where there are only a finite set of distinct covariate settings $\{x_i : i = 1, \ldots, N\}$ and one measurement of the response Y_i at each x_i.

In the case-studying we set $x_i = WF_d$, $Y_i = AUR_d$, $\forall d = 1, \ldots, N$, where $N = 8747$ and d is the index indicating the day and the hour of the measurement.

In order to univocally indicate the day and the hour of the measurements (for the actual upward regulation) and the forecasts (for the wind power) we used a number between 732313 (1st January 2005, 00:00) and 732677.958 (31st December 2005, 23:00), as used in the software Matlab.

We have implemented in R the function rq setting the parameters in the following way

$> fit = rq(y \sim x, tau = taus)$

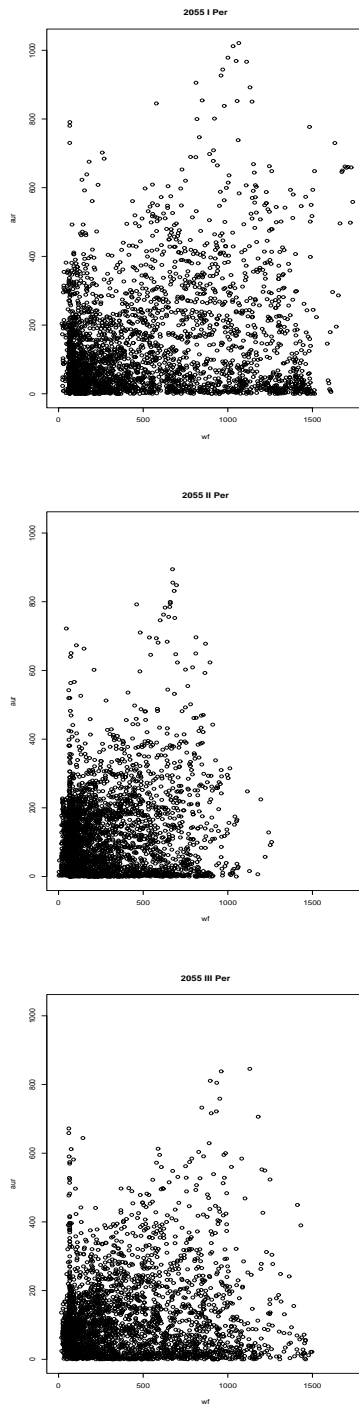where $taus = \{0.05, 0.1, 0.25, 0.75, 0.9, 0.95\}$.

Figure 5.3: Scatterplots of the AUR data versus the WF data after being divided in three main periods. Starting on the left: data referring to the I period (January - February - March - April), data referring to the II period (May - June - July - August), data referring to the III period (September - October - November - December).

Moreover we implements the function lm (used to fit linear models) and, again, rq with $tau = 0.5$.

As we can see in figure (5.5), a linear regression model can't fully illustrate the dependency between the explanatory variable, **WF**, and the response variable, **AUR**. The effects on the response variable are not linear, so in this case the linear quantile regression model fails.

5.4 Nonparametric Additive Quantile

As second step, the quantiles have been modeled as a sum of non-linear smooth functions of the explanatory variables: **WF** (wind forecasts), **D** (time index corresponding to the the day and the hour of the measurement) and **C** (consumptions), following the methodology seen in [19].

We consider the model

$$y = \alpha + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p) + \epsilon \qquad (5.3)$$

where f_i and α can be estimated based on data. The estimated function $\hat{f}_i(x_i)$ can reveal possible nonlinearities in the effect of the x_i on the response variable.

Moreover, in this kind of models, it is possible to mix linear terms and nonlinear terms in a very flexible way, as seen in [21], approximating each of the smooth function as a linear combination of basis functions depending only on known quantities (using splines bases with 10 degrees of freedom).

We imposed the functions to be zero at the lower boundary knots by using natural spline bases without intercept. We placed the boundary knots at the limits of the data and the internal knots according to the quantiles of the individual explanatory variables. As said in [19], in this way the model allows for more flexibility where the observations are relatively dense.

In Figure 5.5 we can see the resulting model for the 90% quantile. The figure shows the effect of each variable. We can easily see that the consumption has the smallest effect on the model (the curve is flatter than the other ones). Indeed, in Figure 5.6, where the quantile is estimated after excluding the consumptions from the model, it's sufficiently possible to see the small changes in the estimated quantile without the consumptions in the model.
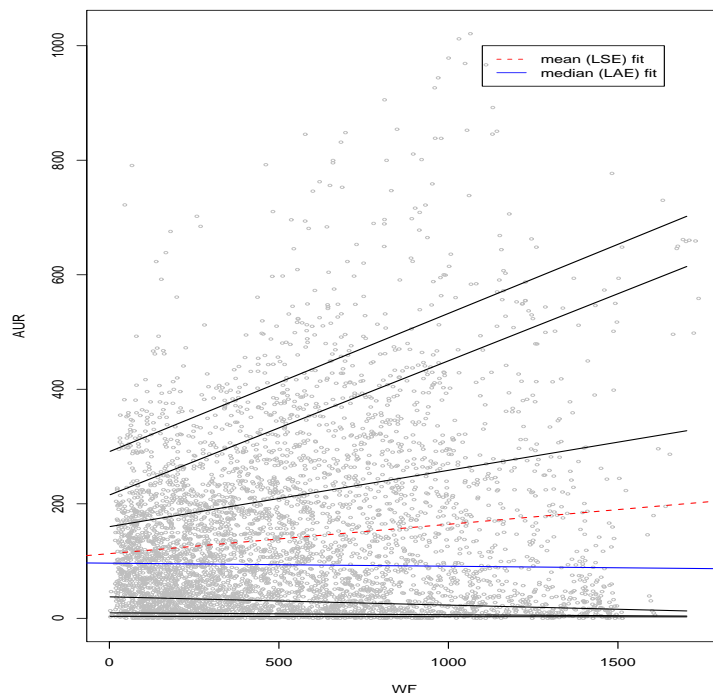
Figure 5.4: Estimated quantiles of the explanatory variable, in comparison with the MEAN fit and the MEDIAN fit
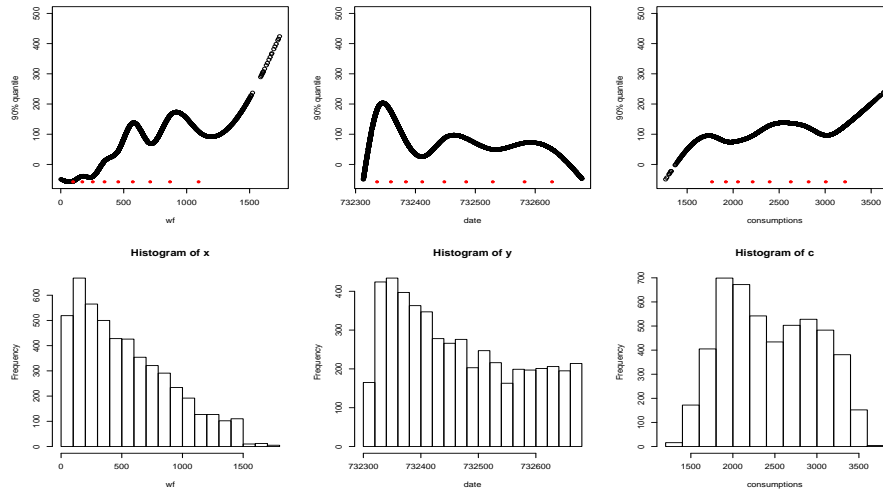
Figure 5.5: Estimated 90% quantile (top row) and histograms (bottom row) of the explanatory variable. The red bullet in the top row of plots indicate the placement of knots
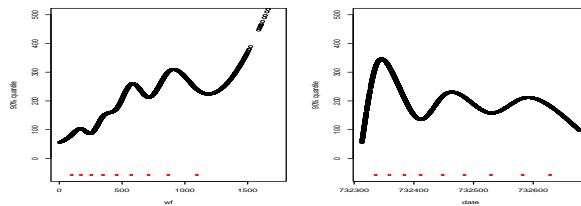

Figure 5.6: Estimated 90% quantile when excluding the consumptions **C** from the model. The red bullet in the top row of plots indicate the placement of knots
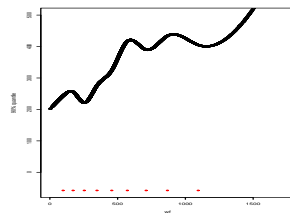

Figure 5.7: Estimated 90% quantile when excluding the consumptions **C** and the time index **D** from the model. The red bullet in the top row of plots indicate the placement of knots

In other respects, in Figure 5.7 we can easily see how the model changes after excluding the explanatory variable **D** (time index) from the model itself. The curve is much "flatter" and less significant than the previous one, in which the time index was part of the model as one of the explanatory variables.

After these considerations we decide to keep in the model, as explanatory variables, the Wind Forecasts (**WF**) and the Time Index (**D**).

In the next Chapters, we will describe the following models:

- a basic model in which the only explanatory variable is Wind Forecasts (**WF**)

- a full model in which there are all the explanatory variables seen earlier: the Wind Forecasts (**WF**) and the Time Index (**D**).

A Basic Model with one explanatory variable: the Wind Forecasts

6.1 Introduction

We analyze the data modeled on a nonparametric regression model with one explanatory variable, the Wind Forecasts. The response variable is the Actual Upward Regulating power.

6.2 The Model

First, we consider the following model

$$AUR = \alpha + f(WF) + \epsilon \qquad (6.1)$$

Using the function *rqss* in R, we want to estimate the quantile smoothing spline

which minimise

$$R_{\tau,\lambda}(f_P) = \sum_{i=1}^{n} \rho_\tau \{aur_i - f_P(wf_i)\} + \lambda \int_0^1 |g''(wf)| dwf \qquad (6.2)$$

The knots for the optimal spline coincide with the observed wf_i (see [13]).

The *rqss* function automatically removes the repetitions in the vector of the observed value. In this way from a vector of length 8747 the function obtain a vector of length 4991 for the observed wf.

Remember that λ is a non-negative smoothing parameter that must be chosen by the data analyst, as we said in Subsection 2.5.2. λ governs the tradeoff between the goodness of fit to the data and smoothness of the function (see [21]). Larger values of λ force f to be smoother. In fact, the interpolating curve corresponds to $\lambda = 0$ at one extreme, and the straight line fit is the limit as $lambda \to \infty$, as we can easily see in the plots in Figure 6.1.

Anyway, for any value of λ, the solution to (6.2) is a linear spline (see [13]).

The problem of computing a family of solutions for various values of λ is a parametric linear program in the parameter λ, in fact, the objective function, solution of the minimising problem in (6.2), is linear in λ.

As said in [13], as λ changes, the orientation of the linear function changes, and we move from one adjacent vertex of the constraint polytope to the next. The solutions $\hat{g}_{\tau,\lambda}(\cdot)$ are thus piecewise constant in λ; that is there exists a mesh $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_j$, such that $\hat{g}_{\tau,\lambda}(\cdot)$ solves the minimising problem for all $\lambda \in [\lambda_{i-1}, \lambda_i]$.

In Figure 6.2 we can see the estimated 80% quantile of the **WF** for very different value of the smoothing parameter λ. For value of λ too small, the estimated quantile function is not enough smooth. Otherwise, for bigger values of the smoothing parameter the estimated quantile function is too smooth and, as we can see in the last plot, for $\lambda = 20000$ the estimated quantile function is a line, loosing in this way many relevant informations.

In order to keep a good balance between the influence of the penalty term and the influence of the fidelity term, we choose the smoothing parameter in the
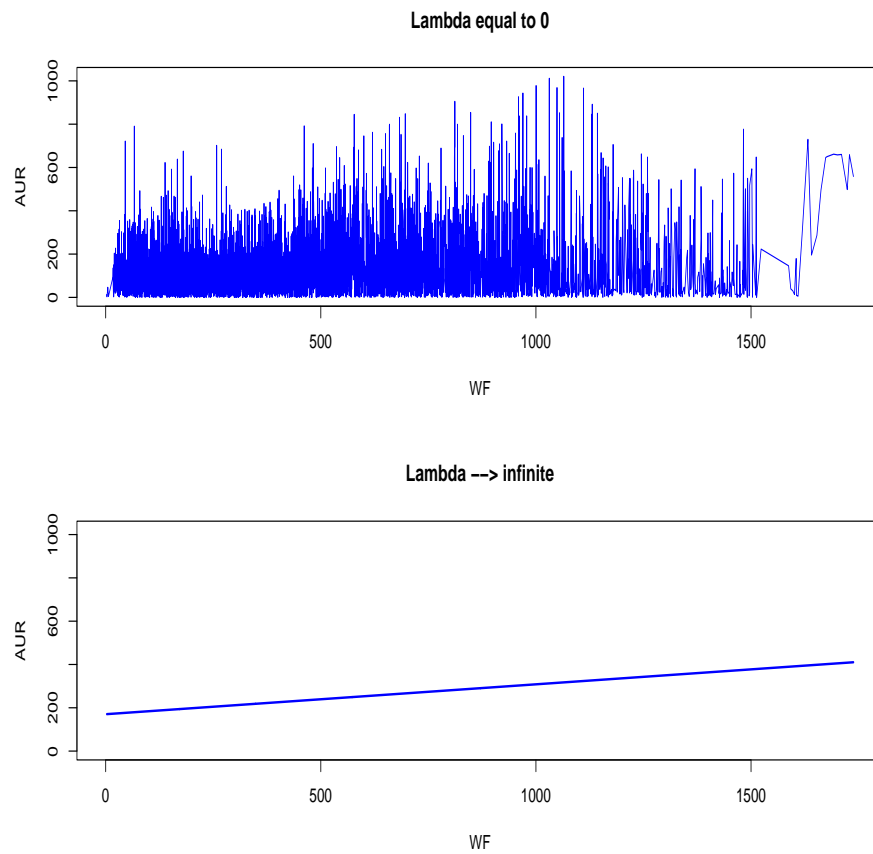
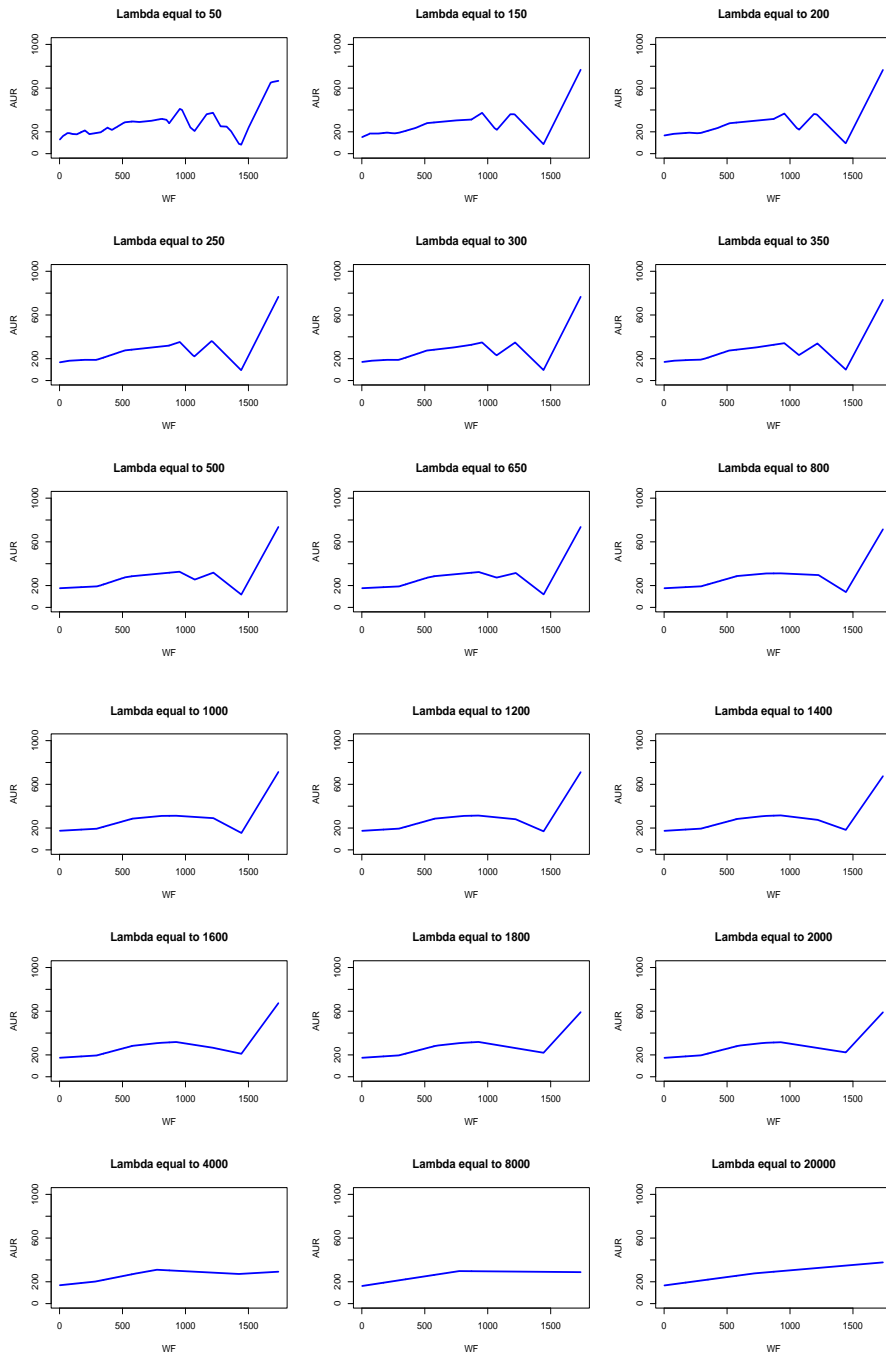Figure 6.1: Estimated 80% quantile of **WF** for two different values of the smoothing parameter *lambda*.

Figure 6.2: Estimated 80% quantile of the WF for different value of the smoothing parameter λ.

following way:

$$\lambda = 650 \tag{6.3}$$

In Figure 6.3 we can see the 80% estimated quantile on the scatterplot of points (wf_i, aur_i), set $\lambda = 650$.

In Figure 6.4 we fit the model for the following values of τ:

$$\tau = (0.5, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95) \tag{6.4}$$

6.2.1 Introducing the time index D in the basic model

In order to better understand the influence of the time index **D** on the dependency of **AUR** on **WF**, we divided the data referring to all the year 2005 in three periods. Then, we added the time index to the model rewriting the model in (6.1) as follows:

$$AUR_P = \alpha_P + f_P(WF) + \epsilon \tag{6.5}$$

where P refers to either the I, the II or the III period.

As we can see in Figure 6.5 the three periods estimated quantile function present some regularities. To better understand that, we plot all the three periods curves together in Figure 6.6.

The first and the third period presents similarities with the all year data plot. The second period presents lightly differences.

Again, we see that the time index is an explanatory variable that has to me added to the model, in order to make it more coherent with the real situation.

6.2.2 Days of the week

Looking for regularities depending on the Time Index, we split up the data by the days of the week.
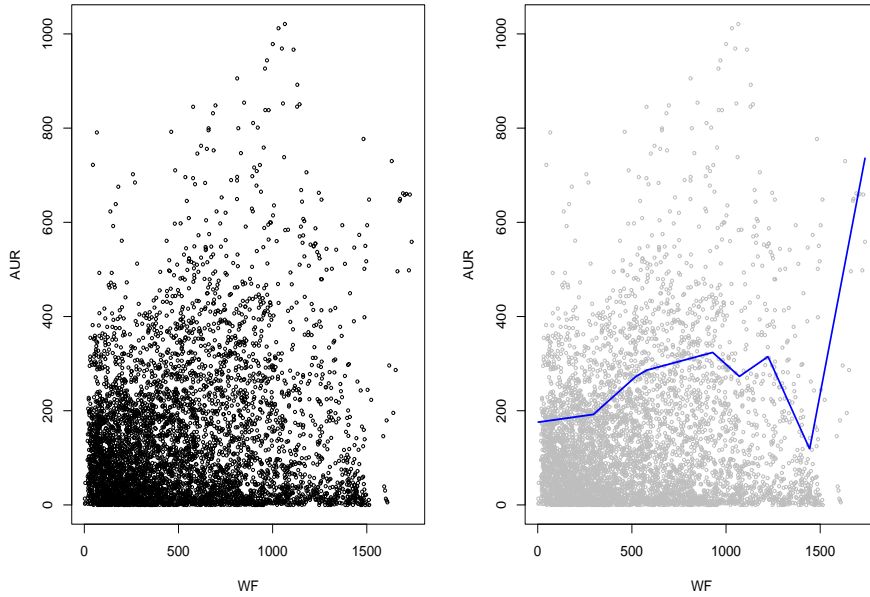
Figure 6.3: The plot on the left is a scatterplot of **AUR** plotted against the predictor **WF**. In the plot on the right, we added a linear smoothing spline (estimating the 80% quantile, with $\lambda = 650$) to describe the trend of **AUR** on **WF**.
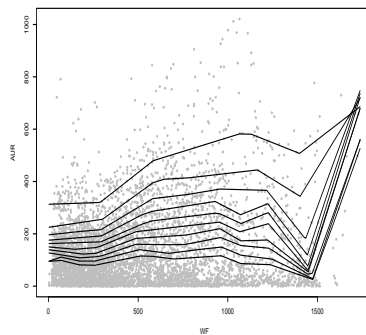

Figure 6.4: Scatterplot of **AUR** plotted against the predictor **WF** in which we added linear smoothing splines estimating different $\tau\%$ quantile. $\lambda = 650$.
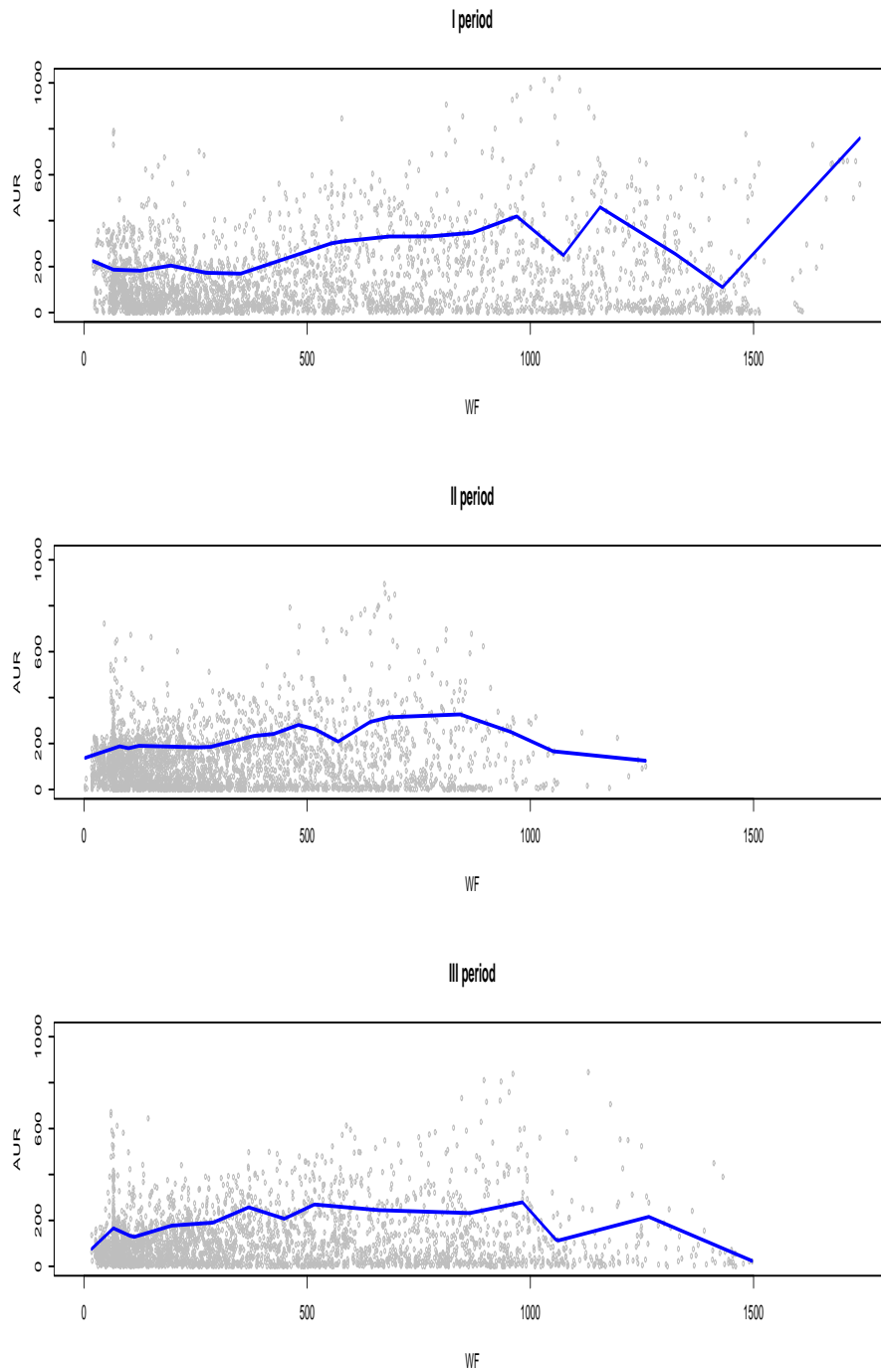
Figure 6.5: The plot on the top is a scatterplot of **AUR** plotted against the predictor **WF** referred to the I period, where we added the estimated 80% quantile. The plot in the middle is a scatterplot of **AUR** plotted against the predictor **WF** referred to the II period, where we added the estimated 80% quantile. The plot on the bottom is a scatterplot of **AUR** plotted against the predictor **WF** referred to the III period, where we added the estimated 80% quantile. In each plot $\lambda = \frac{650}{3}$.

In Figure 6.7 we can see that actually there are some regular features recurring for each day of the week, even if the small number of data (the number of the 2005 data divided by seven is around 1200 data for each day of the week) doesn't allow us to get several conclusions.

Anyway, we can say that the similar features are more evident for small value of the Wind Forecasts, while, for higher values of the Wind Forecasts, the regularities are less evident, this fact probably due to the *rqss* model estimation, not so reliable with such a small number of data.

In Figure (6.8) we can see all the estimated regression quantile functions for each day of the week on the same plot, for the estimated 80% quantile and the estimated 90% quantile.

6.2.3   Dependency of the smoothing parameter on the number of data

Consider again (6.2), if we increase the number of the data from n to N, we can rewrite the equation in the following way:

$$R_{\tau,\lambda}(f_P) = \sum_{i=1}^{N} \rho_\tau \{aur_i - f_P(wf_i)\} + \Lambda \int_0^1 |g\prime\prime(wf)| dwf \qquad (6.6)$$

where

$$\Lambda = N\lambda \qquad (6.7)$$

thus

$$\lambda = \frac{\Lambda}{N} \qquad (6.8)$$

In this way, it is possible to obtain the new value of the smoothing parameter λ when we want to increase or decrease the dimension of the sample that we are working on.
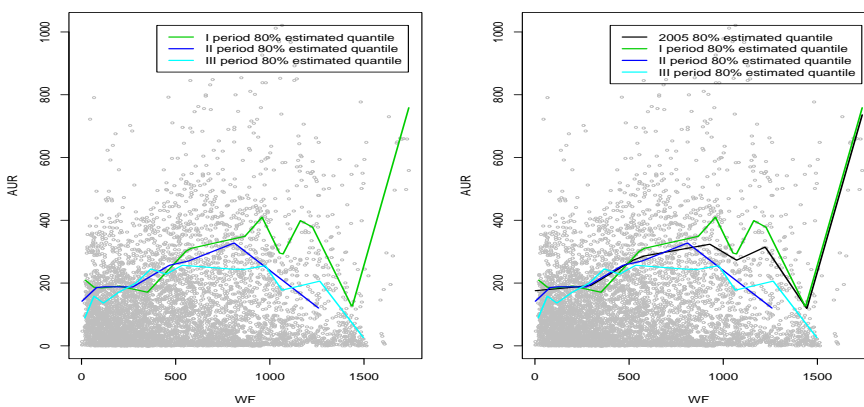
Figure 6.6: On the left: scatterplot of **AUR** plotted against the predictor **WF** referred to all the 2005 data, where we added the estimated 80% quantile for the I, II and III period. On the right: scatterplot of **AUR** plotted against the predictor **WF** referred to all the 2005 data, where we added the estimated 80% quantile for the I, II and III period and for all the 2005 data. In each plot $\lambda = 650$ for the 2005 data estimation, $\lambda = \frac{650}{3}$ for each period estimation.
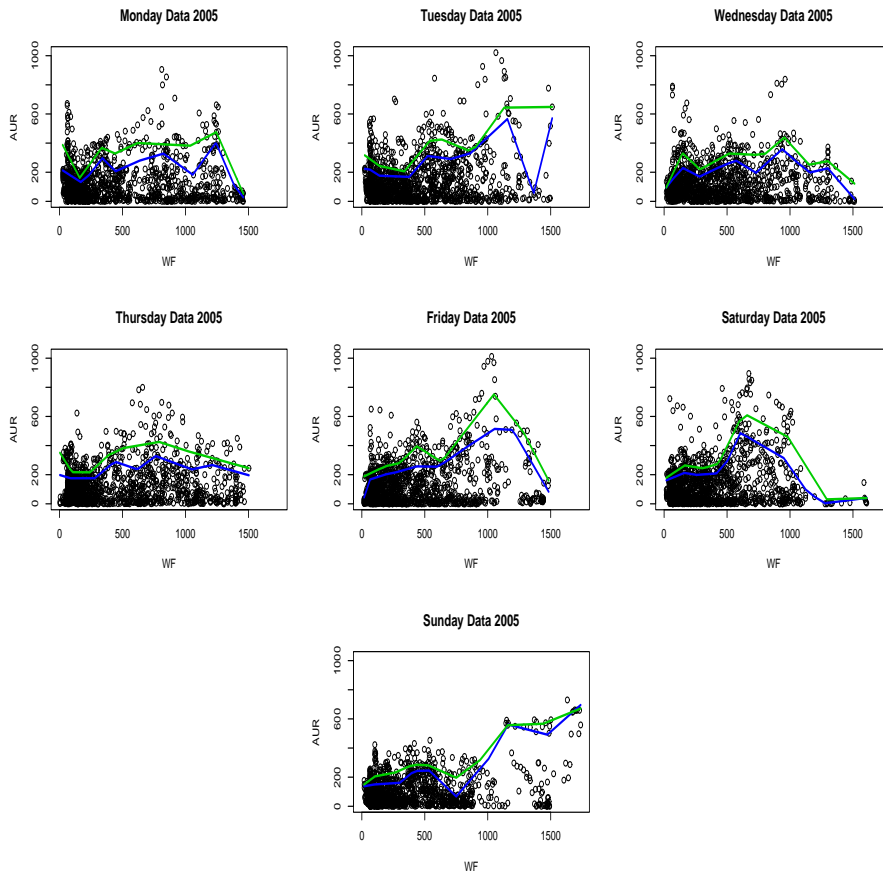
Figure 6.7: Scatterplot of **AUR** plotted against the predictor **WF** referred to each weekday data, where we added the estimated 80% quantile and the estimated 90% quantile.
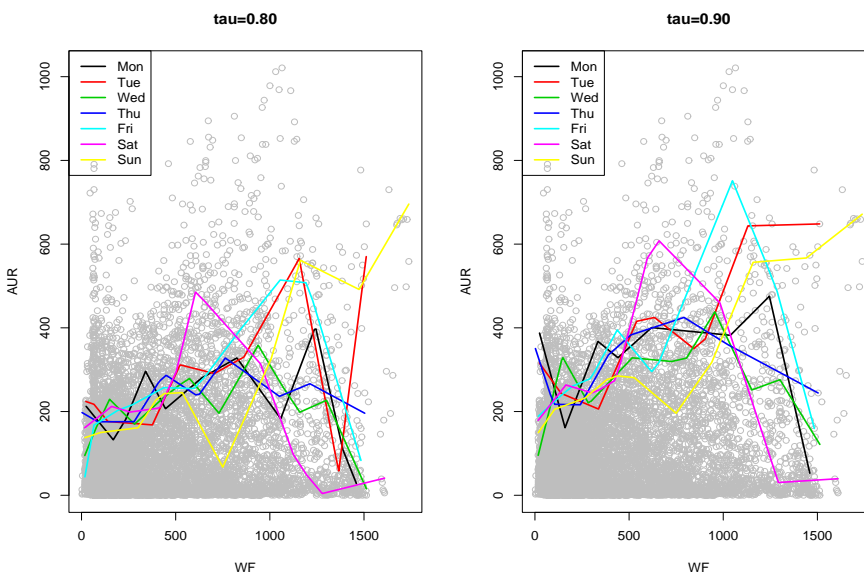
Figure 6.8: On the left: scatterplot of **AUR** plotted against the predictor **WF** referred to all the 2005 data, where we added the estimated 80% quantile functions referring to each day of the week. On the right: scatterplot of **AUR** plotted against the predictor **WF** referred to all the 2005 data, where we added the estimated 90% quantile functions referring to each day of the week.

A Full Model: Wind Forecasts, Time Index and Consumption

7.1 Introduction

We analyze the data modeled on a nonparametric regression model with two explanatory variable: the Wind Forecasts and the Time Index. The response variable is, again, the Actual Upward Regulating power. Then we analyze the data modeled on a nonparametric regression model with three explanatory variable: the Wind Forecasts, the Time Index and the consumptions, dividing the data in the three periods, as previously described.

7.2 Estimation of the extreme quantile value

The main task of this thesis is to estimate a model that explains the dependency of the upward regulating power on the wind forecasts. In order to do that in a way as good as possible, we would like to have a model that can forecast the

upward regulating power needed with a good precision, let us say corresponding to **one "risk situation"** per month.

In other words, we want to estimate the following quantile:

$$((1 - \frac{12}{8760}) \times 100)\% \tag{7.1}$$

where 12 is the number of the hour, at most, in which the model is accepted to fail in a year and 8760 is the number of hour in all the year.

From (7.1) we have

$$\bar{\tau} = 0.9986 \tag{7.2}$$

We want to estimate the $\bar{\tau}$th quantile, $\hat{q}_{\bar{\tau}}$.

Seen Figure 7.1, we assume that the distribution of the response variable right tail is an exponential distribution with density:

$$f_{AUR}(x) = ae^{-bx} \tag{7.3}$$

where $a, b \in \mathbb{R}$, $a, b > 0$.

Thus, the corresponding probability distribution function (p.d.f.) is

$$F_{AUR}(x) = 1 - \frac{a}{b}e^{-bx}, x \geq \bar{x} \tag{7.4}$$

In order to determine the value $\bar{x}$, we apply the normalization condition

$$\int_{\bar{x}}^{+\infty} f_{AUR}(x)dx = 1 \tag{7.5}$$

Since (7.4), we have

$$
\begin{aligned}
F_{AUR}(+\infty) - F_{AUR}(\bar{x}) &= 1 - 1 + \frac{a}{b}e^{-b\bar{x}} \\
&= \frac{a}{b}e^{-b\bar{x}}
\end{aligned}
$$

$$(7.6)$$

thus

$$
\frac{a}{b}e^{-b\bar{x}} = 1 \tag{7.7}
$$

and

$$
\bar{x} = -\frac{1}{b}log\frac{b}{a} \tag{7.8}
$$

if $a > b$, otherwise $\bar{x} = 0$.

In order to determine the parameters a, b we consider the estimated quantile regression functions $\hat{Q}(\tau_1)$ and $\hat{Q}(\tau_2)$ where

$$
\begin{aligned}
\tau_1 &= 0.8 \\
\tau_2 &= 0.9
\end{aligned}
$$

In any point of the function we proceed as follows.

Recalling that

$$
q_\tau(x) = F_X^{-1}(\tau) = inf\{x : F_X(x) \geq \tau\}. \tag{7.9}
$$

we have that

$$\frac{a}{b}e^{-bq_{90}} = 0.1$$
$$\frac{a}{b}e^{-bq_{80}} = 0.2$$

$$(7.10)$$

Given $\hat{q}_{\tau_1}$ and $\hat{q}_{\tau_2}$, ensuing from the nonparamatric quantile regression, from (7.10) we have

$$\hat{a} = \frac{1}{10}\hat{b}e^{\hat{b}\hat{q}_{\tau_2}}$$
$$\hat{b} = \frac{log2}{\hat{q}_{\tau_2} - \hat{q}_{\tau_1}}$$

$$(7.11)$$

such that

$$F_{AUR}(x) = 1 - \frac{\hat{a}}{\hat{b}}e^{-\hat{b}x}, x \geq \bar{x}$$

$$(7.12)$$

Using the results regarding the nonparametric quantile regression, got by the R function *rqss*, we estimate the parameters a and b in a number of values sufficient to fit the estimated functions $Q(0.95)$ and $Q(\bar{\tau})$. In order to do that, we remember that, for any $\tau \in (0, 1)$,

$$F_{AUR}(q_\tau) = 1 - \frac{\hat{a}}{\hat{b}}e^{-\hat{b}q_\tau}$$
$$= \tau$$

$$(7.13)$$

Thus,

$$q_\tau = -\frac{1}{\hat{b}}log[\frac{\hat{b}}{\hat{a}}(1 - \tau)]$$

$$(7.14)$$

7.3   The Model: WF and D

We consider, now, the following model

$$AUR = f(WF) + g(D) + \epsilon \qquad (7.15)$$

The constant presents in the basic model has been absorbed into one of the function.

Using the function *rqss* in R, we want to estimate the quantile smoothing splines which minimize

$$R_{\tau,\lambda}(f) = \sum_{i=1}^{n} \rho_\tau\{aur_i - (f(wf_i) + g(D_i))\} + \lambda_f \int_0^1 |f''(wf)| dwf + \lambda_g \int_0^1 |g''(D)| dD \qquad (7.16)$$

The knots for the optimal splines coincide with the observed wf_i and D_i respectively (see [13]), considered only once.

In Figure 7.2 we can see the estimated quantile regression functions for $\tau = 0.80$ and $\tau = 0.90$, estimated by R using the function *rqss*. We introduced the second explanatory variable, the time index **D**.

As described above, using the 80% and the 90% estimated quantile functions, we simulated the 95% and the 99.86% estimated quantile functions. The second one, the 99.86% quantile function correspond to the $\bar{\tau}$ quantile function that we were looking for, i.e. the situation in which we allow one error per month in the model.

In Figure 7.3 we can see the conditional quantile function corresponding to the 95%. The 95% quantile function has been obtained by the Extreme Value Theory as explained above. We plotted, in the graph, the values of the Actual Upward Regulating power as well.

In Figure 7.4 we can see the conditional quantile function corresponding to the 99.86%, i.e. the quantile we were looking for. Again, the 99.86% quantile function has been obtained by the Extreme Value Theory as explained above. We plotted, in the graph, the values of the Actual Upward Regulating power as well.
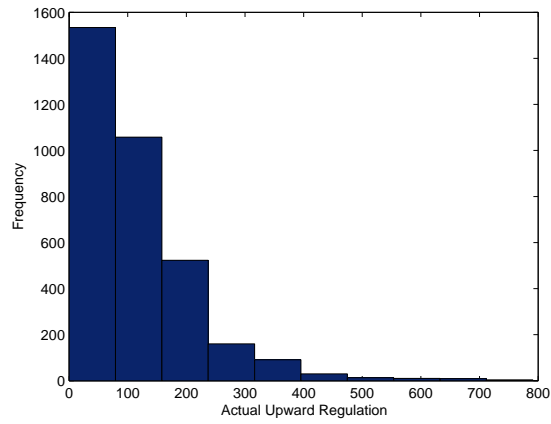
Figure 7.1: Histogram of the right tail of the response variable AUR
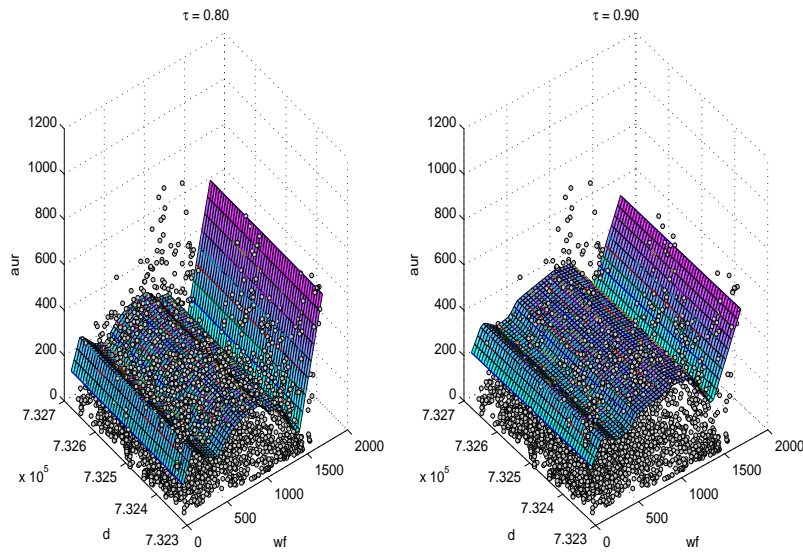

Figure 7.2: The figure illustrates the relationship between Actual Upward Regulation (AUR) and two explanatory variables, Wind Forecast (WF) and Time index (D), for the 80% and 90% quantile.
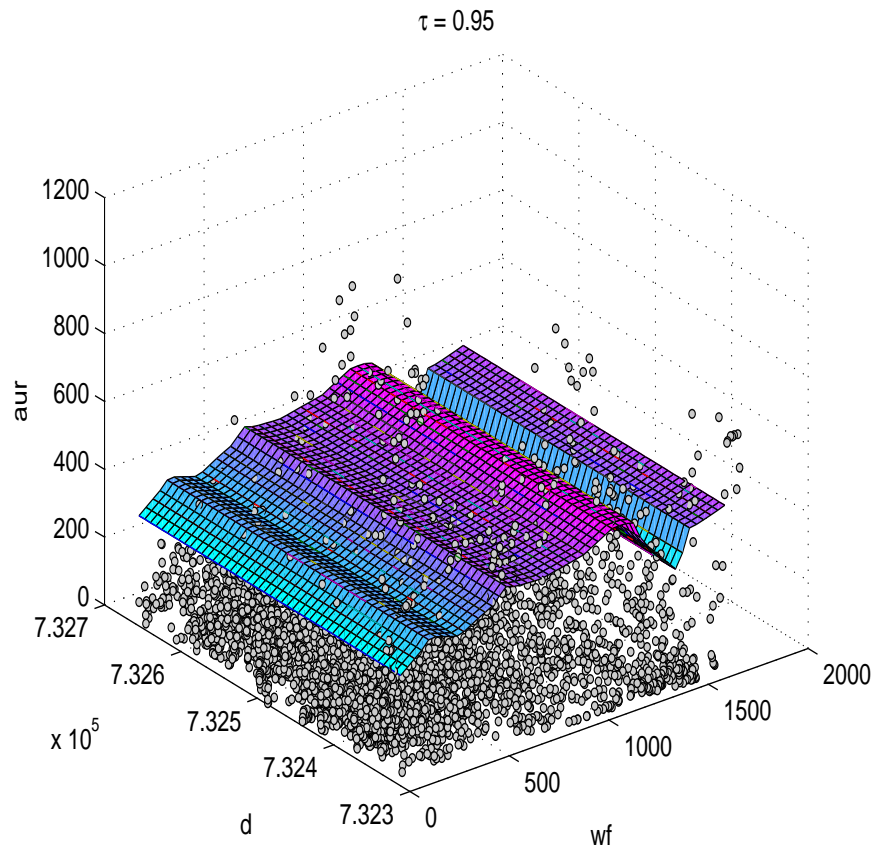
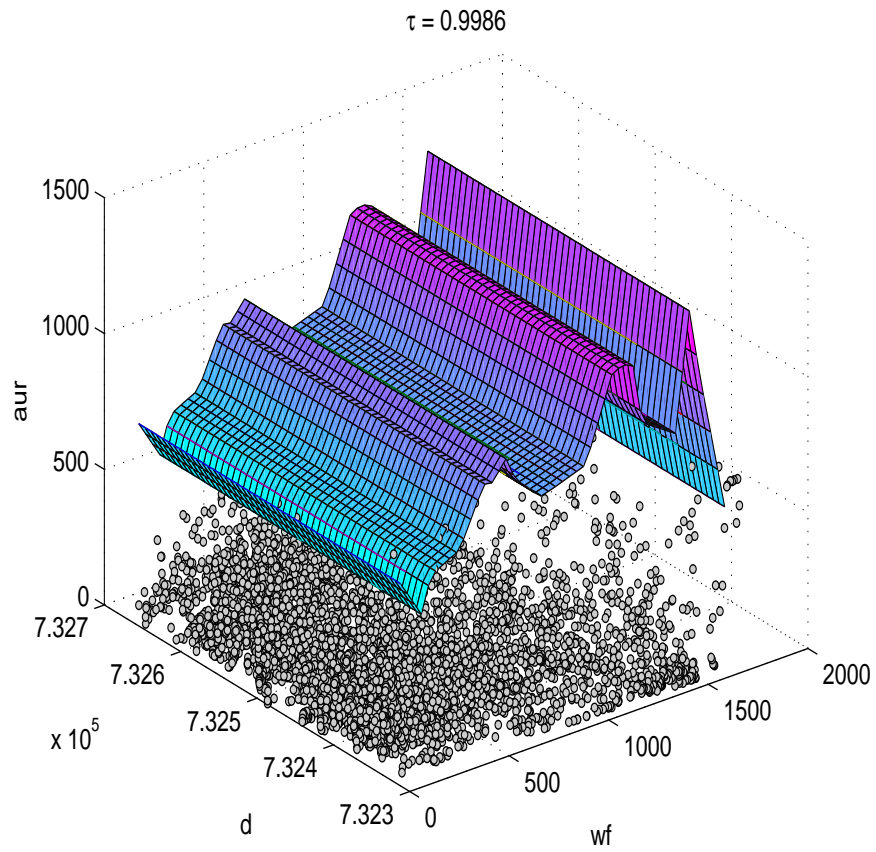Figure 7.3: The figures show the surface corresponding to the 95% conditional quantile function.

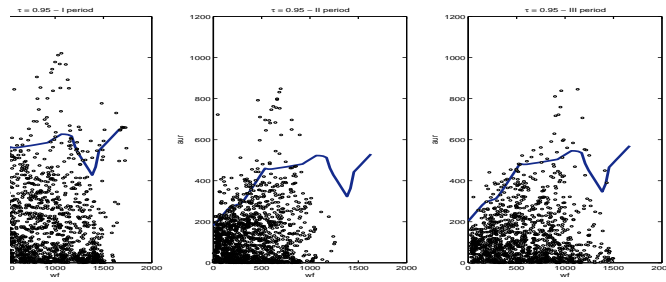Figure 7.4: The figures show the surface corresponding to the 99.86% conditional quantile function.

Figure 7.5: The figure shows the scatterplot of AUR versus WF, with D fixed for each of the three periods, where we added the conditional quantile function corresponding to the 95%.
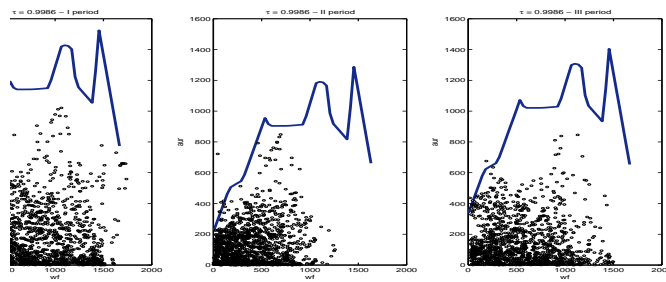

Figure 7.6: The figure shows the scatterplot of AUR versus WF, with D fixed for each of the three periods, where we added the conditional quantile function corresponding to the 99.86%.
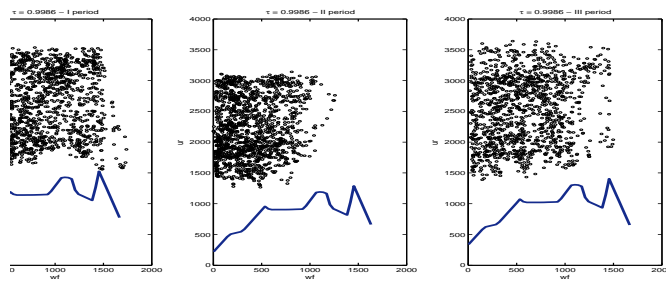

Figure 7.7: The figure shows the scatterplot of UR versus WF, with D fixed for each of the three periods, where we added the conditional quantile function corresponding to the 99.86%.
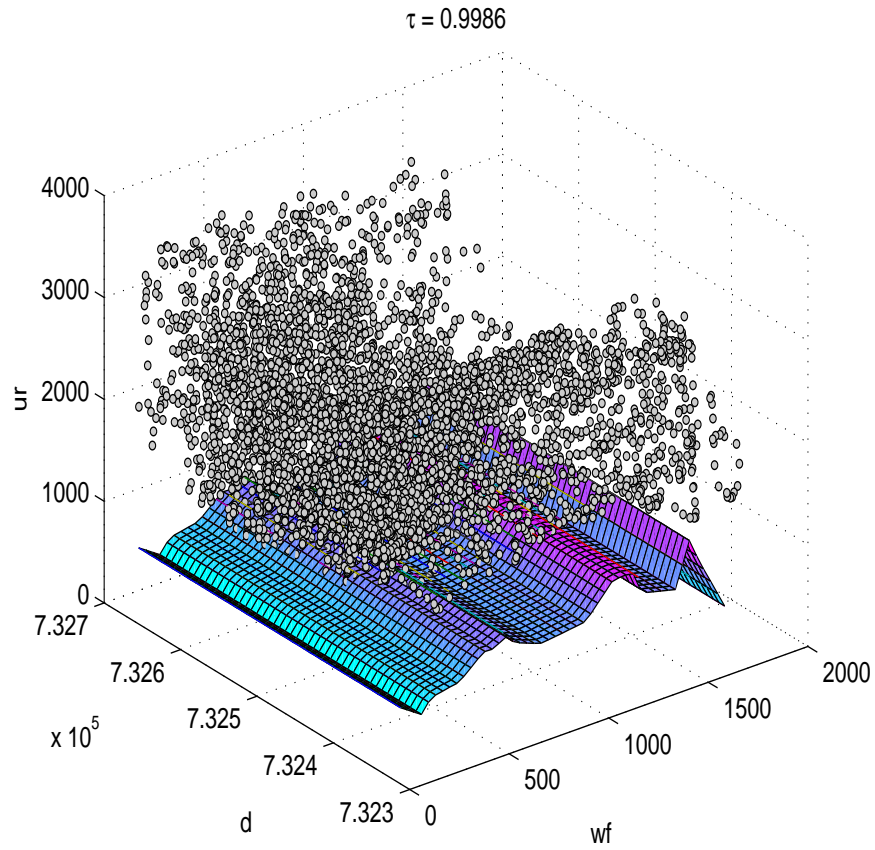
Figure 7.8: The figure shows the scatterplot of UR versus WF and D, where we added the conditional quantile function corresponding to the 99.86%.

In Figure 7.8 we added to the graph the values of the Upward Regulating power. As we can see, the suggestion of this work would allow Energinet.dk to better forecasting the need of regulating power, in order to avoid losses of important and expensive assets like the ancillary services and the regulating reserved.

7.4 The Model: WF, D and C

We consider, now, the following model

$$AUR = f(WF) + g(D) + h(C) + \epsilon \tag{7.17}$$

The constant presents in the basic model has been absorbed into one of the function.

Using the function *rqss* in R, we want to estimate the quantile smoothing splines which minimize

$$R_{\tau,\lambda}(f) = \sum_{i=1}^{n} \rho_\tau \{aur_i - (f(wf_i) + g(D_i))\} + \lambda_f \int_0^1 |f\prime\prime(wf)| dwf + \lambda_g \int_0^1 |g\prime\prime(D)| dD + \lambda_h \int_0^1 |h\prime\prime(C)| dC. \tag{7.18}$$

The knots for the optimal splines coincide with the observed wf_i, D_i and C_i respectively (see [13]), considered only once.

In Figures 7.9, 7.10, 7.11 we can see the estimated quantile regression functions for $\tau = 0.80$ and $\tau = 0.90$, estimated by R using the function *rqss* with three explanatory variables (WF, D, C), for each of the three periods, obtained by fixing the value of the explanatory variable D.

Again, using the 80% and the 90% estimated quantile functions, we simulated the 95% and the 99.86% estimated quantile functions. The second one, the 99.86% quantile function correspond to the $\bar{\tau}$ quantile function that we were looking for, i.e. the situation in which we allow one error per month in the model.

In Figures 7.12, 7.13, 7.14 we can see the conditional quantile function corresponding to the 95% for each period. The 95% quantile function has been
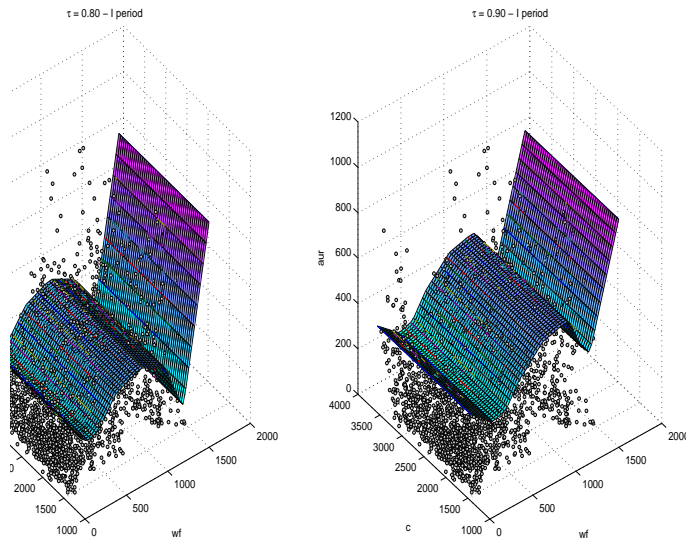
Figure 7.9: The figure illustrates the relationship between Actual Upward Regulation (AUR) and the explanatory variables, Wind Forecast (WF), Consumptions (C) and Time index (D) (fixed, value chosen in the I period), for the 80% and 90% quantile.
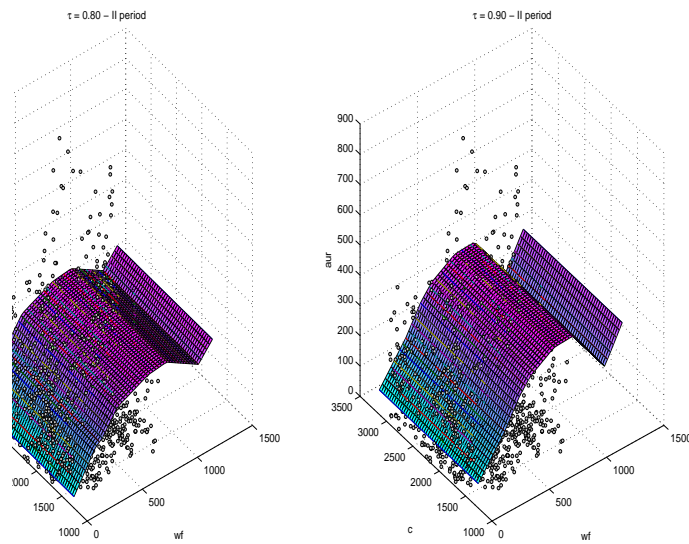
Figure 7.10: The figure illustrates the relationship between Actual Upward Regulation (AUR) and the explanatory variables, Wind Forecast (WF), Consumptions (C) and Time index (D) (fixed, value chosen in the II period), for the 80% and 90% quantile.
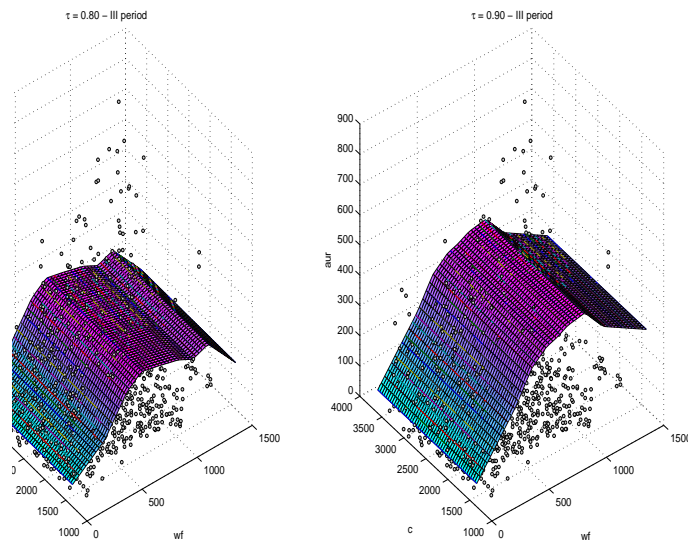
Figure 7.11: The figure illustrates the relationship between Actual Upward Regulation (AUR) and the explanatory variables, Wind Forecast (WF), Consumptions (C) and Time index (D) (fixed, value chosen in the III period), for the 80% and 90% quantile.
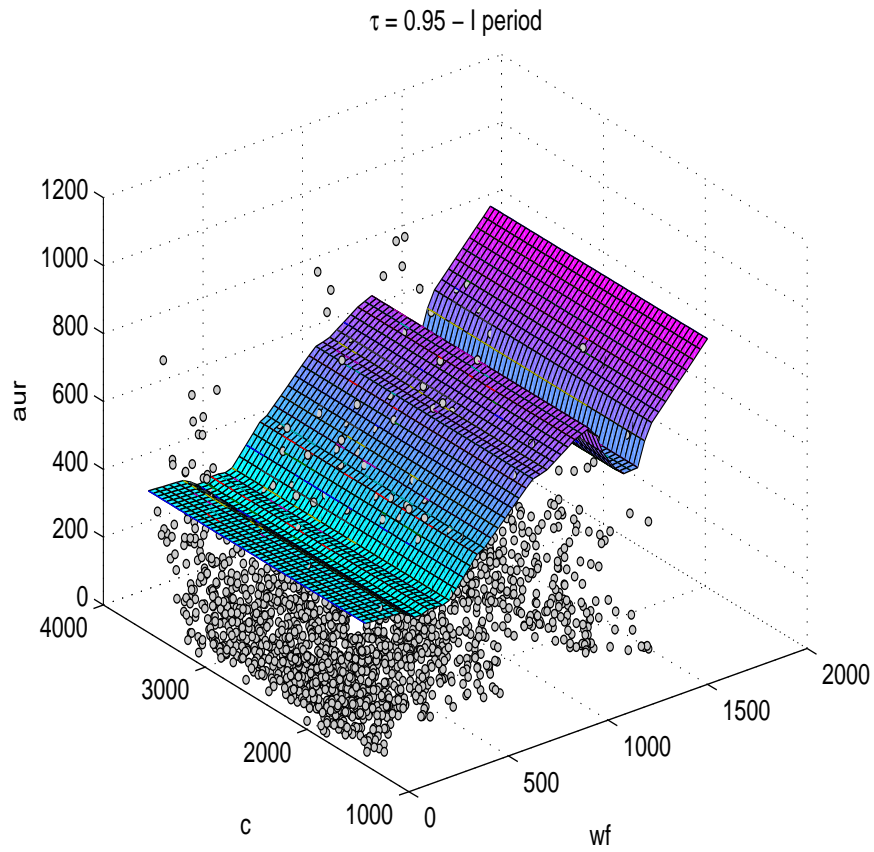
Figure 7.12: The figures show the surface corresponding to the 95% conditional quantile function for the I period.
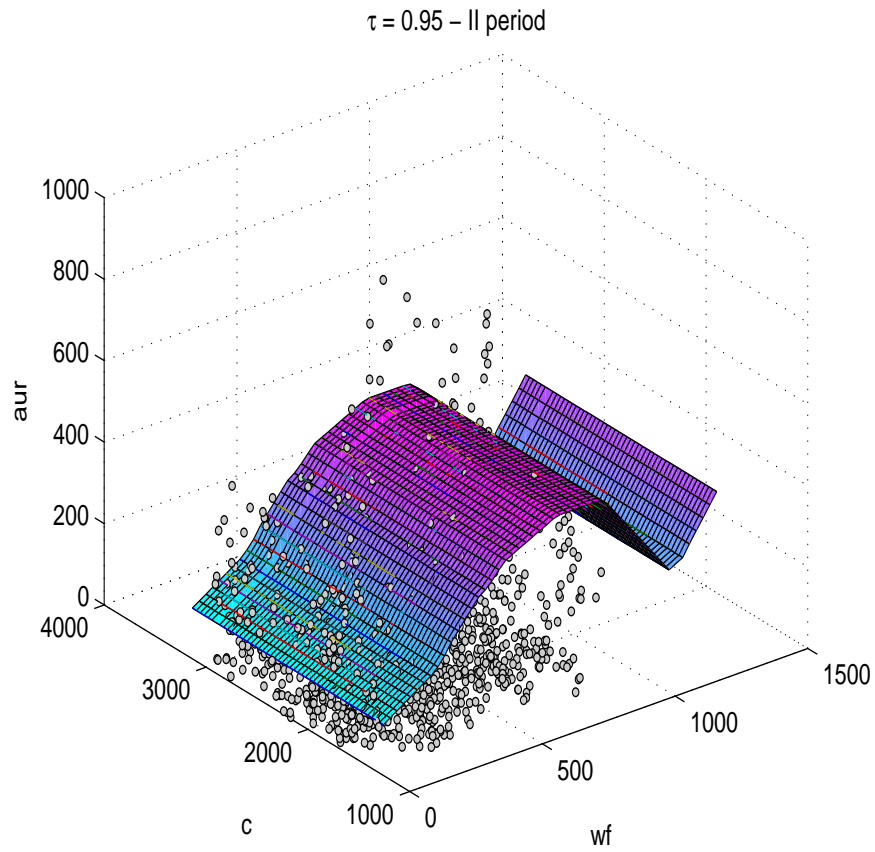
Figure 7.13: The figures show the surface corresponding to the 95% conditional quantile function for the II period.
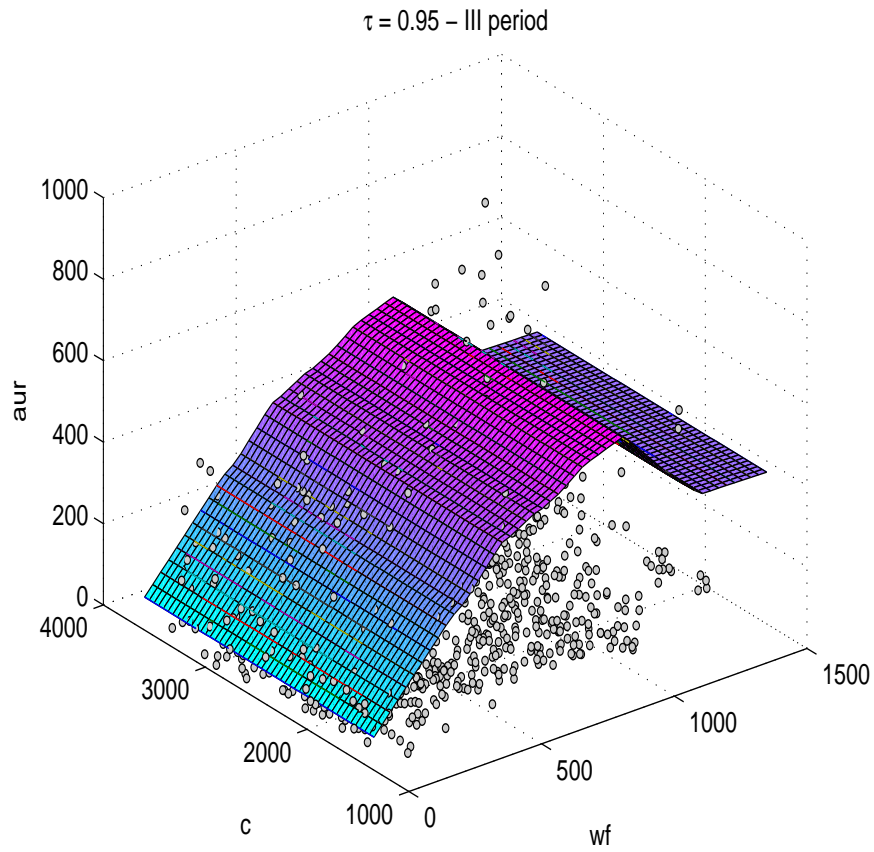
Figure 7.14: The figures show the surface corresponding to the 95% conditional quantile function for the III period.

obtained by the Extreme Value Theory as explained above. We plotted, in the graph, the values of the Actual Upward Regulating power as well.

In Figures 7.15, 7.16, 7.17 we can see the conditional quantile function corresponding to the 99.86%, i.e. the quantile we were looking for, for each period. Again, the 99.86% quantile function has been obtained by the Extreme Value Theory as explained above. We plotted, in the graph, the values of the Actual Upward Regulating power as well.

In Figures 7.18, 7.19, 7.20 we added to the graph the values of the Upward Regulating power. As we can see, again the suggestion of this work would allow Energinet.dk to better forecasting the need of regulating power, in order to avoid losses of important and expensive assets like the ancillary services and the regulating reserved.
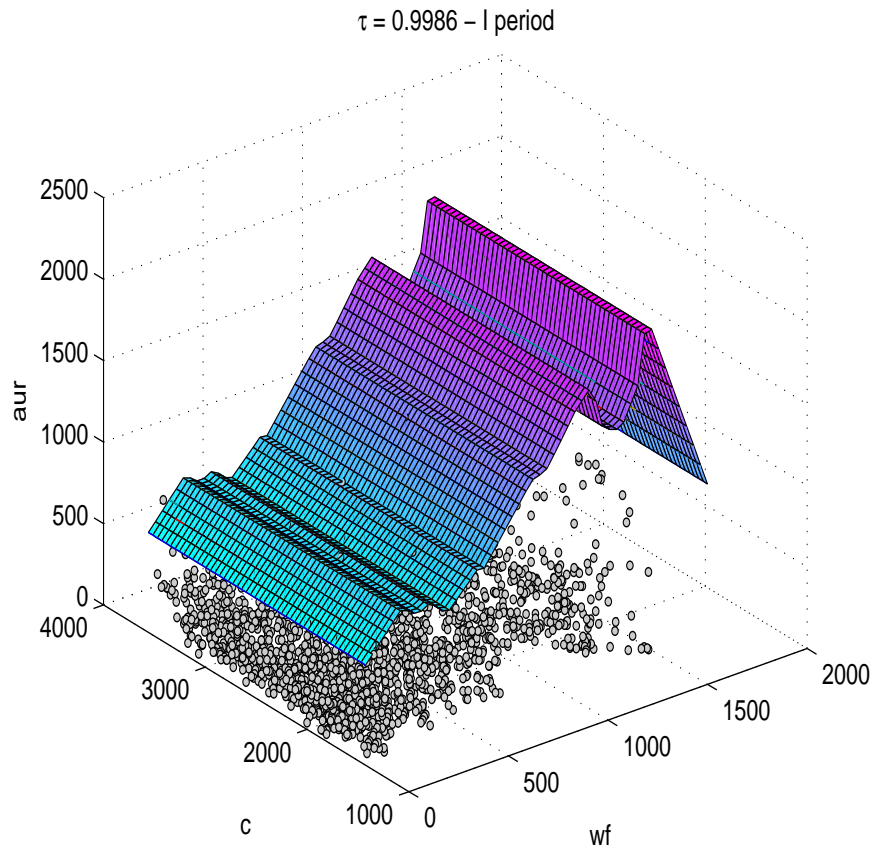
Figure 7.15: The figures show the surface corresponding to the 99.86% conditional quantile function for the I period.

Figure 7.16: The figures show the surface corresponding to the 99.86% conditional quantile function for the II period.

Figure 7.17: The figures show the surface corresponding to the 99.86% conditional quantile function for the III period.
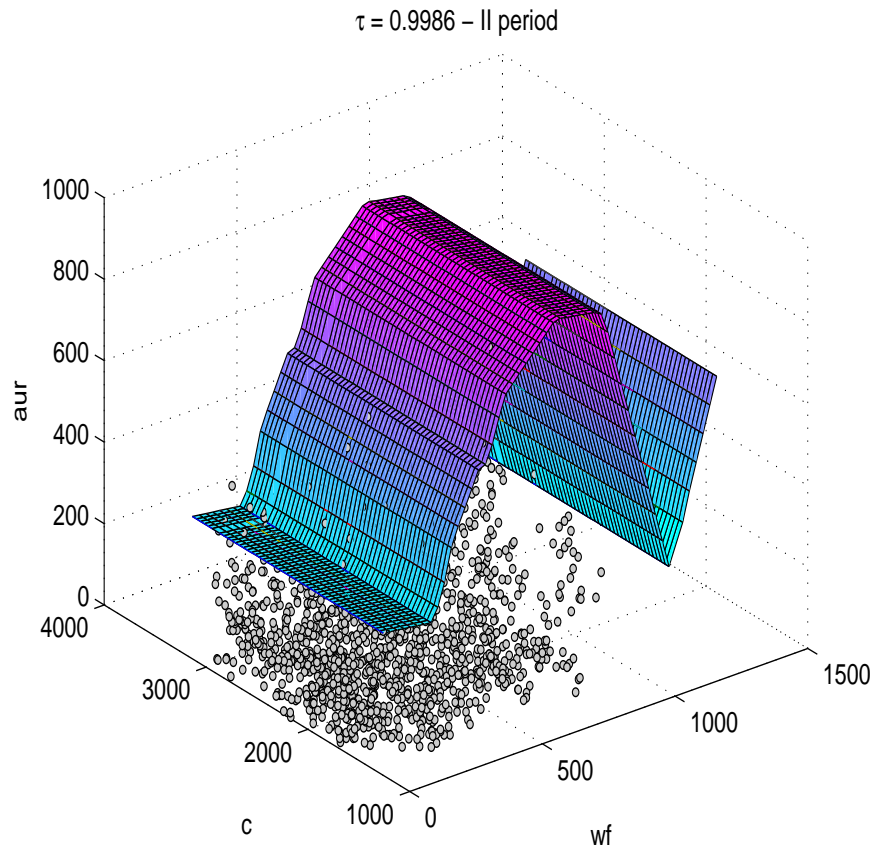
Figure 7.18: The figures show the surface corresponding to the 99.86% conditional quantile function for the I period.

Figure 7.19: The figures show the surface corresponding to the 99.86% conditional quantile function for the II period.
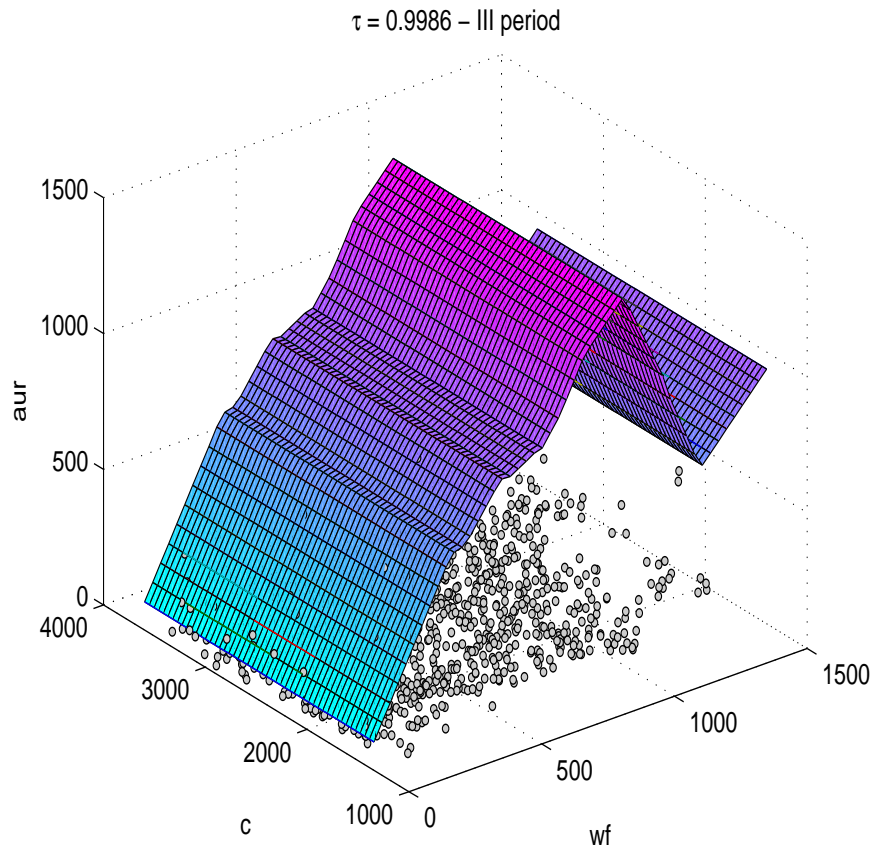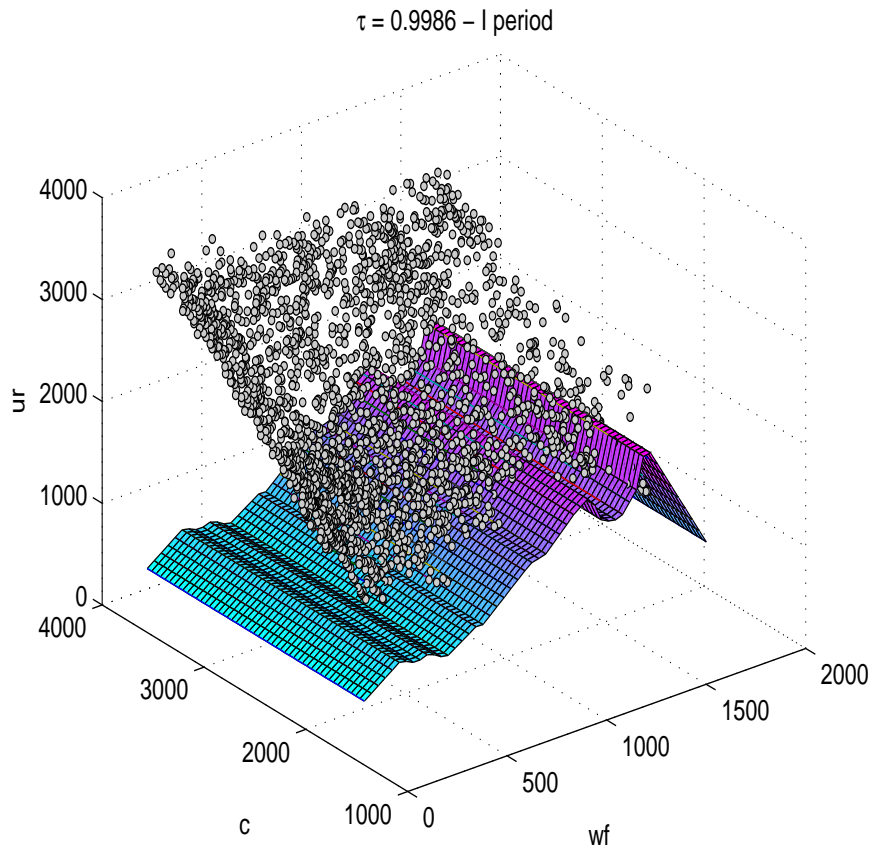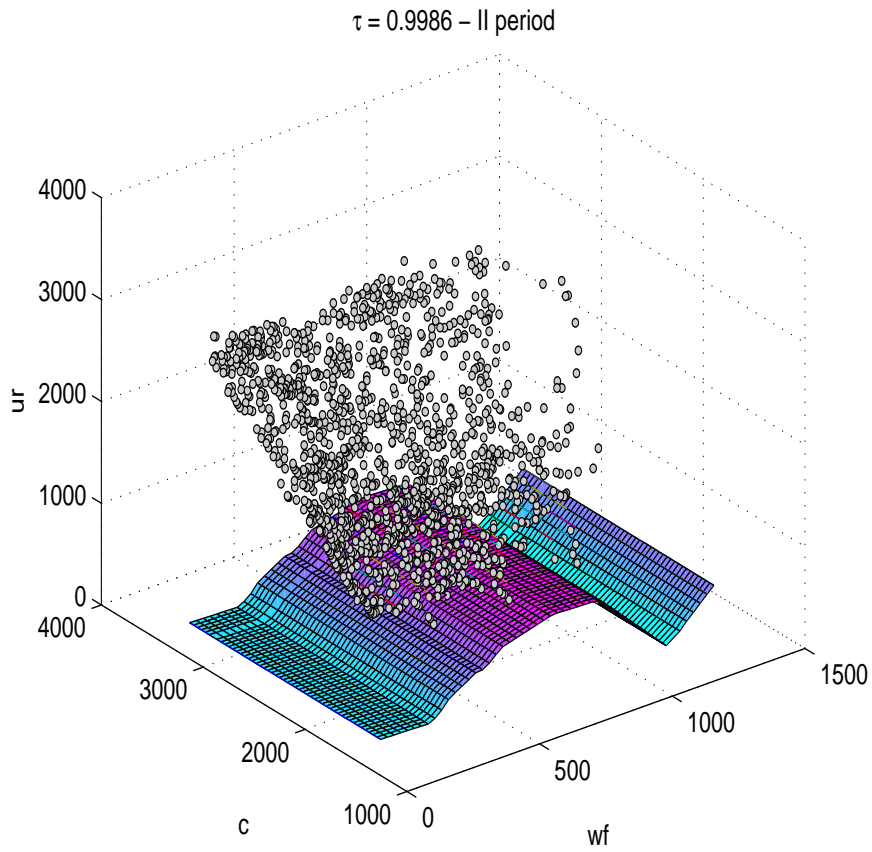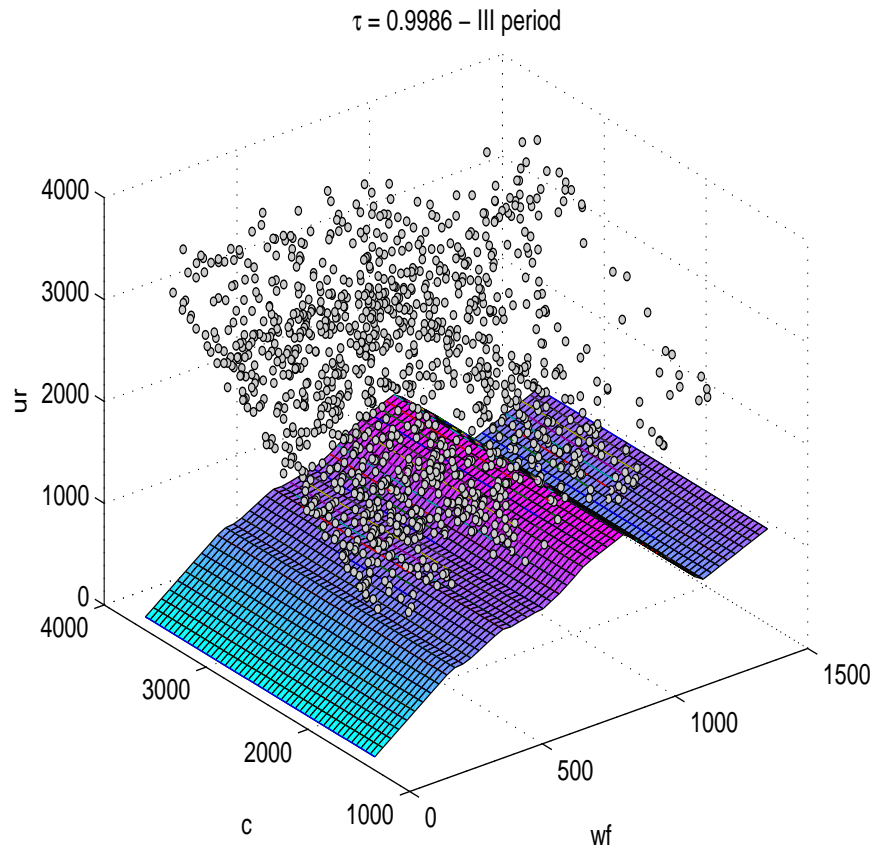
Figure 7.20: The figures show the surface corresponding to the 99.86% conditional quantile function for the III period.

CHAPTER 8

Conclusions

8.1 Results

8.1.1 Linear Quantile Regression

The Quantile Regression method is a very powerful method in order to model the dependency of Actual Upward Regulating power on the Wind Forecasts. Anyway, it has been seen the weakness of linear regression analysis in understanding the model and the need of using smoothing functions to accurately estimate that dependency.

8.1.2 Nonparametric Quantile Regression

Introducing Nonparametric Quantile Regression we have seen that the accuracy of the model has noticeably grown, even if the *rqss* function, function used in R to fit the model, is not really stable and often it needs re-setting of the parameters used, mostly re-setting of the smoothing parameter λ. This parameter, in the current version of the software has to be chosen by the user, and it is desirable that, in the next version, the parameter will be chosen by the R Software itself, in order to make the model more reliable. Otherwise, we suggest to use a

method like cross-validation to choose the smoothing parameter. In this Thesis it hasn't been possible because of the paucity of the number of data. The number of the data, referred just to the 2005 year, didn't allow us to use methods like cross-validation in choosing the smoothing parameter λ and, moreover, it has been the reason because of we couldn't admit less then one error per month in the model. In order to fulfill the demands of Energinet.dk and the Electricity Market in general, we should have admitted at most one/two errors per year, but, to do that, it is mandatory having more data, i.e. data referring to a larger number of years.

8.1.3 The Model

As seen in Chapters 6 and 7, the model that better fit the data is a Generalized Additive Model with two explanatory variable: the Wind Forecasts and the Time Index.

It has been seen that the Time Index must be carefully considered in fitting the model, both the case of the model with one explanatory variable and, obviously, the case of the model with two explanatory variables. Especially if we consider recurring seasonal features (as seen for the data split in the three periods).

8.2 Conclusion

It has been seen that the Actual Upward Regulating power depends on the Winds Forecasts. Moreover it has seen that it has a seasonal component. Applying the Regression Quantile theory in general, and specifically the Nonparametric Extreme Regression Quantile, allowed us to better understand the behavior of the model in the tail and in extreme cases.

Moreover, it has been possible to model the 99.86% quantile for the Actual Upward Regulating power, and, with a larger number of data, it will be possible to predict the behavior of the response variable with a good margin of safety, also for higher quantiles.

8.3   Further Study

Further studies may include more stochastic approaches in finding the smoothing parameter λ, and, with a larger number of data, they may help to fit models for higher quantiles.

Bibliography

[1] The comprehensive r archive network. http://cran.r-project.org/.

[2] The comprehensive tex archive network. http://www.ctan.org/.

[3] Energinet.dk website. http://www.energinet.dk/en/menu/Frontpage.htm.

[4] Nordpool website. http://www.nordpool.com/.

[5] Roger koenker website. http://www.econ.uiuc.edu/ roger/.

[6] Victor chernozhukov website. http://web.mit.edu/vchern/www/.

[7] Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/.

[8] A.J.McNeil. Estimating the tails of loss severity distributions using extreme value theory. Eth zentrum,zurich, Departement Mathematik, December 1996. available at http://www.sfu.ca/ rjones/econ811/readings/McNeil

[9] A.Quarteroni, R.Sacco, and F.Saleri. *Numerical Mathematics*. Springer-Verlag New-York. 2000.

[10] Chung-MingKuan. An introduction to quantile regression. Academia sinica, Institute of Economics, August 2004. available at http://www.sinica.edu.tw/econ/activities/conpdf/chung-ming-kuan.pdf.

[11] E.Castillo. *Extreme Value theory in Engineering*. Academic Press INC., London. 1991.

[12] G.G.Roussas. *A Course in Mathematical Statistics*. Academic Press INC., San Diego. 2000.

[13] H.A.Nielsen, H.Madsen, and T.S.Nielsen. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wiley Interscience (available at www.interscience.wiley.com)*, 2005.

[14] K.Dowd. The extreme value approach to var, 1999. available at http://www.fenews.com/fen11/extreme.html.

[15] M.Yuan. Gacv for quantile smoothing splines. *Computational Statistics and Data Analysis*, 2003. 50(3), 813-829.

[16] R.Doherty. A new approach to quantify reserve demand in system with significant installed wind capacity. *Transactions on Power System*, 2005. 20(2), 587- 595.

[17] R.Koenker. *Quantile Regression*. Wiley. 1990.

[18] R.Koenker. *quantreg: Quantile Regression*, 2006. R package version 3.90.

[19] R.Koenker and G.Bassett. Regression quantiles. *Econometria*, 1978. 46, 33- 50.

[20] R.Koenker and K.F.Hallock. Quantile regression: an introduction, 2001. available at www.econ.uiuc.edu/ roger/research/intro/intro.html.

[21] R.Koenker, Pin Ng, and S.Portnoy. Quantile smoothing splines. *Biometrika*, 1994. 60, pages 537-550.

[22] S.Lee and J.L.Horowitz. Nonparametric estimation of an additive quantile regression model. Econometric Society 2004 Far Eastern Meetings 721, Econometric Society, August 2004. available at http://ideas.repec.org/p/ecm/feam04/721.html.

[23] S.Portnoy. Local asymptotics for quantile smoothing splines. *The Annals of Statistics*, 1997. 25(1), 414-434.

[24] T.J.Hastie and R.J.Tibshiranu. Generalized additive models. *Chapman and Hall*, 1990. 1(3), 297-310.

[25] V.Chernozhukov. Nonparametric extreme regression quantile. Stanford, Ca, Department of Economics, December 1998. 33(2), 806-839.

[26] W.Hardle, M.Muller, S.Sperlich, and A.Werwatz. *Non parametric and Semiparametric Models*. Springer. 2004.

[27] X.He and S.Portnoy. Some asymptotic results on bivariate quantile splines. *Journal of Statistical Planning and Inference*, 2000. 91, 341-349.