MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Analysis and Stochastic Programming in Economics of Wind Energy"

verfasst von / submitted by

Michael Sedlmayer, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von / Supervisor:	Univ.-Prof. Dr. Radu Ioan Bot

Acknowledgements

First of all I want to express my gratitude to my supervisor Prof. Radu Bot who made this thesis possible and was confident about the successful outcome from the very first moment. I want to express my thanks for the faith he put in me in advance and which I hope I was able to confirm during the last months.
Next I want to thank Axel Böhm, who I learned a lot from. I could show up any time and he still was willing to give valuable advice. Our discussions enriched my views, whether on mathematical topics or on everyday life.
A special thanks goes to Günther Hörmann and the administrative staff, who were immensely helpful and friendly every time I came with a request that seemed more or less unsolvable at first.
I would also like to mention my colleagues who made the shared time a precious one. Some of them became close friends, others faded away over the years, but each of them contributed to the fact that I look back fondly.
Of course I have to thank my family and friends, who supported me incessantly and took load off me during stressful times. Together we go through difficult times and together we celebrate accomplishments. Most importantly, they enable me to be myself.
Last but not least I want to express my gratitude to Evelyn Stepancik and Stefan Enzelberger, who convinced me back then that all my (juvenile) life my deepest desire was to explore the fantastic world of mathematics. Without them, I would have never written this thesis in the first place.

Contents

1 Abstract

When producing electricity by wind power stations for a daily spot market, one has to submit a production schedule every day for the following day. Due to different prices for scheduled energy and energy deviating in either way from this announced amount, it may be beneficial to modify the production schedule differing from the initial wind power forecast. This gives rise to "strategic adjustment" of the production schedule depending on the price situation. The prices are published two to three months later, so one has to make predictions of the price situation to develop effective strategic adjustment.

In this thesis we analyse the policies of a company which were used in the past and certain other aspects as limitation of production or adjustment on holidays. We establish a framework for evaluation and compare different policies relative to a "best case". Furthermore we treat two concepts for the forecast of prices - "historic averages" and linear regression models. Subsequently we cover two approaches of strategic adjustment; the first one inspired by the company's heuristic, the second one by solving a stochastic program. For the latter we use a "Sample Average Approximation" approach.

2 Introduction

This Master's Thesis came into being due to a cooperation of the University of Vienna, Faculty of Mathematics with an Austrian energy provider. This company uses wind power stations to produce electrical energy to participate in a daily spot-market. Due to regulations of the market, every participant has to submit a production schedule for the following day. This is done to coordinate the supply of and demand for electricity to avoid shortage or surplus which would lead to black-outs in (at least) major parts of the transmission system and therefore cause personal and property damages. Production deviating from this schedule will be penalised, by lower prices for exceeding energy and higher penalties for deficit energy. These "balancing energy" prices as well as the "day-ahead" price for the scheduled energy are subject to certain fluctuation during a day and have different relative position to each other. For example it can happen that the day-ahead price and the exceeding price are pretty close while the penalty for deficit energy is exorbitantly high, then in practice there is little difference in the revenue whether production exceeds the schedule or not. However deficit energy leads to severe loss in the revenue in this situation. This yields that it may be beneficial to modify the figures of forecast production before submitting them, depending on the price situation. In the above example we would rather decrease the schedule compared to the estimated production forecast to avoid underproduction, as the opportunity cost for exceeding energy is small but the negative impact of deficit energy is immense. This concept will be called "Strategic Adjustment".
So if we knew the prices beforehand, we could adjust the schedule in a way to receive the best possible outcome. However we do not know the prices in the moment we would need them, in particular they are published two to three months later. Therefore one has to predict the incorporated uncertain entities to be able to develop effective strategic adjustment. Another task is to decide *how* to process the forecast prices to receive "optimal" adjustment.

In the following we will start by explaining two of the mathematical constructs we will be using later on. The first one is "Linear Regression", where we start with simple linear regression and some of its useful statistical properties. We continue with generalising to multiple linear regression and its respective analogies. Moreover, we introduce a measure to judge the quality of linear models - the coefficient of determination R^2. After that we proceed with "Sample Average Approximation", which is part of the domain of stochastic programming. This approach makes it possible to easily include uncertainty in an optimisation problem, for instance when numerous historical samples are at hand.

Subsequently we will describe the specific setting we are working in. There are several special features that simplify the problem, e.g., the prices are always positive, leading to maximal production free from voluntary reduction. We will proceed with a descriptive/analytical part, where we cover various questions that the company proposed, e.g., investigating specific adjustment policies on public holidays. Furthermore we will establish a framework for evaluation and introduce a rating to compare different policies. For this we take a reasonable best case as objective reference.
After that we will develop own adjustment policies, where we will start with the simplest approaches and then continue with more complex ones. For this we need to develop certain ways to forecast the prices, as they determine how we should adjust the submitted schedule. In a first attempt we use "historic averages", where we take the average of historical data from reasonable periods of time. Secondly, we employ univariate linear models to predict each of the prices.
Then we continue with establishing two different approaches of strategic adjustment, both using the several types of forecasts developed before. The first one is to derive so-called "adjustment coefficients" by a heuristic, to adjust the schedule relative to the initial production forecast. The second one is to solve a stochastic optimisation problem, to obtain already absolute figures for the schedule.

We conclude this thesis by a short discussion of the obtained results and address several topics that may be subject of future works to develop further improvements.

3 Preliminaries

Some of the ideas and concepts we will use in the following work arise straight forward. Others, on the other hand, are well established mathematics and to some extent part of the specific domain knowledge, but might need some explanation beforehand. This will be done in the following section.

3.1 Linear Regression

The first non-trivial topic we encounter is “Linear Regression”. Describing (and predicting) data mathematically is one of today’s biggest issues - employing linear models is one of the easiest and (therefore) most popular means to do so.
In the following we will start by introducing the one-dimensional case of “Simple Linear Regression” to motivate and arrive at “Multiple Linear Regression”. We will formulate certain assumptions we suppose that we are able to work properly and receive useful mathematical properties.
Since linear regression is such a well established topic there are many references such as for example [2] or [3], but we will base our essay mainly on [1], in particular chapters 3 and 7.

3.1.1 Simple Linear Regression

In the following we will go by [1], Chapter 3.

Introduction

In the simplest form we study the linear correspondence between two one-dimensional variables $X_i, Y_i \in \mathbb{R}$, namely

$$Y_i = \beta_0 + \beta_1 X_i + U_i, \qquad i = 1, \dots, n. \tag{3.1}$$

Here Y_i stands for the i-th observation of the dependent variable Y, the regressand, while X_i denotes the i-th observation of the independent variable X, the regressor. The number of observations is denoted by n. The two real numbers β_0 and β_1 are called the intercept and the slope of the linear model (3.1), respectively. We assume that β_0 and β_1 are unknown parameters that we want to estimate. Furthermore Y_i is not fully determined by the linear expression $(\beta_0 + \beta_1 X_i)$, it is subject to a random error U_i. We expect this disturbance to fulfil certain conditions, that we will state later.

In applications β_0 and β_1 are unknown and need to be estimated from the observations $\{(X_i, Y_i)\,|i = 1, \dots, n\}$. This means that the true values β_0 and β_1, as well as the true error U_i cannot be determined exactly. However, the aim of linear regression is to find the "best" slope $(\widehat{\beta_0} + \widehat{\beta_1}X)$ explaining the observations (X_i, Y_i). To judge which line is the best one, we need to introduce a measurement of the misfit from the observed Y_i to the estimation $\widehat{Y_i} = \widehat{\beta_0} + \widehat{\beta_1}X_i$.
Here and in the rest of this work, a circumflex denotes an estimator for the original entity, e.g., $\widehat{a}$ is an estimator for a.
So each observation (X_i, Y_i) leads to an observable error

$$\varepsilon_i := Y_i - \widehat{Y_i} = Y_i - (\widehat{\beta_0} + \widehat{\beta_1}X_i).$$

One possibility to derive one single number to judge all errors at once, is to compute the sum of squares of residuals

$$SSR(\beta_0, \beta_1) := \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \widehat{\beta_0} - \widehat{\beta_1}X_i\right)^2. \tag{3.2}$$

Instead of taking the sum of squares of the errors, one could also take the sum of absolute values of the errors. This would lead to other interesting results, which is not subject of this work, though. Now that we chose the measure for the misfit, we can find the "best" slope by minimising the residual sum of squares, which leads to least squares estimation.

Least Squares Estimation

The minimum of the sum of squares of residuals $SSR(\beta_0, \beta_1) = \sum\limits_{i=1}^{n} \varepsilon_i^2$ is characterised by the following so-called "normal equations":

$$\frac{\partial\left(\sum\limits_{i=1}^{n}\varepsilon_i^2\right)}{\partial\widehat{\beta_0}} = -2\sum_{i=1}^{n}\varepsilon_i \overset{!}{=} 0$$

$$\frac{\partial\left(\sum\limits_{i=1}^{n}\varepsilon_i^2\right)}{\partial\widehat{\beta_1}} = -2\sum_{i=1}^{n}\varepsilon_i X_i \overset{!}{=} 0$$

or equivalently

$$\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} Y_i - n\widehat{\beta}_0 - \widehat{\beta}_1 \sum_{i=1}^{n} X_i \stackrel{!}{=} 0 \tag{3.3}$$

$$\sum_{i=1}^{n} \varepsilon_i X_i = \sum_{i=1}^{n} X_i Y_i - \widehat{\beta}_0 \sum_{i=1}^{n} X_i - \widehat{\beta}_1 \sum_{i=1}^{n} X_i^2 \stackrel{!}{=} 0 \tag{3.4}$$

Solving the normal equations (3.3) and (3.4) gives the "ordinary least squares" (OLS) estimators, which we denote by $\widehat{\beta_0^*}$ and $\widehat{\beta_1^*}$. We derive

$$\widehat{\beta_0^*} = \overline{Y} - \widehat{\beta_1^*}\overline{X} \tag{3.5}$$

and

$$\widehat{\beta_1^*} = \frac{\sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2} = \frac{\sum\limits_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum\limits_{i=1}^{n} X_i^2 - n\overline{X}^2} \tag{3.6}$$

where

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \overline{Y} := \frac{1}{n}\sum_{i=1}^{n} Y_i$$

and we assume that all operations here are justified, especially $\sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \neq 0$.

Classical Assumptions

In order to derive some interesting properties of the OLS estimators of β_0 and β_1 we have to impose some further conditions on the linear model (3.1).

Assumption 1. *The random noise has zero mean, i.e.,* $\mathbb{E}[U_i] = 0$ *for all* $i = 1, ..., n$.

This assumption is to make sure, that on the average our estimation is correct. Otherwise we would systematically overestimate or underestimate the true values.

Assumption 2. *The random noise has constant variance, i.e.,* $Var[U_i] = \sigma^2$ *for all* $i = 1, ..., n$.

Assumption 2 ensures that each observation is equally reliable. This property is called homoscedasticity.

Assumption 3. *The random noise is not correlated between different observations, i.e.,* $\mathbb{E}[U_i U_j] = 0$ *for* $i \neq j$, $i, j = 1, ..., n$.

This assumption means, that the noise from any observation does not affect the disturbance of any other one.

Assumption 4. *The regressor variable X is deterministic, i.e., X is not a random variable and fixed in repeated samples. Furthermore,* $\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \neq 0$ *and has a finite limit as $n \to \infty$.*

Assumption 4 implies, that the independent variable X is not correlated with the noise U. Moreover the above assumption ensures that we have at least two distinct values for the explanatory variable X.

In what way these four assumptions hold needs to be investigated in reality. For example Assumption 3 regarding the noise to be uncorrelated, is less likely to be satisfied when the underlying data is a time-series. Consecutive points in time may affect each other and thus could have correlated noise over time.
However we will see whether real properties deviating from the theoretical assumptions leads to considerably worse results.

Properties of Least Squares

Next we see which useful properties we can deduce from the above Assumptions 1-4.

First of all, we want to know whether our estimators are correct on average or if they are biased in a certain way. This motivates the following definition.

Definition 3.1. An estimator $\widehat{\alpha}$ for α is said to be unbiased for α, if $\mathbb{E}\left[\widehat{\alpha}\right] = \alpha$.

This leads us to the first useful property of OLS estimators, which we summarise in the following proposition.

Proposition 3.1. ***Unbiasedness***
The ordinary least squares estimators $\widehat{\beta_0^*}$ and $\widehat{\beta_1^*}$ are unbiased for β_0 and β_1, respectively.

Proof. We already know by (3.6), that

$$\begin{aligned}
\widehat{\beta_1^*} &= \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)Y_i-\overline{Y}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2} \\
&= \beta_1+\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)U_i}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2},
\end{aligned} \tag{3.7}$$

using that $\sum_{i=1}^{n}\left(X_i-\overline{X}\right)=0$ and substituting $Y_i=\beta_0+\beta_1X_i+U_i$ from (3.1). Taking the expected value on both sides of the equation, using Assumption 1 and Assumption 4 we conclude that $\widehat{\beta_1^*}$ indeed is unbiased for β_1.
Similarly one can derive that $\widehat{\beta_0^*}$ is unbiased for β_0. □

The next property concerns asymptotic behaviour and ensures that the estimator converges in an appropriate sense to the true value. We need to clarify what is meant exactly, which is done by the following definition.

Definition 3.2. The estimator $\hat{\alpha}_n$ of α is said to be consistent for α, if $\operatorname{plim}\hat{\alpha}_n=\alpha$, i.e., for all $\varepsilon>0$ we have $\lim_{n\to\infty}P\left(|\hat{\alpha}_n-\alpha|>\varepsilon\right)=0$.

Consistency means that the estimator converges in probability to the true value of the parameter, see [4] for more detailed explanations.

Proposition 3.2. ***Consistency***
The ordinary least squares estimators $\widehat{\beta_0^*}$ and $\widehat{\beta_1^*}$ are consistent for β_0 and β_1.

Proof. A sufficient condition for an estimator to be consistent is that it is asymptotically unbiased and that its variance vanishes asymptotically (see [4]). We already know that $\widehat{\beta_0^*}$ and $\widehat{\beta_1^*}$ are unbiased for β_0 and β_1 by Proposition 3.1. We show that $\widehat{\beta_1^*}$ is consistent for β_1:
First we need to derive the variance of $\widehat{\beta_1^*}$. Substituting (3.7) in the definition of

variance we deduce that

$$
\begin{aligned}
\operatorname{Var}\left(\widehat{\beta_1^*}\right) &= \mathbb{E}\left[\left(\widehat{\beta_1^*} - \mathbb{E}\left[\widehat{\beta_1^*}\right]\right)^2\right] \overset{Prop.\ 3.1}{=} \mathbb{E}\left[\left(\widehat{\beta_1^*} - \beta_1\right)^2\right] \\
&\overset{(3.7)}{=} \operatorname{Var}\left(\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right) U_i}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\right) \\
&\overset{A3}{=} \sum_{i=1}^{n} \operatorname{Var}\left(U_i\right) \frac{\left(X_i - \overline{X}\right)^2}{\left(\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2\right)^2} \\
&\overset{A2}{=} \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2},
\end{aligned}
$$

where we additionally used Assumption 3, i.e., that the U_i are uncorrelated with each other and Assumption 2 about constant variance. Employing Assumption 4, finally we deduce

$$
\lim_{n\to\infty} \operatorname{Var}\left(\widehat{\beta_1^*}\right) = \lim_{n\to\infty} \frac{\frac{\sigma^2}{n}}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = 0,
$$

as the numerator $\frac{\sigma^2}{n} \to 0$ and the denominator $\frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2 \neq 0$ and has a finite limit by assumption as $n \to \infty$. That concludes the proof of $\widehat{\beta_1^*}$ being consistent for β_1. Similarly one can show that $\widehat{\beta_0^*}$ is consistent for β_0. □

The last property we present here is a famous one.

Proposition 3.3. ***Best Linear Unbiased Estimator***
Both $\widehat{\beta_0^*}$ and $\widehat{\beta_1^*}$ are "Best Linear Unbiased Estimators" (BLUE) for β_0 and β_1, respectively.

The above is better known as Gauss-Markov theorem.

Theorem 3.1. ***Gauss-Markov***
We consider the simple linear model (3.1)*. Suppose that Assumptions 1-4 hold. A linear estimator of β_j, $j = 0, 1$, is a linear combination of the form $\widehat{\beta_j} = c_{1j}Y_1 + \ldots + c_{nj}Y_n$, in which the coefficients $c_{1j}, ..., c_{nj}$ are not allowed to depend on the true values β_j, but on the observable variables X_j.*
Then the ordinary least squares (OLS) estimators $\widehat{\beta_j^*}$ are best linear unbiased estimators

(BLUE), i.e., linear unbiased estimators with $\operatorname{Var}\left(\widehat{\beta_j^*}\right) \leq \operatorname{Var}\left(\widehat{\beta_j}\right)$ *for any other linear unbiased estimator* $\widehat{\beta_j}$.

Proof. We show that $\widehat{\beta_1^*}$ is BLUE, the proof for $\widehat{\beta_0^*}$ is similar. First we see that $\widehat{\beta_1^*}$ is linear, as

$$\widehat{\beta_1^*} = \sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)}{\sum\limits_{j=1}^{n} \left(X_j - \overline{X}\right)^2} Y_i = \sum_{i=1}^{n} v_i Y_i, \qquad \text{with } v_i := \frac{\left(X_i - \overline{X}\right)}{\sum\limits_{j=1}^{n} \left(X_j - \overline{X}\right)^2}.$$

Now let $\widehat{\beta_1}$ be any arbitrary linear unbiased estimator $\widehat{\beta_1} = \sum\limits_{i=1}^{n} w_i Y_i$ for β_1. Then we plug in the definition (3.1) of Y_i in $\widehat{\beta_1}$, which gives

$$\widehat{\beta_1} = \beta_0 \sum_{i=1}^{n} w_i + \beta_1 \sum_{i=1}^{n} w_i X_i + \sum_{i=1}^{n} w_i U_i.$$

For $\widehat{\beta_1}$ to be unbiased for β_1 we need

$$\mathbb{E}\left[\widehat{\beta_1}\right] = \beta_0 \sum_{i=1}^{n} w_i + \beta_1 \sum_{i=1}^{n} w_i X_i + \sum_{i=1}^{n} w_i \, \mathbb{E}\left[U_i\right] = \beta_0 \sum_{i=1}^{n} w_i + \beta_1 \sum_{i=1}^{n} w_i X_i \stackrel{!}{=} \beta_1$$

That implies the following,

$$\sum_{i=1}^{n} w_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} w_i X_i = 1,$$

assuming that we have a non-zero intercept β_0. Thus

$$\widehat{\beta_1} = \beta_1 + \sum_{i=1}^{n} w_i U_i.$$

Analogously we get $\sum\limits_{i=1}^{n} v_i = 0$ and $\sum\limits_{i=1}^{n} v_i X_i = 1$. Let $d_i := w_i - v_i$, then $\sum\limits_{i=1}^{n} d_i = 0$ and $\sum\limits_{i=1}^{n} d_i X_i = 0$. Hence we have by definition of v_i that

$$\sum_{i=1}^{n} v_i d_i = \frac{\sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right) d_i}{\sum\limits_{j=1}^{n} \left(X_j - \overline{X}\right)^2} = \frac{\sum\limits_{i=1}^{n} X_i d_i - \overline{X} \sum\limits_{i=1}^{n} d_i}{\sum\limits_{j=1}^{n} \left(X_j - \overline{X}\right)^2} = 0.$$

Thus we obtain

$$\begin{aligned}\operatorname{Var}\left(\widehat{\beta_1}\right) &= \operatorname{Var}\left(\sum_{i=1}^n w_i U_i\right) = \sigma^2 \sum_{i=1}^n w_i^2 = \sigma^2 \left(\sum_{i=1}^n v_i^2 + 2\sum_{i=1}^n v_i d_i + \sum_{i=1}^n d_i^2\right) \\ &= \operatorname{Var}\left(\widehat{\beta_1^*}\right) + \sigma^2 \sum_{i=1}^n d_i^2 \geq \operatorname{Var}\left(\widehat{\beta_1^*}\right),\end{aligned}$$

and we see that $\widehat{\beta_1^*}$ has minimal variance amongst all linear unbiased estimators. This finishes the proof that $\widehat{\beta_1^*}$ is a BLUE. □

3.1.2 Multiple Linear Regression

Next we look at the multiple linear model to generalise the insights we made above to more than just one explanatory variables. The model is still univariate, i.e., we only consider one dependent response variable. In this section we mainly follow [1], Chapter 7.

The multiple univariate linear model is

$$y = X\beta + U \tag{3.8}$$

where

$$y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}.$$

The number of observations is still denoted by n, with k being the number of independent regressor variables and $n \geq k$. This means that each row of X stands for an observation of all the independent variables, while a column denotes all observations for a specific one.

"Multiple" here means that we now have multiple regressor variables, contrary to the simple linear model which only employs one explanatory variable. "Univariate" implies that we still estimate only one regressand for each observation. On the other hand we would have a "multivariate" model, if we estimated a vector of dependent variables for each observation. However we will not consider this type of models in this work.

Again we ask ourselves how to find estimators that minimise the least squares problem of the resulting residuals.

$$SSR(\beta) = \|y - X\beta\|_2^2 = (y - X\beta)^T (y - X\beta) = y^T y - 2\beta^T X^T y + \beta^T X^T X\beta \quad (3.9)$$

We see that (3.9) is a convex problem.

This is as every norm $f(t) = \|t\|$ is convex, $g(t) = \begin{cases} 0 & \text{for } t < 0 \\ t^2 & \text{for } t \geq 0 \end{cases}$ is convex and non-decreasing, thus $g \circ f$ is convex. Furthermore the composition $h \circ T$ of a convex function h and an affine function T is convex as well. Here we have $T(\beta) = y - X\beta$, which is affine apparently. So we conclude, that $SSR(\beta) = (g \circ f \circ T)(\beta)$ is convex. Differentiating (3.9) with respect to β and setting the equation to zero gives the first order conditions for a minimiser,

$$\frac{\partial SSR(\beta)}{\partial \beta} = -2X^T y + 2X^T X\beta \overset{!}{=} 0,$$

which is clearly equivalent to the ordinary least squares (OLS) normal equations

$$X^T X\beta = X^T y. \quad (3.10)$$

As long as X has full column rank k, then $X^T X$ is non-singular and the OLS solution $\widehat{\beta^*}$ is

$$\widehat{\beta^*} = \left(X^T X\right)^{-1} X^T y. \quad (3.11)$$

Under the Assumptions 1-4 we encountered in the previous section, we again can derive the same useful properties for multiple linear regression.

1. $\widehat{\beta}$ is an unbiased estimator for β.
2. $\widehat{\beta}$ is a consistent estimator for β.
3. $\widehat{\beta}$ is a best linear unbiased estimator (BLUE) for β.

Proofs for the above statements can be found for example in [1].

3.1.3 Coefficient of Determination - R^2

The above is motivated by the goal of estimating the dependent variable and then predicting unknown outcomes. To assess the quality of predictions made by the model, we would like to judge how good a certain linear model describes the data. We accomplish this by introducing the "coefficient of determination", denoted by R^2, which we define as the proportion of the squared sum explained by the regression compared to the entire squared sum to be explained. The coefficient of determination indicates how much of the variance in the regressand can be predicted from the regressors. Thus this entity should be able to measure how well the given data is described. In the following we will go by [5], Chapter 3.6 and [1], Chapter 4.5.

Definition 3.3. The coefficient of determination is formally defined as

$$R^2 = \frac{\sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2},$$

where $\widehat{Y}_i = X\widehat{\beta}$ and $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

We assume $\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 \neq 0$, which is satisfied as long as there exist $i, j \in \{1, ..., n\}$ such that $Y_i \neq Y_j$. Equivalently, we get

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2} \tag{3.12}$$

as

$$\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n} \left(\left(\widehat{Y}_i - Y_i\right) - \left(\overline{Y} - Y_i\right)\right)^2 = \sum_{i=1}^{n} \left(\widehat{Y}_i - Y_i\right)^2 + \sum_{i=1}^{n} \left(\overline{Y} - Y_i\right)^2.$$

Let SST, SSE and SSR denote the sum of squared total deviations, the sum of squared explained deviations and the sum of squared residuals, respectively. We have

$$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2, \quad SSE = \sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y}\right)^2, \quad SSR = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2.$$

Then the definition is equivalent to $R^2 = \frac{SSE}{SST}$, exactly what was our motivation. The second formulation is equivalent to $R^2 = 1 - \frac{SSR}{SST}$.

Thus, $0 \leq R^2 \leq 1$. On the one hand if we have vanishing residuals, then $SSR = 0$ and $R^2 = 1$. This means that the data is described perfectly by the regression. On the other hand if the explained variance is vanishing, i.e., the explanatory variables do not contribute to explaining the total variance. This implies $SSE = 0$ and thus $R^2 = 0$.

Therefore, R^2 measures the proportion of explained deviations relative to the total deviations. As the ordinary least squares problem minimises the sum of squares of residuals (SSR), adding additional explanatory variables to the linear model will lead to SSR less or equal compared to before. That means that incorporating more regressors potentially increases the coefficient of determination. This becomes relevant when we want to compare several models with different amounts of regressors, since this would potentially favour models with more independent variables. To take this into account we introduce a modified version, the adjusted coefficient of determination $\bar{R}^2$,

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2} \frac{n-1}{n-k} \tag{3.13}$$

The more explanatory variables we have in our model, the more it will be punished in the derivation of $\bar{R}^2$. This way, we try to balance the gain in the coefficient of determination by adding variables with the loss of degrees freedom. Hence an added variable will increase $\bar{R}^2$ only if the reduction in the residuals outweighs this loss. Using the definition we deduce the following relation between $\bar{R}^2$ and R^2,

$$1 - \bar{R}^2 = \left(1 - R^2\right) \frac{n-1}{n-k} \tag{3.14}$$

The above considerations assumed that the linear model includes an intercept. The statements are no longer true, when there is none. For example all of a sudden there are possible situations with $R^2 < 0$.

Definition 3.4. To deal with this, we define the un-centred coefficient of determination,

$$R_u^2 = \frac{\sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n} Y_i^2}. \tag{3.15}$$

3.2 Sample Average Approximation

The second and final topic we are going to cover in this section is the so-called “Sample Average Approximation”. It is part of the domain of stochastic programming, which is a framework to treat optimisation problems that involve uncertainty. Stochastic programming in general is an immensely current subject and there is a lot of literature that evolved over the last decades. In this work we will mainly follow [6], in particular Chapter 5, where all mentioned theorems and proofs can be found. We start with the formulation of the type of problem which will be relevant for this work and conclude this chapter with establishing some interesting statistical properties.

3.2.1 Formulation of SAA

Consider the general two-stage stochastic program, which can be written as follows

$$\min_{x \in X} \; \{f(x) := \mathbb{E}\left[F(x,\xi)\right]\}, \tag{3.16}$$

where $F(x,\xi)$ is the optimal value of the second-stage problem

$$\min_{y \in \mathcal{G}(x,\xi)} \; g(x,y,\xi). \tag{3.17}$$

Let $X \neq \emptyset$, $X \subseteq \mathbb{R}^n$ be a non-empty closed subset of $\mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^m \times \Xi \to \mathbb{R}$ the objective function of the second-stage problem and $\mathcal{G} : \mathbb{R}^n \times \Xi \rightrightarrows \mathbb{R}^m$ denote the set-valued function describing the constraints of the second-stage problem. Additionally, ξ is a random vector with probability distribution p, which has support on the set $\Xi \subseteq \mathbb{R}^d$, and $F : X \times \Xi \to \mathbb{R}$. In the following we assume that the expected value $f(x)$ is well defined and finite valued for all $x \in X$. This means that $F(x,\xi)$ is finite for almost every $\xi \in \Xi$ and for all $x \in X$.

In the above formulation (3.16)-(3.17) we call $x \in \mathbb{R}^n$ the first-stage decision variable and $y \in \mathbb{R}^m$ the second-stage decision variable. At the first stage we have to make a "here-and-now" decision x before the particular realisation of the random vector ξ is available. After we get to know the realisation of ξ, i.e., at the second stage, we solve the regarding optimisation problem to derive an optimum.
The underlying idea of such a two-stage stochastic program is, that we should base optimal decisions solely on data which is available at the time we have to make the decision and cannot depend on future observations. This type of problem is widely used in stochastic optimisation, as it already incorporates many essential aspects of dealing with uncertainty while still maintaining a rather simple structure (compared to more complex multi-stage problems).
One of the most looked at problems in the history of two-stage programs is the two-stage stochastic linear program which was originally investigated by Dantzig and Beale

(see [7] and [8], respectively),

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^T x + \mathbb{E}\left[Q(x,\xi)\right] \\ \text{s.t.} \quad & Ax = b, x \geq 0, \end{aligned} \tag{3.18}$$

where $Q(x, \xi)$ is the optimal value of

$$\begin{aligned} \min_{y \in \mathbb{R}^m} \quad & q(\xi)^T y \\ \text{s.t.} \quad & T(\xi)x + W(\xi)y = h(\xi), y \geq 0. \end{aligned} \tag{3.19}$$

We see that we obtain the linear problem (3.18)-(3.19) from the general problem (3.16)-(3.17) by setting

$$\begin{aligned} & X := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}, \\ & g(x,y,\xi) := c^T x + q(\xi)^T y, \\ & \mathcal{G}(x,\xi) := \{y \in \mathbb{R}^m : T(\xi)x + W(\xi)y = h(\xi), y \geq 0\}. \end{aligned}$$

Already in its simplest form (3.18)-(3.19) we ask ourselves, how to derive a solution in the easiest way.
Suppose that the distribution of ξ has finite support, i.e., ξ has a finite number of possible realisations $\xi^1, \ldots, \xi^J$ with respective probabilities $p_1, \ldots, p_J$. Then the expected value in the first stage problem (3.18) simplifies to a finite sum

$$\mathbb{E}\left[Q(x,\xi)\right] = \sum_{j=1}^{J} p_j Q(x, \xi^j). \tag{3.20}$$

This means, that for given x the expected value $\mathbb{E}\left[Q(x,\xi)\right]$ is equal to the optimal value of the linear program

$$\begin{aligned} \min_{y_1,\ldots,y_J \in \mathbb{R}^m} \quad & \sum_{j=1}^{J} p_j q(\xi^j)^T y_j \\ \text{s.t.} \quad & T(\xi^j)x + W(\xi^j)y_j = h(\xi^j), \\ & y_j \geq 0, \quad j = 1, \ldots, J. \end{aligned}$$

Moreover, the two-stage stochastic linear program (3.18)-(3.19) collapses to one (potentially very) large deterministic linear program

$$
\begin{aligned}
\min_{x,y_1,\ldots,y_J} \quad & c^T x + \sum_{j=1}^{J} p_j q_j^T y_j \\
\text{s.t.} \quad & T_j x + W_j y_j = h_j, \quad j = 1, \ldots, J, \\
& Ax = b, \\
& x \geq 0, \quad y_j \geq 0, \quad j = 1, \ldots, J,
\end{aligned}
\tag{3.21}
$$

where $q_j := q(\xi^j)$, $T_j := T(\xi^j)$, $W_j := W(\xi^j)$ and $h_j := h(\xi^j)$.
The same idea can be used for the general two-stage problem (3.16)-(3.17) as well. This means, by discretising the expected value we arrive at a deterministic problem which can be solved (at least numerically). However even crude discretisation of some continuously distributed random vector $\xi \subseteq \mathbb{R}^d$ leads to an exponential growth of the amount of scenarios with increase of its dimension d. One possible approach of dealing with this issue is to produce a manageable number of "representative" scenarios.

For instance, we can randomly generate a sample of N realisations $\xi^1, \ldots, \xi^N$ of the random vector ξ. This sample could either be produced in the computer, e.g., by Monte Carlo sampling techniques, or interpreted as historical data of observations of the random vector ξ.
For any $x \in X$ we can estimate the expectation $f(x)$ by taking the average of $F(x, \xi^j)$, $j = 1, \ldots, N$. This motivates the so-called "sample average approximation" (SAA),

$$
\min_{x \in X} \left\{ \widehat{f}_N(x) := \frac{1}{N} \sum_{j=1}^{N} F(x, \xi^j) \right\}, \tag{3.22}
$$

of the original first-stage problem (3.16). Furthermore, the sample average function $\widehat{f}_N$ can be viewed as the expectation with respect to the empirical distribution $P_N = \frac{1}{N} \sum_{j=1}^{N} \delta(\xi^j)$, where $\delta(z)$ denotes measure of mass one at the point z, as

$$
\widehat{f}_N(x) = \mathbb{E}_{P_N}\left[F(x, \xi)\right].
$$

Thus, for a given sample we can interpret the SAA problem (3.22) as a stochastic program with scenarios $\xi^1, \ldots, \xi^N$ and respective probabilities $p_1 = \cdots = p_N = \frac{1}{N}$ (compare (3.20)).

3.2.2 Statistical Properties of SAA

In the following the sample $\xi^1, \dots, \xi^N$ can denote either a sequence of random vectors, or a particular realisation of that sequence. Which of the two meanings is used in an individual situation will be clear from the context.
We assume that each of the random vectors ξ^j has the same distribution as the original vector ξ, i.e., they are identically distributed. We say that the sample is "independently identically distributed" (iid), if in addition each sample ξ^j is distributed independently of the others.

As the sample average function $\widehat{f}_N(x)$ was motivated by discretising the true expected value function $f(x)$, we are interested in whether the sample average approximation gives the true values in the long run.
By the "Law of Large Numbers" (LLN) we obtain the following properties.

Proposition 3.4. *Suppose the samples $\xi^1, \dots, \xi^N$ are iid, then for every fixed $x \in X$ the following statements hold.*

(a) $\lim\limits_{N\to\infty} \widehat{f}_N(x) = f(x)$ almost surely (a.s.), i.e., we have pointwise convergence a.s..

(b) $\mathbb{E}\left[\widehat{f}_N(x)\right] = f(x)$ for all $N \geq 1$, i.e., $\widehat{f}_N(x)$ is an unbiased estimator of $f(x)$.

Under some additional conditions we even get uniform convergence. The following theorem is taken from [6], Chapter 7, where a proof can be found, too.

Theorem 3.2. *Let X be a non-empty compact subset of $\mathbb{R}^n$ and suppose that:*

(a) For any $x \in X$ the function $F(\cdot, \xi)$ is continuous at x for almost every $\xi \in \Xi$.

(b) $F(x, \xi)$, for $x \in X$, is dominated by an integrable function.

(c) The samples are iid.

Then the expected value function $f(x)$ is finite valued and continuous on X, and $\lim\limits_{N\to\infty} \sup\limits_{x\in X} |\widehat{f}_N(x) - f(x)| = 0$ a.s., i.e., we have uniform convergence a.s..

As we saw above, we have pointwise as well as uniform convergence of $\widehat{f}_N(x)$ to $f(x)$ a.s. for $N \to \infty$. Furthermore, the sample average function is an unbiased estimator for the true expected value function. Therefore, it is natural to hope for the optimal value and the set of optimal solutions of the SAA problem (3.22) to converge to their respective counterparts of the original problem (3.16). We denote by ϑ^* and S the

optimal value and the set of solutions of the true problem (3.16), respectively. By $\widehat{\vartheta}_N$ and $\widehat{S}_N$ we denote the optimal value and the set of solutions of the sample average problem (3.22), respectively.
In the following we will discuss statistical properties of the SAA estimators $\widehat{\vartheta}_N$ and $\widehat{S}_N$.

Consistency of Solution Value

First, let us consider the question of *consistency* of the SAA estimator of the optimal value ϑ^*. Recall, that an estimator $\widehat{\alpha}_N$ of a parameter α is consistent if $\widehat{\alpha}_N$ converges to α in an appropriate sense (compare Definition 3.2). For any fixed $x \in X$ we have $\widehat{\vartheta}_N \leq \widehat{f}_N(x)$. Thus, if we have pointwise convergence due to the Law of Large Numbers then

$$\limsup_{N\to\infty} \widehat{\vartheta}_N \leq \lim_{N\to\infty} \widehat{f}_N(x) = f(x) \quad \text{a.s.}.$$

Therefore, under the assumption of pointwise convergence we obtain

$$\limsup_{N\to\infty} \widehat{\vartheta}_N \leq \vartheta^* \quad \text{a.s.}. \tag{3.23}$$

Without additional conditions on $F(x,\xi)$, the above inequality can be strict. The following theorem and its proof can be found in [6], Chapter 5.

Theorem 3.3. *Suppose that $\widehat{f}_N(x)$ converges to $f(x)$ uniformly on X almost surely (a.s.).*
Then $\widehat{\vartheta}_N$ converges to ϑ^* a.s. as $N \to \infty$.

Proof. Uniform convergence a.s. of $\widehat{f}_N(x) = \widehat{f}_N(x,\omega)$ to $f(x)$ (here we make the dependence an the element ω of the probability space Ω explicit) means that for any $\varepsilon > 0$ and almost every $\omega \in \Omega$ there is some index $N^* = N^*(\varepsilon,\omega)$ such that the following inequality holds for all $N \geq N^*$:

$$\sup_{x\in X} |\widehat{f}_N(x,\omega) - f(x)| \leq \varepsilon.$$

Then it follows for almost every ω that $|\widehat{\vartheta}_N(\omega) - \vartheta^*| \leq \varepsilon$ for all $N \geq N^*$, which completes the proof. □

Consistency of Solution Set

Next we are going to show that the SAA estimators for the optimal solutions, i.e., the set of solutions S, are consistent as well. To measure the deviation between two sets A and B, we introduce the following entities.

Definition 3.5. Let $A, B \subseteq \mathbb{R}^n$ be sets.

1. For $x \in \mathbb{R}^n$ we denote the distance between x and A by

$$\operatorname{dist}(x, A) := \inf_{y \in A} \|x - y\|.$$

2. We denote the deviation of the set A from the set B by

$$\mathbb{D}(A, B) := \sup_{x \in A} \operatorname{dist}(x, B).$$

To establish consistency of $\widehat{S}_N$ we need stronger conditions than for $\widehat{\vartheta}_N$. The following theorem as well as its proof can also be found in [6], Chapter 5.

Theorem 3.4. *Suppose that there exists a compact set $C \subseteq \mathbb{R}^n$ such that the following conditions hold.*

(a) The set S of optimal solutions of the true problem is non-empty and is contained in C.

(b) The function $f(x)$ is finite valued and continuous on C.

(c) $\widehat{f}_N(x)$ converges to $f(x)$ uniformly in $x \in C$ a.s., as $N \to \infty$.

(d) With probability 1 for N large enough the set $\widehat{S}_N$ is non-empty and $\widehat{S}_N \subseteq C$.

Then $\widehat{\vartheta}_N \to \vartheta^*$ and $\mathbb{D}(\widehat{S}_N, S) \to 0$ a.s. as $N \to \infty$.

Proof. Assumptions (a) and (d) yield that the true problem as well as the SAA problem can be restricted to the set $X \cap C$. Thus, without loss of generality we can assume that already the set X is compact
By Theorem 3.3 we conclude that $\widehat{\vartheta}_N \to \vartheta^*$ a.s., for $N \to \infty$.
It suffices to show that $\mathbb{D}(\widehat{S}_N(\omega), S) \to 0$ for almost every $\omega \in \Omega$, such that $\widehat{\vartheta}_N(\omega) \to \vartheta^*$ and the assumptions hold, as $N \to \infty$. For simplicity of notation we omit the attribute of "almost surely" in the following. We now prove the statement by contradiction.
Let us suppose that $\mathbb{D}(\widehat{S}_N, S) \nrightarrow 0$ as $N \to \infty$. Then there exists a subsequence $(\widehat{x}_{N_k})_{k \in \mathbb{N}}$, where $\widehat{x}_N \in \widehat{S}_N$ for $N \geq 1$, such that $\operatorname{dist}(\widehat{x}_{N_k}, S) \geq \varepsilon$ for all $k \geq 1$ and some $\varepsilon > 0$. As X is compact we can assume that $\widehat{x}_{N_k} \to x^*$ as $k \to \infty$ for some point

$x^* \in X$, if necessary by passing to another subsequence. Obviously we have $x^* \notin S$ and thus $f(x^*) > \vartheta^*$. Additionally we have $\widehat{\vartheta}_{N_k} = \widehat{f}_{N_k}(\widehat{x}_{N_k})$ for all $k \geq 1$. This yields for all $k \geq 1$,

$$\widehat{f}_{N_k}(\widehat{x}_{N_k}) - f(x^*) = (\widehat{f}_{N_k}(\widehat{x}_{N_k}) - f(\widehat{x}_{N_k})) + (f(\widehat{x}_{N_k}) - f(x^*)).$$

The first term of the right-hand side goes to zero by Assumption (c), while the second term vanishes due to continuity of $f(x)$ (Assumption (b)). This implies that $\widehat{\vartheta}_{N_k}$ tends to $f(x^*) > \vartheta^*$ as $k \to \infty$, which is a contradiction. □

These results close the section about necessary preliminaries. Next we are going to explain and analyse the problem we are going to work on. Subsequently we are going to use the mathematical concepts established above to obtain satisfactory outcomes.

4 Problem and Results

4.1 Problem description

4.1.1 Setting

Before we go into further details we need to fully understand the problem and what exactly is the task.

The electricity producing company we worked with runs three wind parks, which all participate in the daily spot market. This means the electricity sold concerns the production of the next day only - opposing to a futures market in which goods are delivered at a (much) later date. As a third type there is the intra-day market, dealing with sales within the day the energy is produced.
To ensure the stability of the electrical grid, every participant of the spot market has to announce the next day's scheduled hourly production values to the transmission system operator (TSO). The TSO is an entity responsible to transmit the electrical power from the power generating plants to regional electricity distribution operators who then distribute the power locally. The TSO's task is to coordinate the supply of and demand for electricity to avoid shortage or surplus which would lead to black-outs in (at least) major parts of the transmission system and therefore cause personal and property damages.

After every participant submitted their estimated production for the next 24 hours, a day-ahead price per MWh separately for each hour of the following day is determined. Each participant receives money according to the scheduled amount of energy at this day-ahead price. If one produces more energy than announced, one gets a (usually) lower price for the amount of exceeding energy and thus faces opportunity costs. If one produces less, then a (usually) higher penalty for each MWh of deficit energy is inflicted.
Due to netting effects with other participants in the same balancing zone, one gets a crucially higher price for exceeding energy and lower penalty for deficit energy compared to the price/penalty one would get in the control area alone. A balancing zone is a union of several electricity producing participants within one control area. These zones are the basis for determination of the actually necessary balancing energy for each zone, which is usually less than the total amount of all participants separately. The company's wind parks operate all in the same balancing zone, thus the prices and penalties are the same for each park. In the following we will only consider the prices

and penalties for the balancing zone.
Because of the modalities of the spot market and the billing inside the balancing zone we get to know the day-ahead price on the same day already, but the price for exceeding energy and the penalty for deficit energy only about two months later. We get to know the day-ahead price after submission of the production schedule, though. Thus the most recent price data available before submitting a schedule is from the previous day regarding the day-ahead price and from 2-3 months ago regarding balancing energy prices.

In general there is an intra-day market in which exceeding and deficit energy is bought and sold within one day in real-time. However in our specific setting, this is not relevant for us and it would be out of scope to take into account. This simplifies the situation considerably - we submit 24 values of estimated electricity production once a day and then we cannot intervene any more during the same day.
Furthermore we could have either positive or negative prices for day-ahead and exceeding energy, as well as either positive or negative penalties for deficit energy in general. In our situation we only have positive prices and penalties, though. Moreover the balancing energy prices are not symmetric. More precisely the opportunity cost for exceeding energy is not the same as the loss due to deficit energy.
This again implies a simplification - the wind power stations produce the entire amount of energy that is possible at any time. Due to both the day-ahead price and the price for exceeding energy always being positive, one never decreases the production of the wind power stations voluntarily. This is because every MWh produced gains more revenue - no matter whether it is exceeding our schedule or not.
As a consequence, we do not have to think about sticking to a certain schedule and how to control production. Only due to external influences a decrease of production is carried out. These can be for example technical failure (like a broken generator or icing of the wind turbines' rotor blades) on the one hand, as well as limitation by the TSO on the other hand. The TSO has the authority to decrease one's production to the submitted value in case of an imminent collapse of the transmission system to regain balance of the electrical grid.

In addition to the revenue generated by selling electrical energy on the spot-market, one earns a so-called *Green Certificate* (GC) for each MWh of submitted and actually produced energy. The number of GCs earned is determined as follows.
For each month the total amount of scheduled and produced energy is aggregated and

then rounded down to the nearest integer less or equal this number. This then is the quantity of GCs for that month. To simplify our considerations we ignore the rounding at the end of the month and calculate with fractions of GCs for each hour. This is not too far off, since the difference is less than one GC for a whole month at most.
For the following, we assume one GC is worth a fixed price g approximately equal to one MWh at an average day-ahead price. This means that Green Certificates are quite valuable and influence the total revenue significantly.

4.1.2 Strategic Adjustment - Conceptualisation

To recap the situation, we submit an estimated production forecast for the next day and deviation from it will be penalised. We sell the amount of scheduled energy at the day-ahead price da one day in advance. The surplus compared to the schedule is sold at the price exc and for deficit energy - energy we predicted to produce but actually did not - we have to pay the penalty def.

Definition 4.1. We summarise,
da := day-ahead price per MWh,
exc := price for exceeding energy per MWh,
def := penalty for deficit energy per MWh.

As mentioned above, the TSO is responsible to maintain stability of the grid. To motivate the participants to approximately stick to their schedules and thus make it easier for the TSO, deviation from the submitted schedule is disadvantageous. That means that exceeding energy is worth less than scheduled one and the penalty for deficit energy is higher than the day-ahead price.
Therefore we usually have that $exc < da < def$.

Definition 4.2. Now we can define the two spreads
a := da - exc,
b := def - da.

The entity a corresponds to revenue which was not realised, i.e., the opportunity cost, for each MWh of produced energy exceeding the amount that was estimated the day before. Spread b is the penalty, i.e., the loss, for each MWh of predicted but not produced energy. To clarify, we sell the entire scheduled energy at the day-ahead price da and only the deviating amount of energy gets penalised in one of the two ways.

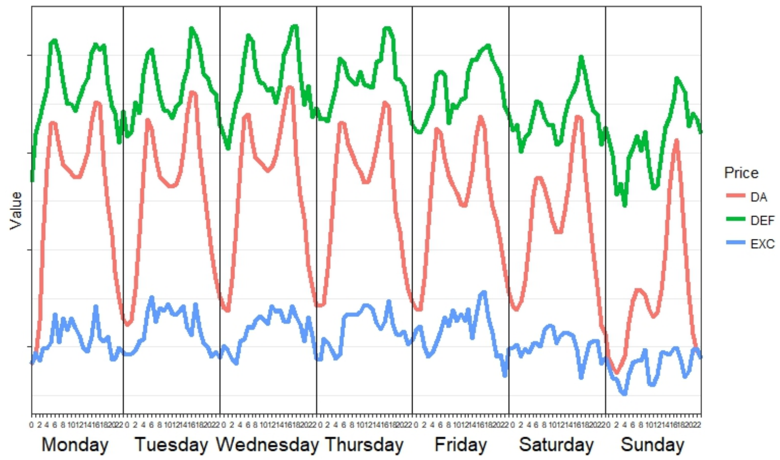

Figure 4.1: Average market prices in 2016

We see in Figure 4.1 that throughout a day the three entities da, def and exc behave differently and have different relative position to each other. We also recognise that usually we have $a > 0$ and $b > 0$. This is to ensure that all participants submit realistic schedules and are not tempted to speculate against the market. For example one could hope that $def < da$ and thus submit maximal possible schedules without any practical foundation. However if we have a look at Figure 4.2, we see that $a < 0$ or $b < 0$ can occur as well - even on average over a whole year.

Nevertheless we distinguish two different cases:

1. $a > b$: The opportunity cost is greater than the loss for deficit energy which means that we rather try to avoid overproduction.

2. $a < b$: The loss for deficit energy is greater than the opportunity cost which implies that we rather want to avoid underproduction.

As a matter of fact, production of electrical energy by wind power stations is subject to uncertainty. We receive a wind power forecast for the next day computed by a specialised company. Naturally, the actual production deviates from the forecast. The more exact we know the next day's production the less loss we have due to opportunity costs or deficit penalties. Thus it is highly beneficial to have the best possible forecast and this could be an issue to pursue. However developing a better wind power forecast is out of scope.

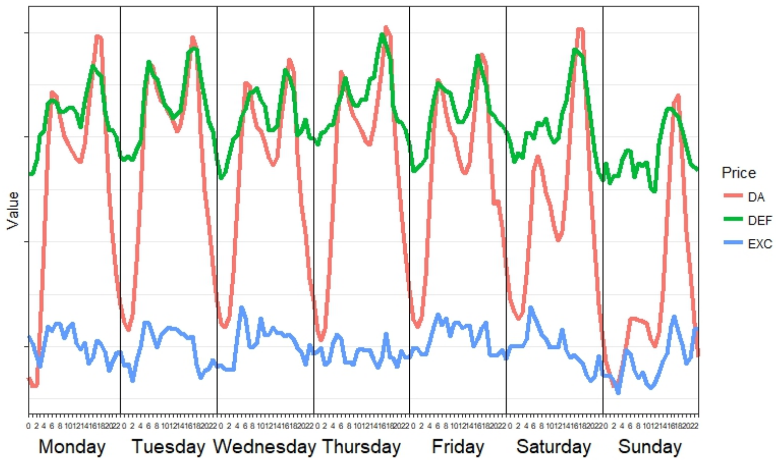

Figure 4.2: Average market prices in 2014

As participants of the market we are interested in maximising revenue and thus minimising penalisation. Unfortunately we cannot improve the wind power forecast by ourselves, but we know about the price development throughout different weekdays. We already figured out in which times it is better to avoid underproduction or overproduction. Therefore the idea is to make a "strategic adjustment" of the wind power forecast before submitting the production schedule to the TSO - solely depending on the current price situation of the market.

We obtain the following strategies for the two cases mentioned above

1. $a > b$: We rather avoid overproduction, so we increase the scheduled production compared to the forecast before submission.

2. $a < b$: We rather avoid underproduction, so we decrease the scheduled production compared to the forecast before submission.

Moreover, we already saw that GCs are highly relevant for the revenue. Thus we intend not to miss any possible GC due to produced, but not scheduled energy. We obtain a third strategy for strategic adjustment.

3. GC: We try to avoid missing possible Green Certificates, so we increase the scheduled production compared to the forecast before submission.

These considerations give rise to establish an "optimal" strategic adjustment policy.

4.1.3 Company's Status Quo

Before we think of own ideas about how to construct sensible adjustment policies, it may be advantageous to analyse how the company's strategic adjustment was carried out in the past.
Energy economics is an immensely fast-moving domain, so the more recent the data we use the better we can estimate the current behaviour. Here lies the main problem - the moment we have to announce the production schedule for the next day we know the previous day's da, but exc and def only from two to three months before. Therefore we need to forecast the prices in a certain way.

This was done so far by calculating an average historic week of the current quarter, separately for each of the entities da, exc and def. That means that same weekdays have the same forecast, so we display the prices' behaviour on different weekdays. However we cannot take into account changing conditions during a quarter of a year. From these hourly values a set of $168\,(= 7 \cdot 24)$ adjustment coefficients, from Monday 00:00(-01:00) until Sunday 23:00(-24:00), was computed to increase/decrease the estimated production relative to the forecast before submission. These coefficients are the same for all three parks, since they all participate in the same balancing zone which means they are subject to the same prices.
Besides that the company utilises a domain-dependent weighting for their adjustment coefficients, too. The idea is to decrease the absolute change in the high-end domain of production where relative adjustment coefficients have more (absolute) impact.

The wind power forecast we receive from the specialised company is for turbine level only. That means that the "raw" forecast needs a pre-adjustment itself to incorporate the losses due to physical reasons, e.g., losses during transmission from the turbine to the entry point into the grid. These losses are subject to the specific energy level at that time and are integrated by a heuristic that proved itself plausible over the years. The raw forecast less all losses is then the actual estimation what we may produce the next day.

Definition 4.3. We call this entity
meta := The raw production forecast less all potential losses.

For our investigations this is the essential value, since we have to accept the physical losses. Thus we exclusively consider *meta* for strategic adjustment.

Definition 4.4. In addition we call
final := The strategically adjusted *meta*, i.e., the value that is submitted to the TSO.

Strategic adjustment has two natural bounds, though. On the one hand $final$ must always be non-negative, on the other hand we must not submit $final$ greater than the physical upper production limit of each park. This is the amount of energy one park can generate at full-load during one hour.

Definition 4.5. Moreover we call
prod := The actual production, i.e., the amount of energy that was measured at the entry point of the grid.

4.1.4 Delivered Data

At the beginning of the cooperation we received several types of necessary data from the company:

1. Production data:
   a) hourly $final$ values from 2014-01-01 (and 2014-06-01) until 2018-03-12
   b) hourly $meta$ values from 2016-03-17 until 2018-03-12
   c) hourly $prod$ values from 2014-01-01 (and 2014-06-01) until 2017-12-31
2. Price data:
   a) hourly da, exc and def values from 2012-01-01 until 2017-12-31
3. Adjustment data:
   a) adjustment coefficients for each hour of each weekday for all quarters of the year from 2014-01-01 until 2018-03-12
   b) adjustment algorithm to compute $final$ with $meta$ and an adjustment coefficient as input

One of the three wind parks went on line later than the others, so we have reliable data for $final$ and the actual production starting at two different dates.
The data was partitioned in several hundred different files, with time-stamps in a couple of time zones and various file structures even within the same type of data due to style changes throughout the years. The first main task was to read in, tidy up and standardise all the data before doing further actions.

Towards the end of the work on this subject we additionally got time series of potential production at turbine level that indicate at which time actual limitation by the

TSO occurred. They also carry the information when there was potential limitation, i.e., there was no limitation in reality, but if the actual production would have exceeded the scheduled production then there would have been limitation. However this data is not $prod$ at the entry point of the grid, but the raw production data.

Future Recommendations 1. Utilisation of this data would have needed quite a lot more consideration. For example one would need to find a domain-dependent loss function to propagate the potential values from turbine level to the entry point for each park separately. We came across this pretty late, so it was not taken into account in the following. However this could be covered in another work in the future.

4.1.5 Goals

The company wanted several topics to be analysed and investigated regarding potential improvements.

- Strategic adjustment on holidays
- Limitation by the TSO
- Optimal strategic adjustment

4.2 Analysis

4.2.1 Framework for Evaluation

Before we start analysing the concept of "strategic adjustment" and investigating the different topics we are confronted with, we need to establish a framework for evaluation. Of course we can compare the relative differences in the revenue due to two different adjustment policies, but we would like to have an objective reference to judge the merit of policies independently. A possible "best case" would lend itself to this task, so we need to figure out what a best case scenario can be. The idea is to rate every policy according to the percentage of revenue it gains, e.g. the best case itself has a rating of 100 while zero revenue has a rating of 0.

That means we need to clarify of which parts the revenue consists. Since billing takes place on an hourly basis, we assume $meta$, $final$ and $prod$ as well as da, exc and def to be regarding one certain hour. Actually, the above entities are different for every hour so they should have a subscript indicating a time stamp. Since all the problems are independent of each other, though, we omit the index for simplicity.

Definition 4.6. Let $final$ be the scheduled production, $prod$ the actual production and da, exc and def the prices. The price for one GC is denoted by g.

1. The revenue for the produced energy is

$$\texttt{rev}_{NRG} := final \cdot da + [prod - final]^{+} \cdot exc - [final - prod]^{+} \cdot def, \tag{4.1}$$

where $[a]^{+} = \max(a,\, 0)$ denotes the positive part of a.

2. The revenue for the earned Green Certificates is

$$\texttt{rev}_{GC} := \min(final,\, prod) \cdot g. \tag{4.2}$$

3. Then the total revenue is

$$\texttt{rev} := \texttt{rev}_{NRG} + \texttt{rev}_{GC} \tag{4.3}$$

We see that

$$\begin{aligned}
\texttt{rev} &= \texttt{rev}_{NRG} + \texttt{rev}_{GC} \\
&= final \cdot da + [prod - final]^{+} \cdot exc - [final - prod]^{+} \cdot def + \ldots \\
&\qquad \ldots + \min(final,\, prod) \cdot g \\
&= final \cdot (da + g) + [prod - final]^{+} \cdot exc - [final - prod]^{+} \cdot (def + g),
\end{aligned} \tag{4.4}$$

since

$$\begin{aligned}
\min(a,\, b) &= a + \min(b - a,\, 0) \\
&= a - \max(a - b,\, 0) \\
&= a - [a - b]^{+}.
\end{aligned}$$

We can formulate the problem including GCs in the same way as the original one, but with shifted day-ahead price $\overline{da} = da + g$ and deficit penalty $\overline{def} = def + g$. Thus we do not have to include GCs with a separate expression, which is pretty convenient for the formulation of the problem. Also excluding GCs from our considerations would lead to potentially different results, since the value of one Green Certificate is approximately as much as one MWh at an average day-ahead price.

4.2.2 Best Case

Now we can think about a best possible case.
If we have a normal market situation (see Figure 4.1) then we have $da > exc$ as well as $da < def$. This implies that we get the best result if we schedule the exact amount that is actually produced on the next day, which requires perfect foresight of the next day's production.

Definition 4.7. We obtain the above described best case by
$final^b := prod.$

However this is not the absolutely best possible revenue we can get. If you look at Figure 4.2, you can see that there are hours when $da \leq exc$ or $da \geq def$. This implies, that it may be beneficial to announce less than produced (because of a higher price for exceeding energy) or as much as we are allowed - the physical upper production limit (because of a lower penalty than the day-ahead price). Although this would require not only perfect foresight about production but also about the prices we get to know up to three months later. Assuming perfect foresight of nature is already quite unrealistic. Additional perfect foresight regarding all other participants, which would be necessary to determine exc and def, is just overcharged - even for theoretical considerations.

This best case now serves as objective reference for our policies. We do this by aggregating one policy's revenue over a fixed period of time and then taking the percentage compared to the best-case's revenue over the same period as a rating.

4.2.3 Strategic Adjustment in the Past

First we want to know whether strategic adjustment was beneficial at all and if so, how big the difference is to no adjustment before submission. For this, we need to know the original $meta$ for all hours and take them as (not-)adjusted scheduled production $final$.

Hence for all those hours in which we do not have $meta$ available, we need to invert the adjustment algorithm that takes $meta$, the adjustment coefficient and a domain-dependent factor as input. Unfortunately this algorithm contains a step that is not injective, thus there are potential cases with ambiguous inversion. By convention we take the lowest possible value and this shows to produce $meta$ close enough to the original ones - at least in the periods in which we can compare. "Close enough" in this case means that the revenue gained after using the calculated $meta$ is approximately the same compared to the original ones. To be consistent we take the recalculated *meta* values for the whole time span we investigate.

The introduction of strategic adjustment was due to several reasons combined. First of all it should capture the behaviour of the market and reduce the risk of inflicting greater loss than necessary. Moreover it was set to be non-negative throughout many periods in the past to on the one hand do not risk the gain of Green Certificates (we cannot get more than the number of MWh we scheduled) and on the other hand to avoid limitation of production by the TSO (they can only limit us to the submitted value).

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
No Adjustment	91.02	90.98	89.82	90.45

Table 4.1: Ratings in percentage of the best case (including GCs)

To judge the results displayed in Table 4.1 correctly, we need to consider that for the absolute figures even tenths of percentage points of the best case lead to significantly more revenue. So we see that strategic adjustment over the last four years was highly beneficial, since the improvement in percentage points of the best case was 1.98 %, 1.15 %, 0.35 % and 1.00 % in 2014, 2015, 2016 and 2017, respectively.

Adjustment Policy	2014	2015	2016	2017
Status Quo	86.49	84.47	83.19	88.37
No Adjustment	84.24	83.42	82.73	87.12

Table 4.2: Ratings in percentage of the best case (excluding GCs)

Also if we compare these ratings with those excluding the Green Certificates (see Table 4.2), we see that the value of the earned Green Certificates is immense. In the past there were periods where GCs had a valuation of around zero in practice, so it is pleasant to see that strategic adjustment also works in this setting. Furthermore we see that including them into computation of the revenue shrinks the difference between the two policies in general. The improvement in 2014, 2015, 2016 and 2017 was 2.25, 1.05, 0.46 and 1.25 percentage points of the best case, respectively.
To avoid confusion we will look at policies including Green Certificates only in the following.

4.2.4 Strategic Adjustment on Holidays

One of the company's main issues is strategic adjustment on public holidays. The experience shows, that some of the days with the worst revenue are on holidays and these days may need special treatment. The reason for this is that holidays simply do not behave like normal weekdays. Therefore we investigate whether different adjustment on holidays may be beneficial.

Unfortunately the total number of public holidays is not large enough to specifically forecast prices on these days, so we need to help ourselves with a rather sensible alternative solution. In many areas of everyday life holidays and Sundays are treated in a similar way. Hence we conjecture that also the market behaves similarly on holidays and Sundays. Thus we frankly use the Sunday's adjustment coefficients of the current period in which the holiday lies.
Moreover we want to know the best possible case on holidays to be able to correctly judge the merit of specific holiday adjustment. If we look at Table 4.3, we see that compared to the Status Quo there is indeed some additional potential for improvement. Contrary to this, the suggested type of strategic adjustment on holidays does not work out as we assumed.

Adjustment Policy	2014	2015	2016	2017
Status Quo - Best Case on Holidays	93.25	92.44	90.43	91.84
Status Quo	93	92.13	90.17	91.45
Status Quo - Holiday Adjustment	93.01	92.13	90.15	91.41

Table 4.3: Ratings regarding Holiday Adjustment

Perfect foresight on holidays compared to the recent "normal-weekday-adjustment" in 2014, 2015, 2016 and 2017 is 0.25, 0.31, 0.26 and 0.39 percentage points of the best case better, respectively. On the other hand, the Sunday-like adjustment on holidays leads to an improvement in percentage points of the best case of only 0.01 in 2014, in 2015 there is no difference at all, and in 2016 and 2017 it leads to a worsening of -0.02 and -0.04 percentage points, respectively. We see that there is room for significant improvement on holidays. However if we take the amount of holidays each year and compute the proportional improvement by the best case scenario on the same number of average days, we basically get the same results as for "Best Case on Holidays". The latter is only hundredths of percentage points better in 2014 and 2015, it is even worse

a few hundredths of a percentage point in 2016 and only in 2017 it is better in the order of tenths of a percentage point. This means it appears to be that strategic adjustment on holidays is not significantly worse than on any other day.
However we did not investigate this any further, since other approaches were more promising.

Future Recommendations 2. Our attempt of using Sunday's adjustment coefficients on holidays did not show significant benefit. Hence analysis of the market's behaviour on holidays and development of specific adjustment from that knowledge could be the subject of future work.

To settle this topic conclusively, we will test our approach of "Holiday Adjustment" on our own, new heuristic later on. If this shows no improvement as well, we need to reject this idea of strategic adjustment on holidays.

4.2.5 Limitation by the TSO

Another big topic for the company is limitation by the TSO. We recall that limitation can occur in case of imminent collapse of the transition system. This means, that they are authorised to reduce the actual production to the schedule. Hence we produce less than we could and thus get less revenue (remember that $exc > 0$).
The data we got at the beginning does not have a mark or flag to indicate actual (or potential) limitation. However measuring actual production $prod$ pretty close to the scheduled production $final$ is a rather strong sign for limitation. If we filter all these hours with $|prod - final| < M$ for some reasonable $M > 0$ big enough, we get at least all limited hours. In the worst case we also get hours with accidental close values, but since this is a worst-case-analysis that would lead to a worse result at most.

In the Figures 4.3-4.6 we have aggregated all those hours of one year during which limitation might have occurred. This is done for all parks together. Note that, theoretically, the number of hours on each day can be up to $72 = (24 \cdot 3)$ because we look at three parks.
The different colours mark different weekdays, to figure out whether there are specific days on which limitation happens more often than on others. Furthermore we only look at data with non-zero $final$ values submitted because "$final = 0$" is a strong sign of non-availability implying no produced energy as well.
We see in Figure 4.3 that limitation hardly occurred in the first half of 2014, increased

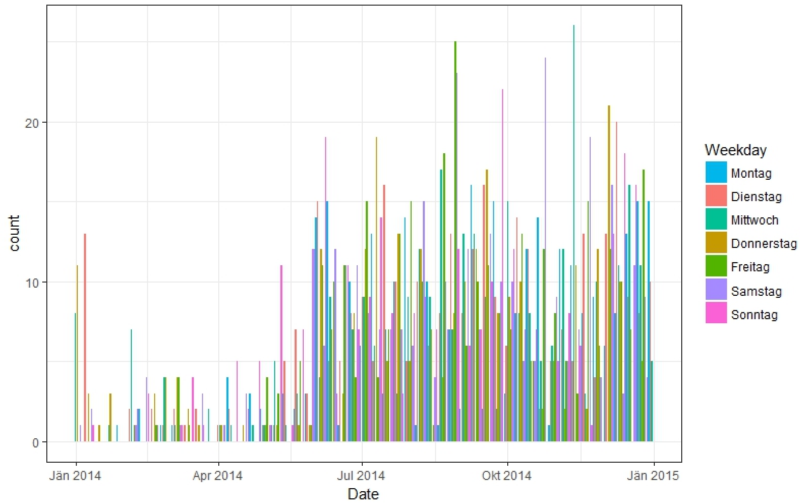

Figure 4.3: Possibly limited hours aggregated for all parks in 2014

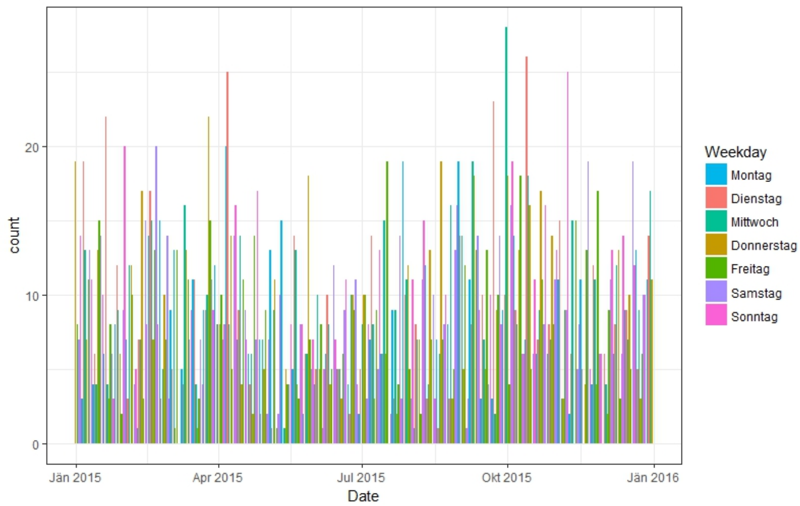

Figure 4.4: Possibly limited hours aggregated for all parks in 2015

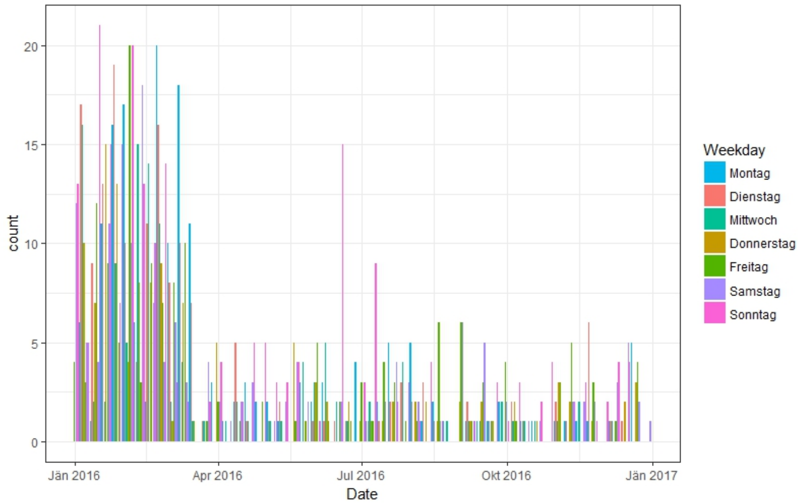

Figure 4.5: Possibly limited hours aggregated for all parks in 2016

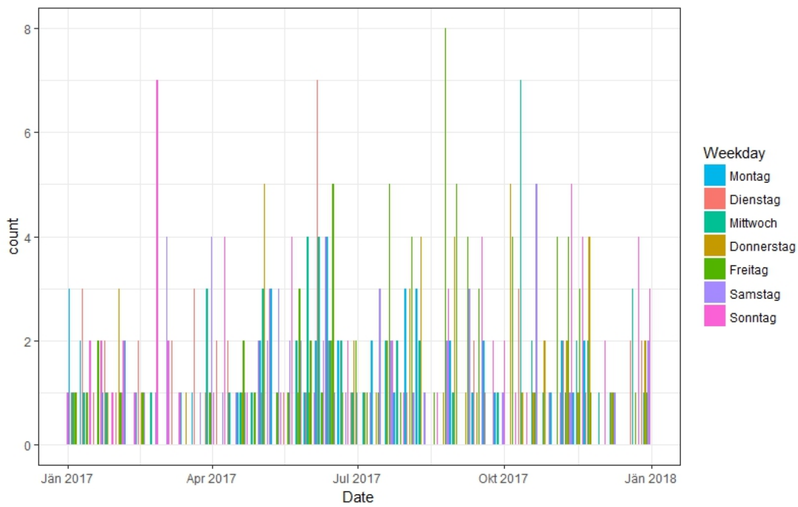

Figure 4.6: Possibly limited hours aggregated for all parks in 2017

in the second half of the same year, remained at the same level until the first quarter of 2016 (see Figure 4.4 and Figure 4.5) and then decreased again to a very low level of limitations. This trend continues even further in 2017, if we look at Figure 4.6. However we do not recognise a specific pattern for limitation on certain weekdays or seasons.

Next we look at the capacity utilisation rates whenever limitation occurs. This rate denotes the proportion of the actual production *prod* relative to the physical upper production limit of the park. E.g., if the physical upper production limit was 100MWh and we had an actual production of 25MWh then the corresponding capacity utilisation rate is 0.25. This entity is of interest for us since limitation at high rates is not as harmful as when the parks could possibly produce much more energy before reaching the natural bound. On the other hand, limitation at low rates does not immediately yield a bad result as the theoretically producible energy could be only slightly higher.

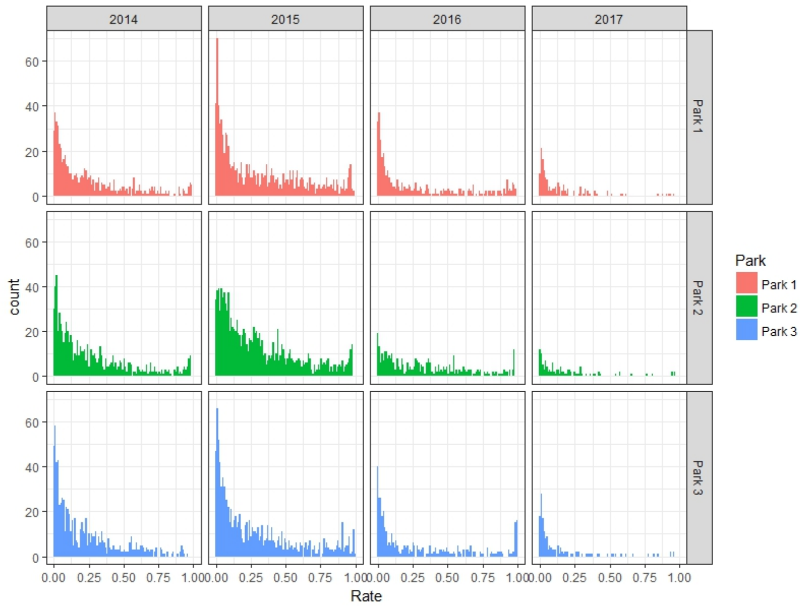

Figure 4.7: Capacity utilisation rates at possible limitation

If we look at Figure 4.7 we notice something rather astounding. In each year for all parks the most limited hours occur in the lower sector. Moreover, there is only a small peak at the high end. This behaviour may be due to successful strategic adjustment to avoid limitation in the upper sector, opposing to too cautious adjustment in the lower one.

4.3 Mathematical Modelling

By now we have finished the purely descriptive part of this work and we proceed with finding some strategic adjustment by ourselves. We can partition our endeavours into two main tasks:

1. How do we get a reasonable forecast of the prices?
2. How do we get reasonable strategic adjustment?

In the following we start with the easiest approaches to answer both of the questions and then continue with more sophisticated ones.

4.3.1 Forecast of Prices

Firstly we need to think about how to forecast the three entities da, exc and def, which is a major difficulty in our task of finding good strategic adjustment policies. Even though def denotes the penalty for deficit energy, both exc and def are the prices for positive and negative balancing energy. Thus if we speak of "prices" we generally mean all three entities, including def in specific.

Historic Average

After gaining insight into the company's status quo, we opted for a very simple historic average as our first means of finding predictors $\widehat{da}$, $\widehat{exc}$ and $\widehat{def}$, i.e., forecasts of the prices da, exc and def. We decided to do this on a monthly basis, hoping we can capture current events better using more recent data as the energy market is a fast-moving domain.
We compute the historic averages in two versions, the second one being a slight variation of the first one.

1. For each hour of the week we compute an average of the current month over the last three years. For example if we want to forecast exc on a Tuesday in March 2016 at 17:00, we compute the average of all values for exc on Tuesdays in March in 2015, 2014 and 2013 at 17:00.
   Strategic adjustment policies using this version of **historic averages** get the label **HA**.

Taking the average over three years was a purely ad-hoc decision at the beginning of the work on this topic. Due to rather acceptable results and our intentions to develop more complex ideas, we did not try to find an "optimal" configuration for the historic

average. For example we did not try out another number of averaged years.
Averaging over one month of the last three years means that we take into account only about 12 values, as for each hour of a week we have approximately four weeks per month times three years only. This is not a very broad foundation for a profound forecast. Moreover single values fluctuate extraordinarily from time to time, leading to completely different averages. So our attempt is on the one hand to smoothen the hourly values and on the other hand to gather more samples from which we calculate the average. The choice so far to compute a set of adjustment coefficients once a quarter of a year was arbitrary, and so is the choice of doing so once a month. The prices do not think in months or years, the behaviour is a continuous development over time. Therefore we considered a second version of historic average.

2. For each hour of the data we compute the average of this hour, the previous hour and the one afterwards, e.g., if we again want to forecast *exc* on a Tuesday in March 2016 at 17:00, we compute the average of the values from 16:00, 17:00 and 18:00.
   Then we take the average of the "rolling quarter" surrounding the looked at month of the last three years for each hour of the week. That means for our example about forecasting *exc* on a Tuesday in March 2016 at 17:00 that we compute the average of the averaged value for 17:00 on Tuesdays in February, March and April each in 2015, 2014 and 2013.
   Strategic adjustment policies using this more **stable** version of historic averages get the label **HAs**.

The historic averages compute an average week for one month, so each weekday has the same forecast of da, exc and def throughout a month. The earliest prices available for us are from 2012, so the first year we can forecast by these historic averages is 2015.

Linear Regression

Our second approach of forecasting da, exc and def is by linear regression. The idea is to fit the prices of the current year with the data from the year before by a linear model. This way we obtain coefficients for each year to predict the consecutive year in the same way.
In the following we will only consider univariate, multiple linear regression. That means we use multiple independent variables (regressors) to fit one dependent variable (regressand). This is done for each of the prices separately.

Future Recommendations 3. However one could consider one multiple multivariate model, i.e., several regressors and more than one regressand, to fit all prices at the same time. This could lead to overall improvement, as the behaviour of one particular price could contain information about the other prices. We recommend that these considerations should be treated in a future work.

We again have two slightly different versions.

1. We do this by using the prices of 53, 52 and 51 weeks (i.e., one year) ago, 27, 26 and 25 weeks (i.e., half a year) ago and 14, 13 and 12 weeks (i.e., one quarter) ago as explanatory variables to describe the current price. This is done for each of the three entities - da, exc and def - separately. The resulting predictors are $\widehat{da}$, $\widehat{exc}$ and $\widehat{def}$ for da, exc and def, respectively.
   Strategic adjustment policies using this version of linear regression, i.e., a **linear model**, get the label **LM**.

We look at one of the prices $p \in \{da, exc, def\}$, denote its value at hour t by p_t. To simplify notation we introduce $w := 24 \cdot 7 = 168$, the number of hours in one week. Then we have the following linear model,

$$\begin{aligned} p_t = \beta_1 p_{t-53w} + \beta_2 p_{t-52w} + \beta_3 p_{t-51w} + \beta_4 p_{t-27w} + \beta_5 p_{t-26w} + \dots \\ \dots + \beta_6 p_{t-25w} + \beta_7 p_{t-14w} + \beta_8 p_{t-13w} + \beta_9 p_{t-12w} + U_t^p, \end{aligned} \tag{4.5}$$

where p_{t-mw} denotes the price with a lag of m weeks and U_t^p is random noise at hour t. Because of the random noise, the price is a random variable as well. Therefore we need to distinguish between the random variable P_t, describing the price at hour t, and its realisation p_t. We use upper case for the first and lower case for the latter.
Some trials at the beginning of consideration of linear models have shown, that better results are achieved with no intercept in our case. Therefore we do not consider an intercept in the following. A possible reason is that there is no secured minimal price greater zero that we would sell the energy at.

We already know that energy economics is a fast-moving domain, so we always try to use the most recent data possible. Using the values of exc and def from a quarter ago is the best we can do, but we have more recent data of da. This motivates the second version.

2. The linear model for exc and def is the same as in LM. However for da we additionally use the value of one week and one day ago as explanatory variables.

Strategic adjustment policies using this version of linear regression incorporating the most **recent** data get the label **LMr**.

Therefore, the linear models for exc and def are the same as in “LM” (see (4.5)). Only da has a different model, which is as follows,

$$\begin{aligned} da_t = \beta_1 da_{t-53w} + \beta_2 da_{t-52w} + \beta_3 da_{t-51w} + \beta_4 da_{t-27w} + \dots \\ \dots + \beta_5 da_{t-26w} + \beta_6 da_{t-25w} + \beta_7 da_{t-14w} + \beta_8 da_{t-13w} + \dots \\ \dots + \beta_9 da_{t-12w} + \beta_{10} da_{t-w} + \beta_{11} da_{t-24} + U_t^{da_r}, \end{aligned} \tag{4.6}$$

We see that “LMr” has two additional regressors compared to “LM”. However overfitting should not be a serious concern. Eleven explanatory variables compared to more than 8000 observations we fit during a year are still considerably few.
We employ the well-established assumptions we encountered in Section 3.1.

1. $\mathbb{E}\left[U_t^p\right] = 0$ for all hours t.
2. $\operatorname{Var}\left[U_t^p\right] = \sigma^2$ for all hours t.
3. $\mathbb{E}\left[U_t^p U_s^p\right] = 0$ for all hours $t \neq s$.

This implies that by solving the ordinary least squares problem we obtain unbiased estimators $\widehat{\beta}_j$ for β_j.

Even though the observed prices are stochastic themselves, we assume for simplicity they are fixed for the regression. Corresponding statements are also true, if you condition on the certain realisation so this would not lead to drastic modifications.
Furthermore we assume that we are able to “catch” the entire deterministic contribution by the prediction of our linear model, only perturbed by random noise ε_t^p with $\mathbb{E}\left[\varepsilon_t^p\right] = 0$. That means we have

$$P_t = \widehat{p}_t + \varepsilon_t^p \quad \text{and thus} \quad \mathbb{E}\left[P_t\right] = \widehat{p}_t. \tag{4.7}$$

The linear models compute forecasts $\widehat{da}$, $\widehat{exc}$ and $\widehat{def}$ for da, exc and def, respectively, separately for each hour. Thus same weekdays throughout a month do not have the same values - contrary to the historic averages above. Since price data is available to us as far back as the beginning of 2012, the first year we can fit is 2013 which means the first year we can forecast prices by these linear models is 2014.

In the following we look at the results of fitting the prices from 2016 to obtain coefficients to predict the prices for 2017, which is the most current year we consider. Therefore we look at the results of fitting the prices in 2016. The tables are generated by the *R* command "`summary`", which provides the most important key figures to judge the quality of a regression.
First we look at the linear model for *da*.

```
> summary(fit_DA_2016)

Call:
lm(formula = DA ~ 0 + DA_12 + DA_13 + DA_14 + DA_25 + DA_26 +
     DA_27 + DA_51 + DA_52 + DA_53,
    data = lm_DA %>% filter(year(UTC) == 2016))

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
DA_12   0.106122   0.012380   8.572  < 2e-16 ***
DA_13   0.094242   0.012802   7.362 1.98e-13 ***
DA_14   0.101004   0.012267   8.234  < 2e-16 ***
DA_25   0.010606   0.011562   0.917  0.35899
DA_26   0.031942   0.011758   2.717  0.00661 **
DA_27  -0.078776   0.011522  -6.837 8.64e-12 ***
DA_51   0.211231   0.009941  21.248  < 2e-16 ***
DA_52   0.192522   0.010271  18.744  < 2e-16 ***
DA_53   0.269370   0.010039  26.833  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.05 on 8775 degrees of freedom
Multiple R-squared:  0.9226,  Adjusted R-squared:  0.9225
F-statistic: 1.162e+04 on 9 and 8775 DF,  p-value: < 2.2e-16
```

Listing 1: Summary of linear model fitting *da* in 2016

We mainly look at two things - statistical significance on the one hand and the (adjusted) coefficient of determination R^2 (and $\bar{R}^2$, respectively) on the other hand. We see in Listing 1 that all regressors (except the lag of 25 weeks) are highly significant, only da_{t-26w} slightly less. High significance levels are indicated by a maximum of three asterisks in the rightmost column of the section "Coefficients".
Furthermore, we have $R^2 = 0.9226$ and $\bar{R}^2 = 0.9225$ which are amazingly high values for a regression. This suggests that employing a linear model is justified.

Next we want to investigate the impact of adding two new regressors in the linear model for *da* - we look at “LMr”.

```
> summary(fit_DA_recent_2016)

Call:
lm(formula = DA_recent ~ 0 + DA_1D + DA_1 + DA_12 + DA_13 + DA_14 +
    DA_25 + DA_26 + DA_27 + DA_51 + DA_52 + DA_53,
   data = lm_DA_recent %>% filter(year(UTC) == 2016))

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
DA_1D  0.373205   0.008480  44.010  < 2e-16 ***
DA_1   0.230095   0.009569  24.046  < 2e-16 ***
DA_12  0.043751   0.010439   4.191 2.81e-05 ***
DA_13  0.027892   0.010796   2.584  0.00979 **
DA_14  0.034050   0.010352   3.289  0.00101 **
DA_25  0.010982   0.009697   1.133  0.25745
DA_26  0.031045   0.009867   3.146  0.00166 **
DA_27 -0.059026   0.009680  -6.098 1.12e-09 ***
DA_51  0.084743   0.008639   9.809  < 2e-16 ***
DA_52  0.088833   0.008794  10.102  < 2e-16 ***
DA_53  0.113951   0.008810  12.935  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.429 on 8773 degrees of freedom
Multiple R-squared:  0.9456,  Adjusted R-squared:  0.9455
F-statistic: 1.385e+04 on 11 and 8773 DF,  p-value: < 2.2e-16
```

Listing 2: Summary of linear model fitting *da* with more recent data in 2016

In Listing 2 we see, that the price data from the previous day and one week ago are highly significant. However including the lag of one week and one day also yields that da_{t-13w} and da_{t-14w} are in a lower significance level as before. The situation for the rest of the explanatory variables is the same, though.
As mentioned in Section 3.1, more regressors cannot lead to a lower coefficient of determination, which is $R^2 = 0.9456$ in this case. But a resulting $\bar{R}^2 = 0.9455$, which takes into account the number of independent variables, is pleasing nonetheless.

Next we examine the results for *exc*.

```
> summary(fit_EXC_2016)

Call:
```

```
lm(formula = EXC ~ 0 + EXC_12 + EXC_13 + EXC_14 + EXC_25 + EXC_26 +
    EXC_27 + EXC_51 + EXC_52 + EXC_53,
    data = lm_EXC %>% filter(year(UTC) == 2016))

Coefficients:
       Estimate Std. Error t value Pr(>|t|)
EXC_12  0.16828    0.01122  14.996  < 2e-16 ***
EXC_13  0.10217    0.01142   8.947  < 2e-16 ***
EXC_14  0.13495    0.01140  11.841  < 2e-16 ***
EXC_25  0.08323    0.01120   7.433 1.16e-13 ***
EXC_26  0.04949    0.01126   4.397 1.11e-05 ***
EXC_27  0.07440    0.01113   6.685 2.45e-11 ***
EXC_51  0.18431    0.01201  15.350  < 2e-16 ***
EXC_52  0.16481    0.01216  13.558  < 2e-16 ***
EXC_53  0.13163    0.01215  10.833  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.34 on 8775 degrees of freedom
Multiple R-squared:  0.7679,  Adjusted R-squared:  0.7677
F-statistic:  3226 on 9 and 8775 DF,  p-value: < 2.2e-16
```

Listing 3: Summary of linear model fitting *exc* in 2016

We see the results are a bit different compared to *da*. For example, *all* of the regressors are highly significant. On the other hand, the coefficients of determination are substantially lower than for both versions of *da*.
We have $R^2 = 0.7679$ and $\bar{R}^2 = 0.7677$, which are still reasonably good after all.

Finally, we look at the performance of the linear model fitting def in 2016.

```
> summary(fit_DEF_2016)

Call:
lm(formula = DEF ~ 0 + DEF_12 + DEF_13 + DEF_14 + DEF_25 + DEF_26 +
    DEF_27 + DEF_51 + DEF_52 + DEF_53,
    data = lm_DEF %>% filter(year(UTC) == 2016))

Coefficients:
       Estimate Std. Error t value Pr(>|t|)
DEF_12  0.16791    0.01221  13.748  < 2e-16 ***
DEF_13  0.10372    0.01260   8.231  < 2e-16 ***
DEF_14  0.03127    0.01251   2.499   0.0125 *
DEF_25  0.02940    0.01207   2.436   0.0149 *
DEF_26  0.09404    0.01224   7.684  1.7e-14 ***
```

```
DEF_27  -0.01627     0.01207   -1.348    0.1778
DEF_51   0.17611     0.01266   13.908   < 2e-16 ***
DEF_52   0.19908     0.01281   15.540   < 2e-16 ***
DEF_53   0.24012     0.01282   18.725   < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.09 on 8775 degrees of freedom
Multiple R-squared:  0.913,  Adjusted R-squared:  0.913
F-statistic: 1.024e+04 on 9 and 8775 DF,  p-value: < 2.2e-16
```

Listing 4: Summary of linear model fitting def in 2016

The results for def are overall pretty similar to those for da. Again, we have that one of the regressors from half a year ago is not significant at all - in this case it is p_{t-27w}. Additionally, the lags of 14 and 25 weeks have rather low significance levels. Furthermore we have $R^2 = 0.913$ and $\bar{R}^2 = 0.913$, which have the same immensely high level as for da.

To summarise the results of Listing 1-4, we may conclude that utilising linear models are justified. Furthermore, the regressors seem to be well-chosen in general. Only when taking the prices from half a year ago, we maybe should leave out at least one of the lags of 25, 26 or 27 weeks as for all prices except exc there is one of them with the lowest significance level.

As mentioned above, after fitting the linear model we obtain a vector $\widehat{\beta}$ of coefficients. Then we use these weights to predict the following year's prices with the historic data. We assumed that the prediction by our linear model determines the entire deterministic part of the prices. The next step is to investigate whether this assumption is consistent with the results or not. Therefore we analyse the resulting residuals $\varepsilon_t^p := p_t - \widehat{p_t}$ of the predictions in 2017.

We see in Figure 4.8 that the residuals are approximately centred at 0. For this density plot all residuals from 2014 to 2017 were aggregated. Even though there is a small shift to the right, i.e., our prediction more often underestimates the real occurrences, our assumption does not seem to be far-fetched. Hence, we stick to it and assume $\mathbb{E}[P_t] = \widehat{p_t}$.

The next step is to look at the residuals over time, to see whether there are seasonal dependencies, possibly contradicting the assumption about independence of the errors. Furthermore, we need to check whether the variance of the random noise term is indeed

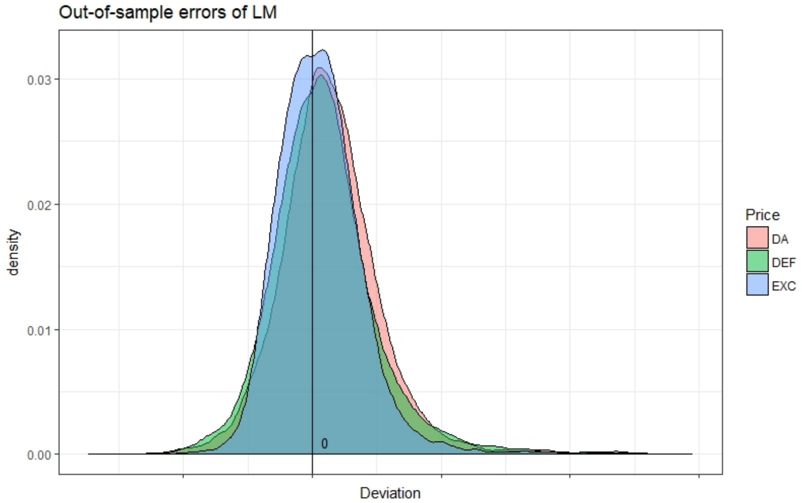

Figure 4.8: Errors made when using LM to find predictors for DA, EXC and DEF

constant throughout all observations.

In Figures 4.9-4.12 we have plotted the residuals for the different models over time. We can clearly see, that our assumption of constant variation is likely to be not entirely correct. For each of the models we have seasonal dependencies, with a bigger/smaller spread or systematically over-/underestimating reality. However, for *da* using the most recent data (Figure 4.10) and for *exc* (Figure 4.11) these unpleasant characteristics are not as strong as for the other models.

A general problem with estimating time-series is, that consecutive points in time have similar characteristics and influence each other most likely. The correct model to use would be an "autoregressive model" (AR model). This type of model describes situations, where the dependent variable depends on its own previous values and an additional stochastic term.

Future Recommendations 4. Future work could focus more on the aspect of time-series and how to cope with the special aspects one has to take into account to treat this properly. For example one could utilise models over shorter periods than a whole year or introduction of dummy variables (variables to incorporate categorical regressors in a model) for distinct seasons could provide remedy. On the other hand, using an AR model as mentioned above seems to be promising, as this type of models is widely used to describe time-series.

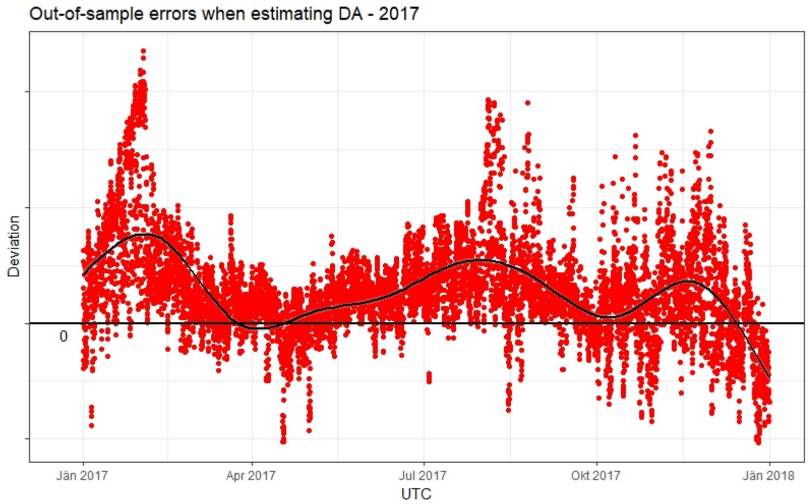

Figure 4.9: Errors made when using LM to predict DA in 2017

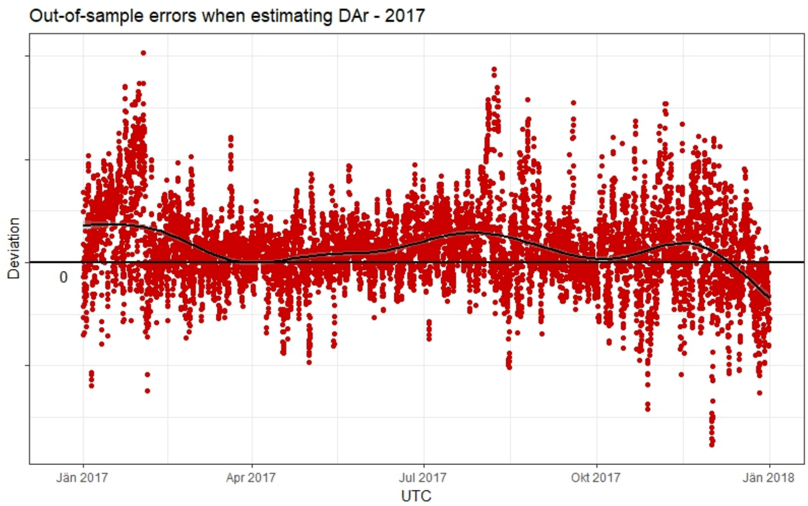

Figure 4.10: Errors made when using LMr to predict DA in 2017

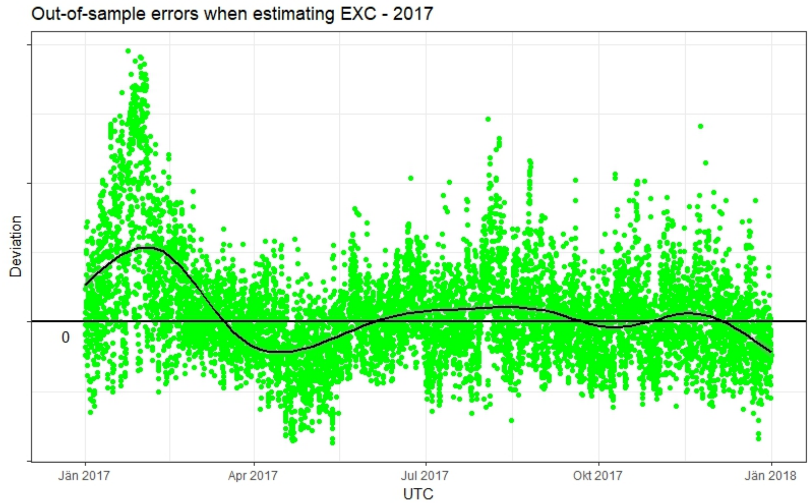

Figure 4.11: Errors made when using LM to predict EXC in 2017

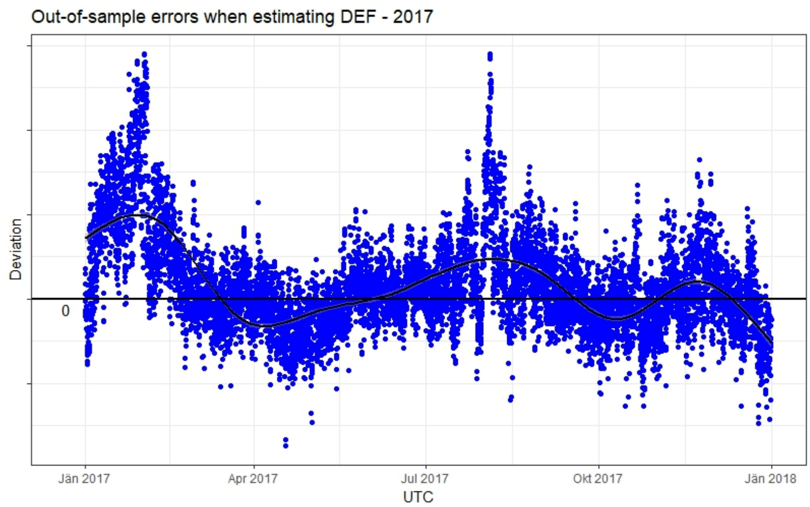

Figure 4.12: Errors made when using LM to predict DEF in 2017

Actual Prices

We have seen already that one of the major problems is that the balancing energy prices exc and def become available only after two to three months. Contrarily, we know the day-ahead price da shortly after. So a naive approach is to use the actual values for da and only forecast exc and def in the ways described above.
It makes sense to use the most recent data available to capture the current behaviour of the market. At first it was not clear whether we can employ the very same day's values, which unfortunately is not possible since the prices are published after submission of the production values. Other sensible attempts are using the actual prices of the exact day before or one week before.
Strategic adjustment policies using actual **day-ahead** prices from the very same day (which is impossible in practice), i.e., from **0** days before, from the previous day, i.e., from **1** day before, and from the week before, i.e., **7** days before, get the additional label **DA0**, **DA1** and **DA7**, respectively.

4.3.2 Strategic Adjustment by Heuristics

In this part we develop a heuristic that mimics strategic adjustment like we considered in Section 4.1.2. Let us recall,

1. If $a > b$, we try to avoid overproduction and therefore increase the $final$ value before submission.
2. If $a < b$, we try to avoid underproduction and therefore decrease the $final$ value before submission.

Under the assumption that the market is in balance, we have $a \geq 0$ and $b \geq 0$. Consider the following expression,

$$\lambda = \frac{a - b}{a + b}, \quad \text{with } a \neq -b. \tag{4.8}$$

We see that (4.8) produces adjustment coefficients exactly the way we want. We check the most extreme cases:

1. If $a > b = 0$, then $\lambda = \frac{a}{a} = 1$ and we increase $final$ with an adjustment coefficient of 1.
2. If $0 = a < b$, then $\lambda = -\frac{b}{b} = -1$ and we decrease $final$ with an adjustment coefficient of -1.

This means that under the expected condition that $exc < da < def$ the above heuristic produces adjustment coefficients in the range of $-1 \leq \lambda \leq 1$. This however does not imply that we adjust by ±100% because of the company's adjustment heuristic involving some domain-dependent weighting coefficient.
In reality it also happens quite often that either $a < 0$ or $b < 0$, though, which leads to adjustment coefficients with absolute value significantly larger than 1. However we force the resulting λ to be of absolute value less or equal to one, by simply capping at ±1. This means that we calculate the final adjustment coefficient $\overline{\lambda}$ as

$$\overline{\lambda} = \begin{cases} \min(\lambda, 1), & \text{if } \lambda \geq 0 \\ \max(\lambda, -1), & \text{if } \lambda < 0 \end{cases}$$

Since submission of awkwardly large or small $final$ values will be prevented by the TSO, extreme adjustment with $|\lambda| \cong 1$ may not be eligible. Thus another capping value of ±0.3 is considered as well.
Strategic adjustment policies using the above **heuristic** to derive adjustment coefficients capped at ±**1.0** and ±**0.3** get the label **HEU(1.0)** and **HEU(0.3)**, respectively.

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
HEU(1.0) HA	NA	93.33	91.33	92.07
HEU(1.0) HA DA0	NA	92.95	88.91	94.81
HEU(1.0) HA DA1	NA	90.82	87.81	91.63
HEU(1.0) HA DA7	NA	90.95	88.99	92.16
HEU(1.0) HAs	NA	93.31	91.52	92.71
HEU(1.0) HAs DA0	NA	93.1	88.93	94.53
HEU(1.0) HAs DA1	NA	90.94	87.9	91.67
HEU(1.0) HAs DA7	NA	91.19	89.31	92.25

Table 4.4: Ratings of HEU(1.0) with Historic Average

In Table 4.4 we have the resulting ratings for the heuristic capped at values of ±1 using historic averages to forecast the three entities da, exc and def. Recall that the ratings are the percentage of the policy's revenue compared to the best case's revenue. We see that the two policies "HEU(1.0) HA" and "HEU(1.0) HAs" already deliver better results than the status quo, however they may not be applicable in reality due to rather extreme adjustments.
Moreover we see that if we are using actual day-ahead prices instead of forecast ones,

the actual prices of the current day ("DA0") would lead to the best results. For both "HA" and "HAs" using "DA0" is at least one percentage point better than "DA1" or "DA7", which is pretty much. Furthermore we notice that "DA7" leads to better results than "DA1", so apparently the current weekday is more important than the recent behaviour of the last day. If we compare the ratings for "HA" and "HAs" we see that the latter generally seems to be more stable as we intended, since we are taking into account much more data. Concluding, our idea which motivated the second version of the historic average appears to be sensible.

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
HEU(1.0) LM	87.33	92.32	91.28	81.42
HEU(1.0) LM DA0	95.79	95.59	87.54	90.56
HEU(1.0) LM DA1	92.49	91.45	86.83	89.33
HEU(1.0) LM DA7	92.7	91.09	88.46	89.67
HEU(1.0) LMr	94.56	92.89	90.42	89.54

Table 4.5: Ratings of HEU(1.0) with Linear Model

In Table 4.5 we have the resulting ratings for the heuristic capped at values of ± 1 using linear models to forecast the prices. First of all we notice that in 2017 all policies using linear models are worse than the status quo. However in 2014 and 2015 "HEU(1.0) LM DA0" is astoundingly superior compared to the currently used adjustment policy. Also "HEU(1.0) LMr" is better in 2014, 2015 and 2016 - so a forecast using linear models seems to be justifiable, even though we might need to change the specific application a bit. In contrary to the heuristic using historic averages, both "DA1" and "DA7" are worse than the status quo throughout all the years.
Note that "LMr DA0", "LMr DA1" and "LMr DA7" are not missing, as "LMr" only differs from "LM" in the way we forecast da which is exchanged if we use the actual day-ahead prices ("DA").

If we look at Table 4.6 we have the resulting ratings for the heuristic capped at values of ± 0.3 using historic averages. Surprisingly the results for "HEU(0.3)" are notably similar to those for "HEU(1.0)". The policies "HEU(0.3) HA" and "HEU(0.3) HAs" produce better ratings than the status quo and should be applicable in reality, too. We also notice that the impact of using the actual da values decreases when capping at ± 0.3 - in both positive and negative direction. Moreover, the difference between using

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
HEU(0.3) HA	NA	93.07	91.87	92.37
HEU(0.3) HA DA0	NA	92.85	90.91	93.48
HEU(0.3) HA DA1	NA	91.87	90.6	92.64
HEU(0.3) HA DA7	NA	91.88	91.05	92.77
HEU(0.3) HAs	NA	93.08	91.83	92.72
HEU(0.3) HAs DA0	NA	92.92	90.9	93.4
HEU(0.3) HAs DA1	NA	91.92	90.61	92.63
HEU(0.3) HAs DA7	NA	91.97	91.13	92.75

Table 4.6: Ratings of HEU(0.3) with Historic Average

"HA" and "HAs" gets smaller as well.
However "HEU(0.3) HA" and "HEU(0.3) HAs" appear to be viable alternatives to the currently used adjustment policy, leading to more revenue in the investigated years.

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
HEU(0.3) LM	91.29	92.63	91.65	87.69
HEU(0.3) LM DA0	93.7	93.47	90.49	91.7
HEU(0.3) LM DA1	92.75	92.07	90.4	91.47
HEU(0.3) LM DA7	93.02	91.94	90.86	91.57
HEU(0.3) LMr	93.48	92.53	91.63	91.29

Table 4.7: Ratings of HEU(0.3) with Linear Model

In Table 4.7 we have the resulting ratings for the heuristic capped at values of ± 0.3 using linear models. As we saw for the historic averages, capping at ± 0.3 leads to less extreme effects. Apparently it manages to improve the worse results on the one hand, but on the other hand it also makes the outstanding good results from before to be not as good anymore. However we do not obtain a strictly better adjustment policy which would be possible in reality.

Strategic Adjustment on Holidays

When we were analysing our attempt for special strategic adjustment on public holidays, we were not able to get consistent significantly better results than normal adjustment ignoring holidays. Recall, our approach of "Holiday Adjustment" is to use the adjustment coefficients for Sundays on holidays. We now want to quickly check whether the results are different for our heuristic.

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
Status Quo - Holiday Adjustment	93.01	92.13	90.15	91.41
HEU(1.0) HA	NA	93.33	91.33	92.07
HEU(1.0) HA - Holiday Adjustment	NA	93.38	91.2	91.89
HEU(0.3) HA	NA	93.07	91.87	92.37
HEU(0.3) HA - Holiday Adjustment	NA	93.07	91.83	92.3

Table 4.8: Ratings regarding Holiday Adjustment for Heuristic

If we look at Table 4.8 we see that the results are basically the same. Even though our attempt of holiday adjustment in 2015 is slightly better for "HEU(1.0) HA" and at least not worse for "HEU(0.3) HA", the rest of the years the results are significantly worse. This strengthens our presumption that (at least this kind of) special holiday adjustment does not lead to reasonable improvements. Thus we will not treat holidays differently from any other days and continue with our search for generally better adjustment policies instead.

4.3.3 Strategic Adjustment by Stochastic Programming

In the following we want to turn away from heuristics and focus on purely mathematically justifiable approaches. So far we tried to find adjustment coefficients to obtain higher revenue. This already yields considerably good results, however the underlying plan is to maximise the revenue. Actually, we do not need adjustment coefficients for this. Instead we want to find the optimal $final$, which maximises the revenue, in a more analytic approach.

In the following DA, EXC and DEF denote the stochastic entities describing the prices with their actual realisations da, exc and def, respectively, while $PROD$ denotes the actual production with realisation $prod$. To distinguish between the random variable and its realisation we use upper case for the first and lower case for the latter. We assume the price $g \geq 0$ for one GC to be fixed.

Due to the stochastic nature of prices and production, in a risk-neutral environment we seek to maximise the expected revenue. We get

$$\max_{final} \; \mathbb{E}\left[\texttt{rev}(final)\right] = final \cdot (\mathbb{E}\left[DA\right] + g) + \mathbb{E}\left[\left[PROD - final\right]^{+} \cdot EXC\right] \dots \\ \dots - \mathbb{E}\left[\left[final - PROD\right]^{+} \cdot (DEF + g)\right], \tag{4.9}$$

subject to the constraint that $final$ must be non-negative and less or equal the physical upper production limit for the certain wind park.

In the following we assume that all the stochastic entities can be estimated deterministically except some additive random noise with zero mean. That means we have

$$DA = \widehat{da} + \varepsilon^{da}, \quad EXC = \widehat{exc} + \varepsilon^{exc}, \quad DEF = \widehat{def} + \varepsilon^{def}$$

and

$$PROD = meta + \varepsilon^{meta},$$

with $\widehat{da}$, $\widehat{exc}$, $\widehat{def}$ and $meta$ being the (deterministic) predictions for DA, EXC, DEF and $PROD$, respectively. Furthermore,

$$\mathbb{E}[\varepsilon^{da}] = \mathbb{E}[\varepsilon^{exc}] = \mathbb{E}[\varepsilon^{def}] = \mathbb{E}[\varepsilon^{meta}] = 0,$$

and thus

$$\mathbb{E}\left[DA\right] = \widehat{da}, \quad \mathbb{E}\left[EXC\right] = \widehat{exc}, \quad \mathbb{E}\left[DEF\right] = \widehat{def} \quad \text{and} \quad \mathbb{E}\left[PROD\right] = meta.$$

Expected Value Solution

A first and simple way to handle the stochastic program (4.9) is to substitute each random variable by its expected value. This widely used approach gives the so-called "expected value solution", as established in [9].
Assume that we have

$$\mathbb{E}\left[DA\right] = \widehat{da}, \quad \mathbb{E}\left[EXC\right] = \widehat{exc}, \quad \mathbb{E}\left[DEF\right] = \widehat{def} \quad \text{and} \quad \mathbb{E}\left[PROD\right] = meta.$$

Then we get the following optimisation problem

$$\begin{aligned}
&\max_{final}\ final \cdot (\mathbb{E}[DA] + g) + [\mathbb{E}[PROD] - final]^+ \cdot \mathbb{E}[EXC] \dots \\
&\qquad\qquad \dots - [final - \mathbb{E}[PROD]]^+ \cdot (\mathbb{E}[DEF] + g) \\
=&\max_{final}\ final \cdot (\widehat{da} + g) + [meta - final]^+ \cdot \widehat{exc} - [final - meta]^+ \cdot (\widehat{def} + g),
\end{aligned} \tag{4.10}$$

still subject to the constraint that $final$ must be non-negative and less or equal the physical upper production limit for the certain wind park. The revenue is determined for each hour and each wind park separately, so the above is just a one-dimensional bound-constrained optimisation problem of the following form

$$\begin{aligned}
&\max_{x \in \mathbb{R}} && \alpha x + \beta[\tau - x]^+ - \gamma[x - \tau]^+, \\
&\text{subject to} && 0 \le x \le L
\end{aligned} \tag{4.11}$$

for some fixed $\alpha, \beta, \gamma, L > 0$ and $\tau \ge 0$. We see that the objective function is continuous, but not differentiable in x due to the positive parts. Moreover, the objective is not concave in general either, since $\beta[\tau - x]^+$ is convex for $\beta > 0$.
However the optimisation problem is one-dimensional and of simple form, so we can solve it numerically with any solver for continuous one-dimensional functions.

Now that we know the kind of optimisation problem we have to solve, we only need to find good estimators for the uncertain entities DA, EXC, DEF and $PROD$. For this we forecast the prices by the linear models "LM" and "LMr" as above and suppose the resulting predictions $\widehat{da}$, $\widehat{exc}$ and $\widehat{def}$ are the expected values of the random variables DA, EXC and DEF, respectively. Developing an individual production forecast would require a lot of technical knowledge, e.g., about the turbines and their power curves, weather data (which we do not have) and much more. Hence generating our own one is completely out of scope. Moreover, the company already has a reliable production forecast, so we simply consider this black-box given $meta$ as an estimator for $PROD$. Furthermore we assume that $meta$ is the expected value for $PROD$. We already saw in Section 4.3.1 that the price forecast by a linear model is reasonable.

Additionally, we see in Figure 4.13 that $meta$ is approximately unbiased and the error $\varepsilon^{meta} := prod - meta$, where $prod$ is the actually happened production, has reasonable small standard deviation. This justifies our above plan to use it as an estimator.

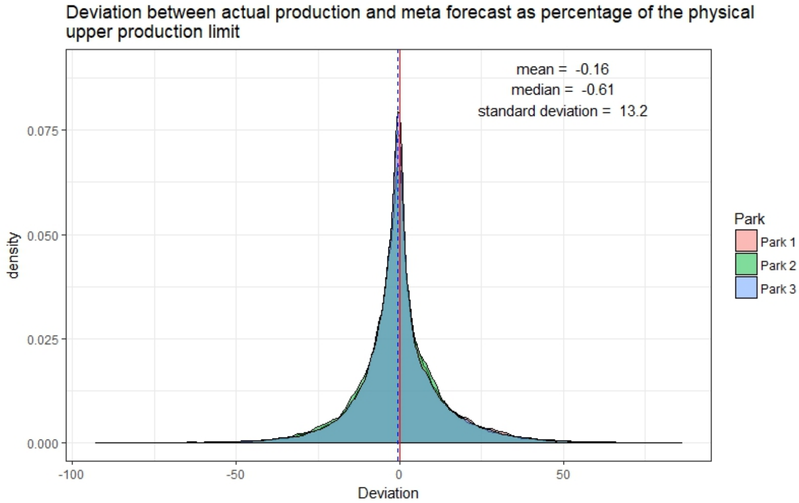

Figure 4.13: Deviation between actual production and *meta* forecast

Strategic adjustment policies solving the above **optimisation** problem (4.10), i.e., computing the **expected value solution**, to derive optimal *final* get the label **OPT(EVS)**.

If we use the actual prices and production instead of the forecasts, we receive the absolute best possible case when solving the optimisation problem. However this would mean we have perfect knowledge of the future - both regarding production and the (balancing) energy prices. This would imply we know the production of all other participants and how the energy prices are determined with this information, which is completely impossible.
Nevertheless the "Absolute Best Case" is the upper bound of what we can earn maximally in theory.

For solving (4.10), we used the standard solver for continuous functions by R, which is a combination of "Golden Section Search" and "Successive Parabolic Interpolation". Golden Section Search is an algorithm that gradually reduces the interval locating the extremum. This is done by generating a sequence of triplets of points which keep the same proportion of spacing relative to each other. The proportion is taken to be the "Golden Ratio" $\varphi = \frac{1+\sqrt{5}}{2}$, which gives the name of the algorithm. We need three points $(x_{i_1} \leq x_{i_2} \leq x_{i_3})$ to determine the interval $[x_{i_1}, x_{i_3}]$ and in which of the two subintervals $[x_{i_1}, x_{i_2}]$ and $[x_{i_2}, x_{i_3}]$ the extremum must be located. For this we compare

the function values at these points. By maintaining the same proportion of spacing throughout the algorithm, we avoid a situation in which x_{i_2} is very close to x_{i_1} or x_{i_3} and guarantee that the interval width shrinks by the same constant proportion in each step.
Successive Parabolic Interpolation is a technique for finding the extremum by fitting parabolas to a function of one variable at three unique points. The algorithm has super-linear convergence, but convergence is not guaranteed if this method is used alone.
However if we alternate parabolic iterations with Golden Section Search to choose iteration points, we have a much better chance to obtain convergence without hampering the rate of convergence significantly.

As we see in Table 4.9 the results for solving the above optimisation problem are not as good as the currently used adjustment policy in general. We see that if we knew prices and production exactly, this would gain about additional 6-15 percentage points compared to the best case ("Absolute Best Case"). Furthermore we see that using the actual day-ahead prices is disadvantageous for "OPT(EVS)". Moreover, neither "OPT(EVS) LM" nor "OPT(EVS) LMr" are even close to the status quo adjustment policy in general. Thus we will ignore actual day-ahead prices in the following. Only in 2014 "OPT(EVS) LMr" is better - in fact even 2.7 percentage points of the best case.

Adjustment Policy	2014	2015	2016	2017
Absolute Best Case	115.62	112.96	106.82	111.81
Status Quo	93	92.13	90.17	91.45
OPT(EVS) LM	92.91	90.65	89.47	91.13
OPT(EVS) LMr	95.69	90.29	86.66	89.4
OPT(EVS) LM DA1	91.46	88.12	83.09	85.63
OPT(EVS) LM DA7	90.3	87.85	85.57	88.02

Table 4.9: Ratings of OPT(EVS)

When looking more closely at the results of the optimisation problem, we saw that in most hours of worst revenue $final$ is equal to the physical upper production limit. This is because our optimisation problem assumes that our forecasts actually happen in reality exactly as we predicted. For example, whenever our forecast predicts $def < da$ then $final$ will always be as big as possible. Even though this scenario happens quite often in reality, many times we predict the wrong hours which leads to significantly less

revenue. The difficulty of solving the optimisation problem is that we do not adjust relative to the forecast of production, but we determine absolute figures for $final$. We see that this approach is very sensitive regarding the forecast prices.
On the one hand we need to make sure that $final$ will be in the range of the originally forecast $meta$, which will be done by a regularisation term in the objective function. On the other hand the next step will be to incorporate uncertainty of the forecasts in the optimisation problem - to get away from "black or white" decisions, which will be done by an "Sample Average Approximation" approach.

Regularisation

In the following we want to use regularisation to obtain $final$ reasonable close to $meta$. This is crucial to avoid the physical upper production limit as $final$ due to mismatch in the prices, which would lead to severe losses with respect to the revenue. This is done by introduction of a regularisation term which inflicts penalty on the revenue whenever $|meta - final|$ is too big.

We get the regularised optimisation problem

$$\begin{aligned} \max_{x \in \mathbb{R}} \quad & \alpha x + \beta[\tau - x]^+ - \gamma[x - \tau]^+ - \kappa|\tau - x|^2 \\ \text{subject to} \quad & 0 \leq x \leq L \end{aligned} \tag{4.12}$$

again for some fixed $\alpha, \beta, \gamma, L > 0$, $\tau \geq 0$ and for some additional regularisation constant $\kappa \geq 0$. To find a suitable choice for κ we simply try different orders of magnitude and see how they influence the results.

If we look at Table 4.10 we see that the approach of regularisation is highly beneficial. Comparing the results for different choices of κ, we recognise that $\kappa = 0.5$ and $\kappa = 1$ lead to best overall ratings. Even though they do not produce the very best results, they manage to maintain a good level throughout all the years. Furthermore we see that "LMr" gives better ratings than "LM" generally speaking.
Also we recognise, that the policy "OPT(EVS) LMr (Kappa = 0.5)" is basically as good as "Status Quo". From 2014 to 2016 it has a higher rating, only in 2017 it is slightly worse by 0.04 percentage points of the best case's revenue.

Sample Average Approximation

As mentioned above, the next (and final) step is to include uncertainty in the opti-

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
OPT(EVS) LM (Kappa = 0)	92.91	90.65	89.47	91.13
OPT(EVS) LMr (Kappa = 0)	95.69	90.29	86.66	89.4
OPT(EVS) LM (Kappa = 0.1)	92.05	92.26	91.1	90.56
OPT(EVS) LMr (Kappa = 0.1)	95.17	92.52	89.73	90.13
OPT(EVS) LM (Kappa = 0.5)	91.28	92.12	90.35	90.47
OPT(EVS) LMr (Kappa = 0.5)	93.01	92.35	90.71	91.41
OPT(EVS) LM (Kappa = 1)	91.16	91.75	90.11	90.46
OPT(EVS) LMr (Kappa = 1)	92.32	91.96	90.46	91.5
OPT(EVS) LM (Kappa = 5)	91.05	91.2	89.88	90.45
OPT(EVS) LMr (Kappa = 5)	91.33	91.31	89.98	90.99
OPT(EVS) LM (Kappa = 10)	91.03	91.1	89.85	90.45
OPT(EVS) LMr (Kappa = 10)	91.18	91.16	89.9	90.75
OPT(EVS) LM (Kappa = 50)	91.02	91	89.82	90.45
OPT(EVS) LMr (Kappa = 50)	91.05	91.02	89.83	90.51

Table 4.10: Ratings of OPT(EVS) regularised

misation problem. We do this by utilising a “Sample Average Approximation” (SAA) approach. We recognise that our stochastic program (4.9) is a two-stage problem as treated in Section 3.2. Recall, that in this type of stochastic program we need to solve

$$\min_{x\in X} \ \{f(x) := \mathbb{E}\left[F(x,\xi)\right]\},$$

where $F(x,\xi)$ is the optimal value of the second-stage problem. For an SAA approach we now replace the expected value function $f(x) = \mathbb{E}\left[F(x,\xi)\right]$ by the sample average function $\hat{f}_N(x) = \frac{1}{N}\sum_{i=1}^{N} F(x,\xi^t)$, where $\xi^1,\dots,\xi^N$ are N samples of the random vector ξ and $N \geq 1$.

To actually compute these sums, we employ the historical errors to draw our random samples from. Therefore we compute the (out-of-sample) error between the actually historically realised value and the forecast for each hour t. Now we explicitly write the subscript for the corresponding hour t to be able to distinguish different samples.
To derive the errors of the estimators compared to the actual values we have for all hours t:

$$\varepsilon_t^{da} := da_t - \widehat{da}_t, \quad \varepsilon_t^{exc} := exc_t - \widehat{exc}_t, \quad \varepsilon_t^{def} := def_t - \widehat{def}_t$$

and

$$\varepsilon_t^{meta} := prod_t - meta_t.$$

In the above definitions da_t, exc_t and def_t denote the actual historic realisations of DA, EXC and DEF, respectively, at hour t. Additionally we have $\widehat{da}_t$, $\widehat{exc}_t$ and $\widehat{def}_t$ standing for the estimators of DA, EXC and DEF, respectively. Recall, $prod_t$ denotes the actual historic production, while $meta_t$ stands for the estimator of the same.

We get the following optimisation problem to obtain an optimum for hour s,

$$\max_{final} \; final \cdot (\widehat{da}_s + g) + \frac{1}{N} \sum_{i=1}^{N} \left[(meta_s + \varepsilon_{t_i}^{meta}) - final \right]^+ \cdot (\widehat{exc}_s + \varepsilon_{t_i}^{exc}) \dots$$
$$\dots - \frac{1}{N} \sum_{i=1}^{N} \left[final - (meta_s + \varepsilon_{t_i}^{meta}) \right]^+ \cdot ((\widehat{def}_s + \varepsilon_{t_i}^{def}) + g) - \kappa \left| meta_s - final \right|^2, \tag{4.13}$$

for some considerably large N and randomly selected hours $t_1, ..., t_N$ and a regularisation constant κ. We have to mind that these sampled hours $t_1, ..., t_N$ have to be earlier than the hour we solve the optimisation problem (4.13) for. In particular, we do not only need $t_i < s$ for $i = 1, \dots, N$ but the historical errors for the balancing energy prices $\varepsilon_{t_i}^{exc}$ and $\varepsilon_{t_i}^{def}$ must be available already. To ensure that we only use data that is known at that time we have to take into account a lag of about three months to be on the safe side. This yields that $t_i \leq s - 13w$, with $w = 168 (= 7 \cdot 24)$, as three months are approximately 13 weeks.
Note that the day-ahead term does not incorporate a sample average, which is not entirely consistent with the SAA approach. We assume that our estimators are unbiased and have errors of zero mean. Hence this should not change the outcome significantly, but saves precious computation time.
Moreover, in theory we would draw N random samples for each hour we solve the optimisation problem (4.13) for. Although in practice we draw N random samples only once for each year to drastically reduce the computational time needed.
Due to runtime issues we only use the candidates for κ which we obtained above for the expected value problem (4.10).

Now we have certain fluctuation in the terms dealing with the balancing energy prices on average. Therefore we can finally avoid "black or white" situations we mentioned before and obtain reasonable intermediate scenarios. The SAA problem (4.13) takes into account the estimated relationships of the prices to each other as well as the possibility to deviate from the forecast.

Strategic adjustment policies solving the above **optimisation** problem (4.13), i.e., using the **Sample Average Approximation**, to derive optimal $final$ get the label **OPT(SAA)**.

Adjustment Policy	2014	2015	2016	2017
Status Quo	93	92.13	90.17	91.45
OPT(SAA) LM (Kappa = 0)	93.56	91.75	84.68	91.37
OPT(SAA) LMr (Kappa = 0)	96.1	91.74	83.31	90.57
OPT(SAA) LM (Kappa = 0.5)	92.99	92.51	90.72	91.08
OPT(SAA) LMr (Kappa = 0.5)	93.83	92.69	90.58	92.07
OPT(SAA) LM (Kappa = 1)	92.78	92.39	91.09	90.94
OPT(SAA) LMr (Kappa = 1)	93.4	92.53	91.05	92.16
OPT(SAA) LM (Kappa = 1.5)	92.6	92.29	91.15	90.85
OPT(SAA) LMr (Kappa = 1.5)	93.18	92.37	91.21	92.11

Table 4.11: Ratings of OPT(SAA) regularised

Finally we have policies that are mathematically consistent and better than the currently used status quo at the same time. Both “OPT(SAA) LM” and “OPT(SAA) LMr” for $\kappa = 0.5$ and $\kappa = 1$ are better throughout all investigated years. Since the reference revenue by the best case is not the same every year, one has the check the absolute figures of the revenue to judge which policy is the better one in total. Moreover, we see that the more recent data for da in “LMr” pays off like in the expected value solution setting - in general it is better than “LM”.

5 Discussion

To conclude this work, let us recap the ideas, concepts and results we came across above.

First of all we started with describing the problem. We developed the idea of strategic adjustment by looking at the behaviour of prices and their relative positions to each other. Further we saw that the problem of finding "optimal" policies splits into two parts - how to predict the prices and what to do with the forecast. Then we built a framework for evaluation and introduced a rating to compare different policies, as objective reference we took a sensible best case. When analysing the company's policies which were used in the past, we saw that strategic adjustment of the schedule gives significantly more revenue than submission of the initial wind power forecast alone and therefore is justified. Moreover, we attempted to find a special adjustment policy for holidays which did not succeed. However we saw, that the potential extra gain on holidays is not significantly more than on average days during a year. We also investigated limitation by the TSO which yielded insight at which capacity utilisation rates the company is limited the most.
Subsequently, we developed two approaches to predict the prices. We had two versions of historic averages on the one hand and employed univariate linear models for each of the prices on the other hand. When establishing own adjustment policies, we saw that already the heuristic to obtain adjustment coefficients produces reasonable results. Finally, we turned away from adjustment coefficients and looked at optimisation problems to generate optimal values for the production schedule.

The final approach by using SAA techniques generated two possible alternatives to the type of strategic adjustment which is currently used by the company. However already the heuristic gives considerably beneficial results, too.

Additionally, we came across several subjects which may be investigated in future works. We saw that one could possibly look at potential limitation and not only the actual limitation that happened in the past, as there is data available to do so. This could give further insight of how to modify the schedule to avoid limitation. Furthermore, as our attempt of holiday adjustment was not significantly beneficial, one could study the behaviour of the prices on holidays to maybe predict them in a better way. Moreover, we used univariate linear models when predicting the prices. Other attempts could be multivariate models incorporating all the prices at once to benefit from possible interconnected information, or autoregressive models to address the time-series aspect of the prices.

6 Appendix

6.1 Deutsche Zusammenfassung (German Summary)

Bei der Stromproduktion für den täglichen Spotmarkt mittels Windkraftanlagen muss man täglich einen Produktionsfahrplan für den folgenden Tag übermitteln. Aufgrund von unterschiedlichen Preisen für vorveranschlagte Energie und davon in beide Richtungen abweichende Energie, kann es vorteilhaft sein, einen Fahrplan abweichend von der initialen Windleistungsvorhersage zu übermitteln. Das ermöglicht „strategische Anpassung" des Fahrplans aufgrund der Preissituation. Die Preise werden erst zwei bis drei Monate später veröffentlicht, daher muss man Vorhersagen über die Preissituation treffen um effektive strategische Anpassung zu entwickeln.
In dieser Arbeit analysieren wir die Strategien, die von einem Unternehmen in der Vergangenheit verwendet wurden, sowie gewisse andere Aspekte wie beispielsweise Abregelung der Produktion oder Anpassung an Feiertagen. Wir begründen zudem einen Rahmen zur Evaluierung und vergleichen verschiedene Strategien relativ zu einem „besten Fall". Darüber hinaus behandeln wir zwei Konzepte für die Vorhersage der Preise - „historische Mittelwerte" und Modelle der linearen Regression. Anschließend bearbeiten wir zwei Herangehensweisen an strategische Anpassung; die erste inspiriert von der Heuristik des Unternehmens, die zweite als Lösung eines stochastischen Optimierungsproblems. Für Letzteres verwenden wir einen Ansatz mithilfe von „Sample Average Approximation".

7 References

[1] Badi Baltagi, *Econometrics*, Springer-Verlag Berlin Heidelberg, Fifth Edition, 2011.

[2] Jürgen Groß, *Linear Regression*, Springer-Verlag Berlin Heidelberg, 2003.

[3] Paul A. Ruud, *An Introduction to Classical Econometric Theory*, Oxford University Press, 2000.

[4] Takeshi Amemiya, *Advanced Econometrics*, Harvard University Press, 1985.

[5] Ludwig Fahrmeir, Christian Heumann, Rita Künstler, Iris Pigeot, Gerhard Tutz, *Statistik - Der Weg zur Datenanalyse* (German), Springer-Verlag Berlin Heidelberg, 8. Auflage, 2016.

[6] Alexander Shapiro, Darinka Dentcheva, Andrzej Ruszczynski, *Lectures on Stochastic Programming - Modeling and Theory*, Society for Industrial and Applied Mathematics and the Mathematical Programming Society, 2009.

[7] George B. Dantzig, *Linear programming under uncertainty*, Management Science, Vol. 1, Issue 3-4, pp. 197-206, 1955.

[8] E. Martin L. Beale, *On minimizing a convex function subject to linear inequalities*, Journal of the Royal Statistical Society, Series B (Methodological), Vol. 17, No. 2, pp. 173-184, 1955.

[9] John R. Birge, Francois Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag New York, 1997.