FACULTY OF MATHEMATICS AND PHYSICS
Charles University

MASTER THESIS

Tereza Novotná

Business Models for Emission Management

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: doc. RNDr. Ing. Miloš Kopa, Ph.D.
Study programme: Mathematics
Study branch: Financial and Insurance Mathematics

Prague 2021

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I would like to dedicate this work to my best friend Berry. He stays next to me for the whole time, has unbelievable patience and makes me happy every day.

Also huge thanks belongs to my family and friends for the support during all my years on university.

Last but not least I would like to thank to the supervisor doc. RNDr. Ing. Miloš Kopa, Ph.D. who was always opened to all my questions and ideas and give me great advice thanks to which I was able to finish this work.

Title: Business Models for Emission Management

Author: Tereza Novotná

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Ing. Miloš Kopa, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis focuses on multistage stochastic programming for CO_2 emissions management models. It defines the multistage stochastic programming in general, portfolio selection problem with utility function but also with other new approaches such as using risk measures, chance constraint or second order stochastic dominance. For testing this models we need to also reformulate our problems to fit scenario tree which are generated. For some models we also need to reduce the dimension of scenario tree. Thus some techniques for scenario tree dimensionality reduction are discussed. We try to apply all these approaches to our data and get results on real data from power energy sector. For this sector, the decisions about allowances might be very important as they are not granted any allowances.

Keywords: emissions management stochastic optimization

Contents

Introduction

As the world tries to deal with global warming, new challenges arise in terms of running a business. In this work, we discuss company´s obligation to cover released greenhouse gasses with emission allowances. Namely, we consider company which produce electric energy in Czech republic. This sector is highly regulated and since year 2020, companies in European Union in this sector do not get any allowances for free but they need to buy all of them. There is multiple random inputs that the company needs to challenge. First, the price of the allowances which in the past showed great volatility and then the market demand for power. The allowances might be traded as other assets for example as spots, futures or options or we can keep them for future use which is called banking.

In this thesis, we are dealing with risk averse company that want to do in some way optimal decisions so it earns as much as possible while avoiding extremely risky situations. The allowances are used at the beginning of each year to cover released CO_2 from production. The decision about how many allowances we want to buy is then changing every year based on the observation of market development. As the company wants to have everything under control, this leads us to multistage stochastic optimization problem. Similar models were previously developed. In [12] a model for a steel company was developed which considers multiple financial instruments as well as both CER and EUA allowances. This model is not, however, multistage but it considers only the decision for the following year. Also the CER allowances can no longer be exchange for EUAs ones. The problem is also discussed in [11] for a steel company. This model is multistage mean-risk model and uses CVAR as the risk measure. It was used as the base model, applied on a new data from different industrial sectors and further enhanced. Our model has 5 stages with 4 decision periods about the amount of production. We have also tried other approaches to risk modelling. The second model introduces relaxation of strict demand constraint with chance constraint formulation. The third model replaces mean-risk model with second order stochastic dominance program which tries to dominate the benchmark strategy. The last model does not use risk measure but instead of it, it uses the power utility function.

The thesis is structured in a following way: In the first chapter, we discuss the carbon trading, regulations and financial instruments. In the second chapter, we give the theoretical background which is necessary to construct models in chapter 3. Here, 4 different models are introduced. In empirical study we first introduce the data that we will be dealing with, generate scenario trees which predicts the future development of both prices and market demand. In chapter 5, we connect previous chapters and run models with these data. Finally, we compare the results and get to some conclusion about the company´s behaviour under different choices of model parameters and different strategies.

1. Carbon Trading

Global warming is often discussed topic world-widely. Scientific consensus on climate states that it is probably caused by greenhouse gasses emissions mainly by excessive emission of carbon dioxide. CO_2 is released by different sources, some of them are natural like respiration of animals or decaying plants, others are produced by human activities, mainly threw burning fossil fuels. The ones that come from nature are removed again by nature itself, but the excess is caused by human activity. Thus people were in need to find solution how to deal with this issue. In 1997 the Kyoto Protocol [8] was adopted and came into force in 2005. It is an international treaty which states emission limitations and guidelines for originally 37 industrialized countries and European Union countries.

1.1 EU ETS allowances

Kyoto Protocol also suggests some ways for countries to help them meet their limits. The biggest operating trading scheme for Carbon Trading is the EU emissions trading system (EU ETS) which will be the trading scheme considered in this thesis. EU ETS is said to be the world's first major carbon trading market. Over 11 000 industrial companies which are based in EU member states and three EEA-EFTA countries has to cover its greenhouse gasses emissions with EU allowances (EUA). They have to monitor emissions for calendar year and till the end of the following April, these has to be covered with EUAs [3]. This trading system works based on "cap and trade" system. Cap is set for total amount of CO_2 (and for other greenhouse gasses) which can be produced by all companies included in heavy energy-using installations in considered countries. Some allowances are given to companies for free others are traded on a market. This cap and its percentage of allowances given for free are reduced over time. At the end of each year, a company has to give enough allowances to cover its CO_2 production. If company has some unused allowances, it can keep them for following years or trade them to another company.

This system was set up in 2005 and its main goal is to achieve climate-neutral Europe till 2050. That should be done in 5 phases. As this thesis is written mainly during year 2021, we are just at the very beginning of phase 4. In this phase , which will apply between years 2021-2030, the annual cut of allowance is going to be more progressive than in previous years. Number of allowances in circulation on market is also regulated and when surplus occurs, there is established mechanism called Market Stability Reserve that deals with it and helps to prevent future market shocks. Another thing worth mentioning for purposes of this thesis is that EUAs are connected to certain phase. Because of that emissions produced in 2020 cannot be covered with allowances for phase 4. On the contrary, emissions produced in year 2021 can be covered with allowances from phase 3. There are some other mechanisms to cover greenhouse gasses emissions. One of them is called clean development mechanism. This mechanism was developed to support countries not included in Kyoto Protocol agreement in emission-reduction projects. From these projects such countries can get CER credits (Certified Emission Reduction credits) which can be traded on a market. In the previous phases, companies were allowed to use part of these to cover its emissions but since phase 4, the usage of international allowances (e.g. CER credits) is prohibited.

This EU system has also some regulatory measures specific for some industries. As we are using data from power generator sector from Czech Republic, these specifics should be mentioned here as power generation sector has many rules by itself which is understandable as energy sector is responsible for great part of greenhouse gasses produced in industry. According to [3], all power generators are obliged to buy all their allowances since 2013 (i.e. since beginning of phase 3). An exception was given to 10 states, that entered EU in 2004 for both phases 3 and 4, they can allocate decreasing number of allowances to existing power plants. 8 out of this 10 states, including Czech Republic, has decided to use this opportunity in phase 3. Since phase 4, only Bulgaria, Romania and Hungary have decided to continue in giving allowances for free to power plants. Consequently, power generators located in Czech Republic has to pay for all allowances they need to cover its emissions production.

1.2 Financial instruments for allowances trading

Carbon allowances including EUA can be traded in a similar way as other financial instruments. Each tonne of CO_2 must be covered with 1 EUA. They can be bought either in primary auction or at secondary market. Prices at auction might differ from those at secondary market, however if the difference is too big, regulators move its price to be closer to the secondary market one. In this thesis we will consider the most common ways for trading such as spot trade, futures and banking. Lets define here this approaches:

Definition 1.

- *spot trade refers to immediate delivery of a financial instrument. Consequently, we do not consider any interest rates or time to maturity.*

- *futures obligate both buyer and seller to make a transaction on a predetermined date for agreed price.*

- *options give to a buyer the possibility of buying or selling the underlying asset for agreed price. Put options give the holder right to buy the asset for agreed price, call options give them the right to sell the asset for agreed price.*

- *banking refers to saving surplus of allowances or credits to cover emissions in following years*

Spot prices can be modeled as time series. Futures might be a good way to reduce market risk. Probably the most general approach to modelling future prices is cost-of-carry model which is widely used, as in general the future prices tends to converge to spot price with closer maturity date. This approach was also used in for example in [12]. Because of the model complexity and based on results from [12] we decided not to consider options.

2. Multistage Stochastic Programming

The aim of stochastic programming in general is to find optimal solution while considering uncertainty. While in some cases it is admissible to imagine our world as deterministic and find answers to our questions without randomness, very often it would be huge negligence and that found solution might be far from reality. In this thesis, we need to deal with random prices for allowances, randomness of demand for company's product and thus random need of production. In this chapter we want to build some solid theoretical foundation for understanding multistage stochastic programming in general to get some better idea of how we can use it in practice.

2.1 Portfolio Selection Problem

Let's consider common financial problem of portfolio selection. We own some capital W_0 which we want to invest in some of N assets. These assets has random return rates $R_i,\ i \in \{1, \dots, N\}$ per year. We are thinking about investments in following $T = 3$ years. At the beginning of the first year, we want to make in some sense optimal decision of amounts $x_{i,1}$ invested into each of considered assets. As one of assets can be money, for invested amounts it holds: $W_0 = \sum_{i=1}^{N} x_{i,1}$. At the end of the first year we observe returns $R_{i,1}$ (i.e. the actual realizations of returns). Thus we possess $W_1 = \sum_{i=1}^{N} \xi_{i,1} \cdot x_{i,1}$, where $\xi_{i,1} = 1 + R_{i,1}$. At the beginning of second year, after observing returns from the first period, we can change our decision and choose different distribution $\mathbf{x}_2 = (x_{1,2}, \dots, x_{N,2})$ of capital W_1 into N assets and then again observe returns. After that the whole reinvestment process is repeated.

We are trying to make an optimal decisions which will lead to maximal value of expected utility U of our final wealth W_3.

$$\max \mathbb{E}[U(W_T)] \tag{2.1}$$

The value of our final wealth W_T depends on previous investment decisions and returns realizations $\mathbf{x}_t, \boldsymbol{\xi}_t,\ t \in \{1, \dots, T\}$ as $W_T = \sum_{i=1}^{N} \xi_{i,T} x_{i,T}$. To stress that decisions made at the end of each stage depends on observed returns the following notation $x_{[t]} = x_t(\xi_{[t]})$, where $\xi_{[t]} = (\xi_1, \dots, \xi_t)$ might be useful. Under this notation we can rewrite 2.1 as stochastis programming problem:

$$\max\ \mathbb{E}[U(W_T)|\xi_{[T-1]}]. \tag{2.2}$$

Constraints for this optimization problem are model specific and are stated later in chapter 3. As one can see, decision made in each stage should only depend on already observed process realizations but not on future. This condition is called the nonanticipativity condition and it is one of the basic characteristic constraints in multistage stochastic programs.

In general as the multistage stochastic program with T stages we are considering some random process $\{\xi_t\}_{t=0}^{T}$ which realizations are trajectories and with some decision vector process x = $\{x_0, \ldots, x_T\}$ that is measurable function of ξ. Then we can use the formulation of general multistage stochastic program which was stated e.g. in [7].

$$\min_{x_0 \in \mathcal{X}_0} f_0(x_0, \xi_0) + \mathbb{E}\left[\inf_{x_{[1]}} f_1(x_2, \xi_2) + \mathbb{E}\left[\cdots + \mathbb{E}\left[\inf_{x_{[T]}} f_T(x_T, \xi_T)\right]\right]\right] \tag{2.3}$$

where $f_t,\ t \in \{0, \ldots, T\}$ are continuous functions such that the expected values in 2.3 are finite.

2.1.1 Utility Functions

Previously we have mentioned the utility function for which we want to find the optimal values. Very often we are dealing with risk averse investor which typically requires as much profit as possible while minimizing the risk. One possible approach is with common utility functions such as power utility function, or the exponential one. Traditional approach to investors risk aversion modeling is using utility functions. We can define the utility function in the following way:

Definition 2. *Let $U : I \to \mathbb{R}$ is continuous and non-decreasing function on $I \subseteq \mathbb{R}$. Then the function u is called the utility function.*

As we mainly want to consider globally risk averse investor, we will work with the utility functions which are strictly concave. One of the most commonly used utility functions is the power utility function $u(w) = \frac{w^{1-\gamma}}{1-\gamma}$ and chosen parameter $\gamma \geq 0,\ \gamma \neq 1$. In the reference literature this parameter is usually chosen as 3 or 4. For $\gamma = 4$ we can use the formula 2.1 and get the optimization problem rewritten as

$$\max \mathbb{E}\left[\frac{W_T^{-3}}{-3}\right] = \mathbb{E}\left[\frac{-1}{3W_T^3}\right]$$

Another common example is the exponential utility function with parameter $\delta > 0$ for risk averse investor which in general take the form $u(w) = -e^{-\delta w}$. This utility function is often used when dealing with normal distribution of the random variable w. In terms of the formula 2.1 for $\delta = 1$ it takes the form

$$\max \mathbb{E}\left[-e^{-W_T}\right]$$

.

2.2 Risk Measures

Nowadays we developed multiple risk measures and often we want them to meet certain conditions. Let´s consider random variable $Z = Z(\omega), X = X(\omega), Y = Y(\omega)$ which are random outcomes from a sample space $(\Omega, \mathcal{F})$, where

$\omega \in \Omega$ and $\mathcal{F}$ is its σ-algebra. According to [2] we define risk measure as function ρ which maps Z into extended real line $\mathbb{R}$. For $p \in [1, \infty)$. We denote $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ the functions of random variables Z, X and Y for which the risk measure $\rho(\cdot)$ is defined. From this definition we assume that random variables has finite $p-th$ moment. w.r.t. probability measure P.

One of typical approaches for using risk measures to select optimal portfolio is considering mean-risk models ,which, as the name suggests, are models which connect expected value of losses and selected risk measure which is in compliance with the initial motivation for developing risk measures. There are several risk measures that are typically used such as variance, standard deviation, semi-deviation, value at risk, conditional value at risk and others. There exists a group of risk measures which behave in some way better and are usually preferably used in practice. These measures are called coherent risk measures and we can define them as:

Definition 3. *Risk measure ρ is called coherent risk measure, if for random loss $X,\ Y \in \mathcal{Z}$, where $X \succeq Y$ denotes the pointwise order, it satisfies conditions :*

1. $\rho(tX + (1-t)Y) \leq t\rho(X) + (1-t)\rho(Y),\ \forall X,\ Y \in \mathcal{Z},\ \forall t \in [0,1]$ *is called convexity condition*
2. *if* $X \succeq Y$ *then* $\rho(X) \leq \rho(Y)$ *is called monotonicity condition*
3. *if* $a \in \mathbb{R}$ *then* $\rho(X + a) = \rho(X) + a$ *is called translation equivariance*
4. *if* $t > 0$ *then* $\rho(tX) = t\rho(X)$ *is called positive homogeneity.*

For more details and further properties see [2].

2.2.1 CVaR

As an example of coherent risk measure, which we are going to work with later, we define Conditional Value at Risk. This measure is also known as Mean Excess Loss or Tail VaR. Under CVaR_α one can understand the expected value of loss which will happen in $(1-\alpha) \cdot 100\%$ cases. Formally we can define it in a following way:

Definition 4. *Let X be a random loss. Then Conditional Value-at-Risk is defined as:*

$$CVaR_\alpha(X) = \inf_{a \in \boldsymbol{R}} \left\{ a + \frac{1}{1-\alpha} E[X - a]^+ \right\}$$

for $\alpha \in (0,1)$ *where* $[\cdot]^+$ *denotes positive real number.*

As previously mentioned, we can use it in some general mean-risk model which for losses Y and portfolio weights x has the form

$$\min_x (1-\lambda)\mathbb{E}[Y^T x] + \lambda \mathrm{CVaR}_\alpha(Y^T x) \tag{2.4}$$

Another advantage of using CVaR for optimization problem is that when used in linear program it keeps the linearity and so it is not computationally that demanding as many others risk measures. If we consider some discrete random distribution with $s \in \{1, \ldots, S\}$ scenarios which are gained with respective probabilities p_s, then this risk measure might be rewritten in the following linear programming form :

$$\begin{aligned} \min \quad & a + \frac{1}{1-\alpha} \sum_{s=1}^{S} p_s z^s \\ \text{s.t.} \quad & z^s \geq x^T Y^s - a \\ & z^s \geq 0 \end{aligned} \tag{2.5}$$

2.3 Chance Constraint

Sometimes it might be convenient or even necessary to deal with random variable also in the model constraints. In some cases we will need some stochastic constraint (of which we know its distribution) to be fulfilled in all cases. This might be however sometimes too strict and we might not found any solution. Or in other cases we do not mind to risk a little and so we can be comfortable with some small chance that the constraint would not state. In these cases, we can use the so called chance constraint. In general the optimization problem with a joint chance constraint for random vector Z and decision vector x takes a form:

$$\min c(x) \tag{2.6}$$

$$s.t.\ P[Z : x \in \mathcal{X}(\mathcal{Z}), j \in \mathcal{J}] \geq 1 - \epsilon \tag{2.7}$$

where $\mathcal{X}(\mathcal{Z}) = \{x \in \mathcal{X}_0 : h_j(x, Z) = 0 \forall j,\ g_k(x, Z) \leq 0 \forall k\}$ is nonempty set, $c : \mathbb{R}^n \to \mathbb{R}$, $h_j(x, Z)$, $g_k(x, Z)$ are model constraints, $\mathcal{K}$, $\mathcal{J}$ are index sets, Z is s dimensional random vector and $0 \leq \epsilon \leq 1$ is parameter that regulates the probability that we want this constraint to be fulfilled.

In practice, to be able to computationally solved chance constrained problems, when dealing with discrete distribution, we take advantage of mixed integer programming which is based on the following theorem:

Theorem 1. *Let f, $g(\cdot, Z) : \mathbb{R}^n \to \mathbb{R}$ be real functions, $X \subseteq \mathbb{R}^n$ and Z be real random vector, with finite discrete distribution $z_1, \ldots, z_M$ with probabilities $p_1, \ldots, p_M$, $\sum_{m=1}^{M} p_m = 1$ then for $\epsilon \in (0,1)$ small then if we are dealing with chance constraint program:*

$$\begin{aligned} & \min c(x) \\ & s.t.\ P[g(x, Z) \leq 0] \geq 1 - \epsilon \end{aligned}$$

which can be rewritten to mixed integer programming as:

$$\begin{aligned} & \min c(x) \\ & s.t \sum_{m=1}^{M} p_m y_m \geq 1 - \epsilon \\ & g(x, z_m) \leq C(1 - y_m) && m = 1, \ldots, M \\ & y_m \in \{0, 1\} && m = 1, \ldots, M \\ & x \in \mathcal{X} \end{aligned}$$

where $C \geq \max_{m=1,\ldots,M} \sup_{x\in\mathcal{X}} g(x, z_m)$

More exhaustive informations to this theorem and other relationships might be found for example in [4].

2.4 Second Order Stochastic Dominance

While considering more portfolios with stochastic variables, sometimes we do not require to reach its best possible value in the terms of some risk measure but we would rather our considered portfolio to be better than some other one. The one which we want to beat is called benchmark strategy. There are multiple approaches in which we can measure whether or not our strategy is better than the benchmark one. Lately stochastic dominance measures became very popular. Usually we speak about first order and second order stochastic dominance. For exhaustive definition see [2]. For our case we choose to use second order stochastic dominance which can be defined as:

Definition 5. *Let* $F_{r^T\lambda}(y)$ *be cumulative distribution function for portfolio with weights* λ *and random returns* r. *Let* τ *be weights of benchmark portfolio. We say that portfolio* λ *dominates* τ *by the second-order stochastic dominance when*

$$F^{(2)}_{r^T\lambda}(y) \leq F^{(2)}_{r^T\tau}(y), \quad y \in [a,\ b]$$

where $F^{(2)}_z(y) = \int_{-\infty}^{y} F_z(x)dx$ *and* $[a,\ b] \subseteq \mathbb{R}$. *Second order stochastic dominance of portfolio* λ *to portfolio* τ *is denoted as* $r^T\lambda \succeq_{SSD} r^T\tau$.

In case of the second order stochastic dominance portfolio selection problem is equivalent to risk averse attitude of the investor. When considering scenarios, at each time period we can act as when dealing with discrete distribution. In this case, following Theorem 2 might be helpful.

Theorem 2. *Let* X *and* Y *be a random variables with univariate discrete distributions with* M *atoms which on its probability space attains values* $X_1 \leq X_2 \leq \ldots, \leq X_M$ *and* $Y_1 \leq Y_2 \leq \ldots, \leq Y_M$ *respectively. Then following equivalency states:*

$$F^{(2)}_X(z) \succeq_{SSD} F^{(2)}_Y(z) \iff \sum_{i=1}^{m} X_i \geq \sum_{i=1}^{m} Y_i \quad m = 1,\ldots,M$$

Another theorems which captures the second order stochastic dominance and can be very useful in scenario approach is using CVaR for which holds:

$$r^T\lambda \succeq_{\mathrm{SSD}} r^T\tau \iff CVaR_\alpha(-r^T\lambda) \leq CVaR_\alpha(-r^T\tau)$$

Then based on [6] theorems 3 and 4 states:

Theorem 3. *For* $\alpha \in \langle \frac{k}{M}, \frac{k+1}{M} \rangle$ *and* $\alpha \neq 1$ *where* k *is the index of the* k*-th smallest element of the scenarios generated from univariate distribution with* M *atoms of random variable* X *i.e.* $X_1 \leq X_2 \leq \ldots, \leq X_k, \ldots, \leq X_M$ *then it holds*

$$CVaR_\alpha(X) = (X_{k+1}) + \frac{1}{(1-\alpha)M} \sum_{i=k+1}^{T} (X_i - X_{k+1})$$

for $k \in \{1,\ldots,M-1\}$ *and* $CVaR_1(X) = X_M$

Theorem 4. *Let X and Y be random variables with discrete univariate distributions with M atoms which take values X_m and Y_m, $m \in \{1, \dots, M\}$ respectively, then*

$$X \succeq_{SSD} Y \iff CVaR_\alpha(-X) \leq CVaR_\alpha(-Y) \quad \forall \alpha \in \left\{0, \frac{1}{M}, \dots, \frac{M-1}{M}\right\}$$

2.5 Scenario Trees

Until now, we did not specified distribution or ways of modelling stochastic process $\{\xi_t\}_{t=1}^T$. As one can imagine, if we would consider some continuous distribution, as it can be suitable for random market returns, it might be very difficult to find optimal solution. Because of that, one of the widely used techniques is to generate scenario trees. We can think of it as the continuous distribution which we consider as known (otherwise we can use some approximation for example based on previously observed values). Let us consider now that we can invest either in stocks or keep our wealth in a bank. For this case, we have $n = 2$ possible options of wealth distribution at each time period. We have some primary observations of considered stock on a market and we noticed that its return rate R_1 is either -0.1 or 0.2 with probability distribution $P[R_1 = -0.1] = p_{11} = 0.4$, $P[R_1 = 0.2] = p_{12} = 0.6$. On our bank account, we have an interest rate $R_2 = 0.04$ which will stays unchanged for at least following 4 years, thus $P[R_2 = 0.04] = p_{2\cdot} = 1$. Under these distributions, we are considering only two possible combinations for returns $\xi_{it} = R_{it}$, either $\xi_t^1 = (0.9, 1.04)$ with probability $p_1 = 0.4$ or $\xi_t^2 = (1.2, 1.04)$ with probability $p_2 = 0.6$. We can now look at scenario tree in figure 2.1. As one can observe, with two children from each node, with $T = 3$ we are dealing with $2^3 = 8$ scenarios which suggests that with each additional time period, problem became much more complex and computationally difficult to find its solution.

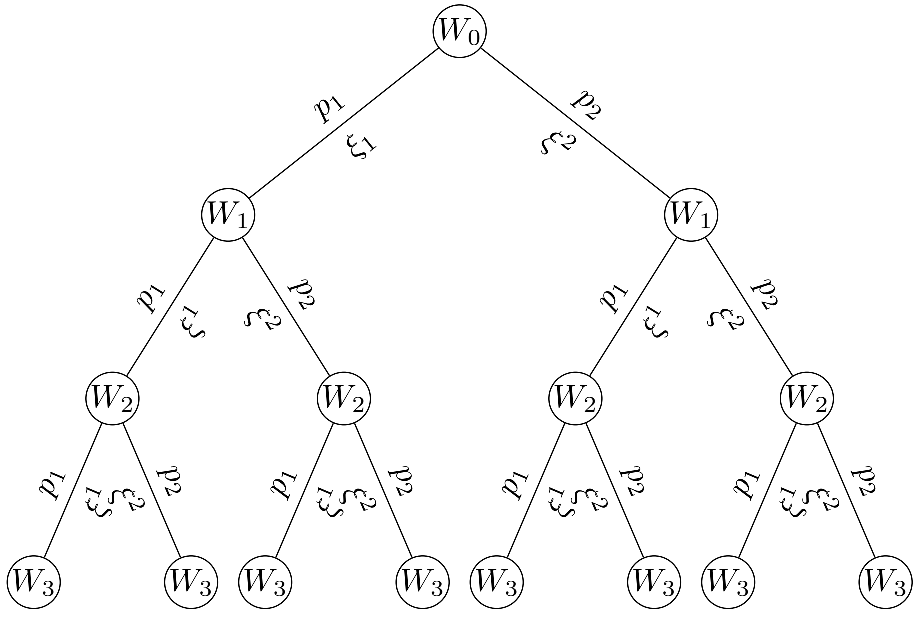

Figure 2.1: Scenario tree (T=3)

As one can notice from the 2.1, each node at time $0 < t < T$ has exactly one ancestor node at time t-1 and multiple children nodes. For root t=0, there is no ancestor and multiple children and for all leaves t=T we have exactly one ancestor and no children. This is the general definition of scenario tree structure. In general, it might happen that 2 nodes at time t has the same value but they are separate because they have different history. As scenario we mean one path from root node t=0 to some leaf t=T, i.e. each scenario represents history of the random process $\{\xi_t\}_{t=1}^T$.

2.5.1 Generating Scenario Trees

Multiple approaches to generating scenario trees might be found in a literature. As the first step, we need to fit the distribution of the data process. This might be some ARCH or GARCH models, ARIMA models, binomial or trinomial models, random walks adopted in Monte Carlo generators. According to [7] with models time discretization, or calibrated models one can generate arbitrarly many sample paths ω. We can also assume that these sample paths are equiprobable.

As the next step, we want to delineate the number of stages and the so-called branching scheme, which tells us with how many children we deal with for each node. In some cases, the time period of each stage is given naturally as we can change our decision once a year, in other cases we have to figure out the most suitable time length for each stage. This also suggest that between each two stages might be different time period. One option is to use clustering method. One of the very popular way is the K-means clustering. In this algorithm, the K stands for the number of desired clusters also, which in our case corresponds to K_t preselected number of scenarios at time t. Each cluster has than one centroid, which is the referential point to which the algorithm assign closest points in the sample space. The problem is, that the number of points that are assigned to each cluster is not given and so one cluster might include hundreds of points and another one is going to include only the centroid itself. When using for generating scenario tree, this might be very inconvenient as the nodes in scenario tree gain significantly different probabilities. This problem solved new approach to K-means clustering called constrained K-means clustering which was introduced in [1]. With this approach we try to solve non-linear mixed-integer program:

$$\min_{C\in\mathbb{R}^{KxM}, Z\in\{0,1\}^{NxK}} \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K} z_{i,k}||x_i - c_k||_2^2 \tag{2.8}$$

$$s.t.\ \sum_{i=1}^{N} z_{i,k} \geq n_k,\ k \in \{1,\dots,K\} \tag{2.9}$$

$$\sum_{k=1}^{K} z_{i,k} = 1,\ i \in \{1,\dots,N\} \tag{2.10}$$

where x_i are the row vectors of sample data matrix X^{NxM}, matrix $Z = (z_{i,k})_{i,k=1,1}^{N,K}$ is the matrix of binary variablech which codes to which cluster (or to which centroid $c_{k,m}$) is the data sample assign and $k \in \{1,\dots,K\}$ is the minimum size of each cluster. For each time period, we run the clustering algorith separately.

If we are dealing with Markov structure of data of the form $\xi_t = P\xi_{t-1} + v_t$, we can discretize the distribution of random disturbations v_t and the rest is deterministic. We can also consider the case where the transition matrix P depends on time t. In [7] a reader can find more exhaustive explanation of commonly used technique which is called sequential importance sampling.

Another one very common technique for generating scenario trees is also matching moments of distribution or some other statistical properties. As stated for example in [7], it can be proven that for m admissible values of moments, the discrete distribution with those moments exists and its support has at most $m + 2$ distinct values.

2.5.2 Scenario Trees Dimensionality Reduction

In some cases we need further reduction of the scenario trees, usually when the computation power is not big enough. There are several techniques which might be used for scenarios reduction, one of the most popular techniques is the single node reduction algorithm which was developed in [10] and further extended in [9] which introduce the following 4-step algorithm:

1. for each pair of nodes (i,j) which has the same parent calculate $\epsilon_{i,j} = p_i||\xi_i - \xi_j|| + \frac{2p_ip_j}{p_i+p_j}$ where p_i is the absolute probability of the node i and ξ_i its value.
2. find the smallest $\epsilon_{i,j}$ for which we denote (k, l) the values of a pair (i, j).
3. set the node k instead of node j as the parent of children's of node l. and set $p_k = p_k + p_l$, remove node l.
4. repeat whole process until we get desired number of scenarios.

To sum it up, the whole idea is to measure the distance between each pair of two nodes that came from the same parent and reduce its count to desired number. In the equiprobable scenario tree, the algorithm above becomes even easier as $p_i = p_j = \frac{1}{M}$ where M is the count of nodes for that time.

Another method for scenario tree dimension reduction is called nodal clustering and it was introduced in [9] where all details can be found. In summary we want to first define the structure of the final scenario tree (i.e. reduced number of nodes in final scenario tree for each time period) and then for each tree level merge the pair of the closest nodes into new one and reduce them into one so we end up with the desired number of clusters.

There are several others algorithms for scenario tree reduction which might be found for example in [9].

3. Models

The company has to decide about its total production for the next yaer at each time $t = 0, \dots, T-1$. They can decide how much electric energy should be produced in considered time period by which of $n \in N$ different types of power generators. In this section we build base model which is further compared to some of its variations. This model is mainly based on [11] but with some modifications as we are considering different phase and different industrial sector.

3.1 Base Model

The decision about the volume of production for time period $[t,\ t+1)$ can be viewed as $x_t = (x_{t1}, \dots, x_{tN})^T$. These decisions are made for time $t \in \{0, \dots, T-1\}$. There are some limitations for how much energy is each generator type able to produce. These are given as production limits $w = (w_1, \dots, w_N)^T$ which are supposed to be constant in all considered time periods. As the factories are not able to plan the production higher than its capacity, naturally we get the limit (3.3) $x_t \leq w$. The first stochastic variable we are going to consider is the total market demand d_t that does not depend on a source by which the energy was produced. Based on this definition, we want condition (3.2) $\sum_{n=1}^{N} x_{tn} \leq \min(d_{t+1}|d_t)$ to be fulfilled as we do not want to produce in total in all kinds of power generators more than the demand will be. For simplicity of this basic model, let us view margins on energy $\mu_t = (\mu_{t1}, \dots, \mu_{tN})^T$ as deterministic vector for time t. As one can notice, this margin vector is dependent on chosen power production source as the production costs differ.

Until now, we have mentioned only factors and variables independent of CO_2 emissions. For each type of power generator we want to consider the produced amount of CO_2. The vector of this greenhouse gas production per unit of energy is denoted as $h = (h_1, \dots, h_N)^T$. At time t we hold certain amount of allowances e_t which is the resulting amount of allowances after covering emissions produced in time interval $[t-1, t)$. This amount is given as $e_t = e_{t-1} + s_t + \sum_{0 \leq \tau < t} f_\tau^t - h^T x_{t-1}$ for $t \in \{1, \dots, T\}$ (3.6) where f_τ^t and s_t denote the amount of futures with maturity at time t and the amount of spot allowances, respectively. For simplicity we suppose that this company is just starting and so at time $t = 0$ we have $e_0 = s_0$ as we consider that no allowances from past are available and no futures have maturity date at time 0. Another assumption is that no short selling is allowed and that in each time period we need to cover all emissions with allowances which results in conditions $s_t,\ e_t,\ f_t^{t+1}, \dots,\ f_t^T \geq 0,\ t \in \{0, \dots, T\}$ (3.11). As goal of this model is to minimize costs with some preferences of risk aversion, we also need to include prices of spot allowances p_t and prices of futures $q_t^{t+1}, \dots, q_t^T$, which we consider as stochastic parameters. If the company needs more allowances than they can buy on spot with its production profit and that they are possessing at that moment (thanks to banking or futures), they have the option of taking a loan with interest rate ρ. When they have some profit which is not used to buy allowances, the company get the same interest rate ρ

from saving that amount of money. Now we have everything to form the equation for considered cash flows z_t. As mentioned above, we consider time $t = 0$ as the start of this company and so $z_0 = -p_0 s_0$. For $t \in \{1, \ldots, T\}$ the equation is $z_t = m_t^T x_{t-1} - p_t s_t - \sum_{\tau=0}^{T-1} q_\tau^t f_\tau^t + \rho c_{t-1}$ (3.10), where $c_t = \sum_{s=0}^{t} z_s$ (3.9) is the amount of wealth, which could also be negative, gained up to time t. To sum it up, the model can be expressed as:

$$\min \; V_\lambda(-z_0, \ldots, -z_T) \tag{3.1}$$

$$s.t. \; \sum_{n=1}^{N} x_t \leq \min(d_{t+1}|d_t) \qquad 0 \leq t \leq T-1 \tag{3.2}$$

$$x_t \leq w \qquad 0 \leq t \leq T-1 \tag{3.3}$$

$$x_t, s_t, f_t^{t+1}, \ldots, f_t^T \in \mathcal{F}_t \qquad 0 \leq t \leq T-1 \tag{3.4}$$

$$e_0 = s_0 \tag{3.5}$$

$$e_t = e_{t-1} + s_t + \sum_{0 \leq \tau < t} f_\tau^t - h^T x_{t-1} \qquad 1 \leq t \leq T \tag{3.6}$$

$$s_0 \geq 0 \tag{3.7}$$

$$z_0 = -p_0 s_0 \tag{3.8}$$

$$c_t = \sum_{s=0}^{t} z_s \qquad 1 \leq t \leq T \tag{3.9}$$

$$z_t = \mu_t^T x_{t-1} - p_t s_t - \sum_{\tau=0}^{t-1} q_\tau^t f_\tau^t + \rho c_{t-1} \qquad 1 \leq t \leq T \tag{3.10}$$

$$s_t,\ e_t,\ f_t^{t+1}, \ldots,\ f_t^T \geq 0 \qquad 0 \leq t \leq T \tag{3.11}$$

We require the nonanticipativity condition to be fulfilled which is done thanks to equation (3.4) where $\mathcal{F}$ stands for filtration defined as $\mathcal{F}_t = \sigma((\pi_\tau)_{\tau \leq t})$ and $\pi_t = (d_t, p_t, q_t)$ for $t \in \{1, \ldots, T\}$. The decision criterion V_λ, where λ stands for risk aversion parameter, is for basic model considered as mean-multiperiod CVaR which results in decision criterion defined as:

$$V_\lambda(-z_0, \ldots, -z_T) = (1-\lambda)\mathbb{E}[\sum_{t=0}^{T} -(1+\rho)^{T-t} z_t] + \lambda R(-z_0, \ldots, -z_T) \tag{3.12}$$

$$R(-z_0, \ldots, -z_T) = \sum_{t=0}^{T} (1+\rho)^{T-t} \mathbb{E}[\text{CVaR}(-z_t|\mathcal{F}_{t-1})] \tag{3.13}$$

To be able to find the solution in some reasonable time we need to rewrite it using scenario tree with M scenarios which speeds up a lot the time needed for computation. To be able to derive the final form of this optimization model, we need also notation for set of nodes at time t. Assume that at time t we are dealing with set $K_t = \{1, 2, \ldots, \mathcal{K}_t\}$ of nodes and that $|K_t| = \mathcal{K}_t$. For the last time period T it states $|M| = \mathcal{K}_T$. Then in combination with the definition 4 of $\text{CVaR}_\alpha(-z_t)$ while considering the univariate distribution of nodes in scenario trees, we can apply following transformations:

$$(1-\lambda)\mathbb{E}[\sum_{t=0}^{T} -(1+\rho)^{T-t} z_t] + \lambda \sum_{t=0}^{T} (1+\rho)^{T-t}\mathbb{E}[\text{CVaR}(-z_t|\mathcal{F}_{t-1})] =$$

$$= -(1-\lambda)(1+\rho)^T z_0 - (1-\lambda)\frac{1}{M}\sum_{m=1}^{M}(1+\rho)^{T-1} z_{1m} +$$

$$+ (1+\rho)^{T-1}\left[\lambda a_0 + \frac{\lambda}{1-\alpha}\sum_{k_1=1}^{\mathcal{K}_1}\frac{1}{\mathcal{K}_1}[-z_{1k_1} - a_0]^+\right]$$

$$... - (1-\lambda)\frac{1}{M}\sum_{m=1}^{M} z_{Tm} + \left[\lambda \sum_{k_{T-1}=1}^{\mathcal{K}_{T-1}} a_{T-1k_{T-1}} + \frac{\lambda}{1-\alpha}\sum_{k_T=1}^{\mathcal{K}_T}\frac{1}{\mathcal{K}_T}[-z_{Tk_T} - a_{T-1k_T}]^+\right]$$

In the equation above the term a_{t-1k_t} represents the solution of formula in definition 4 for parents of the nodes K_t. Then using the form (2.5) we arrive to this final optimization model:

$$\min \quad -(1-\lambda)(1+\rho)^T z_0 + \frac{\lambda}{M}\sum_{t=0,\ m=1}^{T-1,\ M}(1+\rho)^{T-t-1} u_{tm} +$$

$$+ \left[\sum_{t=1,\ m=1}^{T,\ M}(1+\rho)^{T-t}\left(\frac{\lambda-1}{M} z_{tm} + \frac{\lambda}{(1-\alpha)M} b_{tm}\right)\right]$$

$$\begin{aligned}
s.t.\quad & b_{tm} \geq -z_{tm} - u_{t-1m} && 1 \leq t \leq T \\
& b_{tm} \geq 0 && 0 \leq t \leq T-1 \\
& \sum_{n=1}^{N} x_{tmn} \leq d_{t+1m'} && 0 \leq t \leq T-1 \\
& x_{tmn} \leq w_n && 0 \leq t \leq T-1 \\
& e_0 = s_0 && \\
& e_{tm} = e_{t-1m} + s_{tm} + \sum_{0\leq\tau<t} f_{\tau m}^t - h^T x_{t-1m} && 1 \leq t \leq T \\
& s_0 \geq 0 && \\
& z_0 = -p_0 s_0 && \\
& c_{tm} = \sum_{s=0}^{t} z_{sm} && 1 \leq t \leq T \\
& z_{tm} = \mu_t^T x_{t-1m} - p_{tm}s_{tm} - \sum_{\tau=0}^{t-1} q_{\tau m}^t f_{\tau m}^t + \rho c_{t-1m} && 1 \leq t \leq T \\
& s_t,\ e_t,\ f_t^{t+1}, \ldots,\ f_t^T \geq 0 && 0 \leq t \leq T \\
& \text{and the nonanticipativity conditions for } x_t, s_t, f_\tau^t, b_t, u_t &&
\end{aligned}$$

Where all above must stay for $\forall m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, and $\forall m' : x_{tm} = x_{tm'}$.

3.2 Model with Chance Constraints

In the base model we are using the constraints with the demand which is considered as stochastic. This limit might be too strict and we might be willing to consider more relaxed version with only probabilistic constraint. The equation 3.2 can be rewritten as $P(x_t \leq (d_{t+1}|d_t)) \geq (1-\epsilon)$ (in words we want to be sure that for at least in $1-\epsilon \cdot \%$ cases the final production will be fully used as the market demand will be higher). Here, based on the results which are derived in chapter 5, we suppose that if we decide to produce more than the observed demand in the next time period, we will not use the rest of produced energy from coal which will lower our income and because we have already payed for the production, we need to substract the costs γ_1 which is the cost for 1kWh production from coal power generator. If we denote the cumulative distribution function of d_t as $G(x)$, the inequality can be rewritten as:

$$1 - G(x_t) \geq 1 - \epsilon$$
$$G(x_t) \leq \epsilon$$

With this new approach, the final volume of sold products might be differ from the final production and it rather attains value of $\min[\sum_{n=1}^{N} x_{tn}, (d_{t+1}|d_t)]$. After implementing this probabilistic approach to the model, we get :

$$\min \ V_\lambda(-z_0, \ldots, z_T) \tag{3.14}$$

$$s.t. \ G(x_t) \leq \epsilon \qquad 0 \leq t \leq T-1 \tag{3.15}$$

$$x_t \leq w \qquad 0 \leq t \leq T-1 \tag{3.16}$$

$$x_t, s_t, f_t^{t+1}, \ldots, f_t^T \in \mathcal{F}_t \qquad 0 \leq t \leq T-1 \tag{3.17}$$

$$e_0 = s_0 \tag{3.18}$$

$$e_t = e_{t-1} + s_t + \sum_{0 \leq \tau < t} f_\tau^t - h^T x_{t-1} \qquad 0 < t \leq T \tag{3.19}$$

$$s_0 \geq 0 \tag{3.20}$$

$$z_0 = -p_0 s_0 \tag{3.21}$$

$$c_t = \sum_{s=0}^{t} z_s \qquad 1 \leq t \leq T \tag{3.22}$$

$$z_t = \mu_t^T x_t - (\mu_{t1} + \gamma_1)[x_{t1} - d_{t+1}]^+ - p_t s_t - \sum_{\tau=0}^{t-1} q_\tau^t f_\tau^t + \rho c_{t-1} \qquad 1 \leq t \leq T \tag{3.23}$$

$$s_t, \ e_t, \ f_t^{t+1}, \ldots, \ f_t^T \geq 0 \qquad 0 \leq t \leq T \tag{3.24}$$

Where $V_\lambda(-z_0, \ldots, -z_T)$ is defined in the same way as in the base model 3.1. Here we will take advantage of the theorem (1) with sufficiently big constant C

and in the combination of scenario trees we get:

$$\begin{aligned}
\min\ & V_\lambda(-z_0, \ldots, -z_T) \\
s.t.\ & \sum_{m=1}^{M} \frac{1}{M} y_{tm} \geq 1 - \epsilon && 0 \leq t \leq T-1 \\
& x_{tm} - d_{t+1m} \leq C(1 - y_{tm}) && 0 \leq t \leq T-1 \\
& y_{tm} \in \{0, 1\} && 0 \leq t \leq T-1 \\
& x_{tmn} \leq w_n && 0 \leq t \leq T-1 \\
& e_0 = s_0 \\
& e_t = e_{t-1m} + s_{tm} + \sum_{0 \leq \tau < t} f^t_{\tau m} - h^T x_{t-1m} && 0 < t \leq T \\
& s_0 \geq 0 \\
& z_0 = -p_0 s_0 \\
& c_t = \sum_{s=0}^{t} z_{sm} && 1 \leq t \leq T \\
& z_{tm} = \mu_t^T x_{tm} - (\mu_{t1} + \gamma_1)[x_{tm1} - d_{t+1m'}]^+ - p_{tm} s_{tm} - && 1 \leq t \leq T \\
& - \sum_{\tau=0}^{t-1} q^t_{\tau m} f^t_{\tau m} + \rho c_{t-1m} \\
& s_t,\ e_t,\ f_t^{t+1}, \ldots,\ f_t^T \geq 0 && 0 \leq t \leq T \\
& \text{and the nonanticipativity conditions for } x_t, s_t, f^t_\tau, y_t
\end{aligned}$$

Where all above must stay for $\forall m \in \{1, \ldots, M\}$, $n \in \{1, \ldots, N\}$, and $\forall m' : x_{tm} = x_{tm'}$. With this approach, we can observe the effect of this constraint and its dependency on ϵ parameter in following chapters.

3.3 Model with Second Order Stochastic Dominance

As another approach we apply the second order stochastic dominance. We would like that our profit is at each time period better than chosen benchmark strategy in the sense of second order dominance. As our benchmark strategy we choose multiperiod strategy which is based on the assumption that in each time period we need the production to be lower than minimal demand. This production at time t is then distributed between power generators so each one runs with the same ratio of its maximal capacity (i.e. when the minimum of d_{tm} at time t is r_t, then the vector of benchmark portfolio production is $r_t(\frac{w_1}{W}, \frac{w_2}{W}, \ldots, \frac{w_N}{W})$, $W = \sum_{i=1}^{N} w_i$ and it should also hold that $r_t = \min[d_{tm}]$). We also consider only buying spot allowances only in the amount which is exactly needed to cover emissions r_t and so $s_t = h^T x_t, e = 0$. For its corresponding cash flows it holds : $z_t = m_t^T \mu - p_t s_t + \rho c_{t-1}$. Based on [6] we can use the equivalent definition of stochastic dominance based on CVaR value which is mentioned in the previous chapter. Also based on the definition of CVaR in definition 4 we can rewrite the inequality $CVaR_\alpha(-z_t|\mathcal{F}_{t-1}) \leq CVaR_\alpha(-z_t|\mathcal{F}_{t-1})$ as $\min_a \left\{a + \frac{1}{1-\alpha} E[z_t - a]^+\right\} \leq$

$CVaR_\alpha(-z_t|\mathcal{F}_{t-1})$ .This inequality holds if and only if there exist $a \in \mathbb{R}$ for which it holds. For a finite number of scenarios M which has the same probability, we arrive to the inequality $a + \frac{1}{1-\alpha}\sum_{m=1}^{M}\frac{1}{M}[z_{tm} - a]^+ \leq CVaR_\alpha(-z_t|\mathcal{F}_{t-1})$. To arrive to the linear programming form of the inequality, we can introduce additional variables y_{tm} and so we get three inequalities:

$$a + \frac{1}{1-\alpha}\sum_{m=1}^{M}\frac{1}{M}y_{tm} \leq CVaR_\alpha(-z_t|\mathcal{F}_{t-1}),\ z_{tm} - a - y_{tm} \leq 0,\ y_{tm} \geq 0.$$

When we implement all changes we arrive to final optimization problem:

$$\begin{aligned}
\min\ & \frac{1}{M}[\sum_{t=0,\ m=1}^{T,\ M} -(1+\rho)^{T-t} z_{tm}] & \\
s.t.\ & \sum_{n=1}^{N} x_{tmn} \leq d_{t+1m'} & 0 \leq t \leq T-1 \\
& x_{tmn} \leq w_n & 0 \leq t \leq T-1 \\
& e_0 = s_0 & \\
& e_t = e_{t-1} + s_{tm} + \sum_{0\leq\tau<t} f_\tau^t - h^T x_{t-1m} & 0 < t \leq T \\
& s_0 \geq 0 & \\
& z_0 = -p_0 s_0 & \\
& c_{tm} = \sum_{s=0}^{t} z_{sm} & 0 \leq t \leq T \\
& z_{tm} = \mu_t^T x_{t-1m} - p_{tm}s_{tm} - \sum_{\tau=0}^{t-1} q_{\tau m}^t f_{\tau m}^t + \rho c_{t-1m} & 1 \leq t \leq T \\
& s_t,\ e_t,\ f_t^{t+1}, \ldots,\ f_t^T \geq 0 & 0 \leq t \leq T \\
& a + \frac{1}{1-\alpha}\sum_{m=1}^{M}\frac{1}{M}y_{tm} \leq CVaR_\alpha(-z_{\mu t}|\mathcal{F}_{t-1}) & 0 < t \leq T \\
& z_{tm} - a - y_{tm} \leq 0 & 0 \leq t \leq T \\
& y_{tm} \geq 0 & 0 \leq t \leq T \\
& \text{and the nonanticipativity conditions for } x_t, s_t, f_\tau^t, &
\end{aligned}$$

Where all above must stay for $\forall m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, $\forall\alpha \in \{0, \frac{1}{M}, \ldots, \frac{M-1}{M}\}$ and $\forall m' : x_{tm} = x_{tm'}$.

3.4 Model with Utility Function

Instead of using mean-risk model approach, we can try to use the power utility function as we suggest in previous chapter. We would like to solve:

$$\max\ \mathbb{E}\left[-\frac{1}{3W_T^3}\right] \tag{3.25}$$

$$s.t.\ \ x_t \le d_{t+1} \qquad 0 \le t \le T-1 \tag{3.26}$$

$$x_t \le w \qquad 0 \le t \le T-1 \tag{3.27}$$

$$x_t, s_t, f_t^{t+1}, \dots, f_t^T \in \mathcal{F}_t \qquad 0 \le t \le T-1 \tag{3.28}$$

$$e_0 = s_0 \tag{3.29}$$

$$e_t = e_{t-1} + s_t + \sum_{0\le\tau<t} f_\tau^t - h^T x_{t-1} \qquad 0 < t \le T \tag{3.30}$$

$$s_0 \ge 0 \tag{3.31}$$

$$z_0 = -p_0 s_0 \tag{3.32}$$

$$c_t = \sum_{s=0}^{t} z_s \qquad 1 \le t \le T \tag{3.33}$$

$$z_t = \mu_t^T x_{t-1} - p_t s_t - \sum_{\tau=0}^{t-1} q_\tau^t f_\tau^t + \rho c_{t-1} \qquad 1 \le t \le T \tag{3.34}$$

$$s_t,\ e_t,\ f_t^{t+1}, \dots,\ f_t^T \ge 0 \qquad 0 \le t \le T \tag{3.35}$$

$$W_T = [\sum_{t=0}^{T} -(1+\rho)^{T-t} z_t] \tag{3.36}$$

Again, rewritten with using scenarios, it takes the form:

$$max\ \frac{-1}{M}\sum_{m=1}^{M}\left[\frac{1}{3W_{Tm}^3}\right]$$

$$s.t.\ \ \sum_{n=1}^{N} x_{tmn} \le d_{t+1m'} \qquad 0 \le t \le T-1$$

$$x_{tmn} \le w_n \qquad 0 \le t \le T-1$$

$$e_0 = s_0$$

$$e_{tm} = e_{t-1m} + s_{tm} + \sum_{0\le\tau<t} f_{\tau m}^t - h^T x_{t-1m} \qquad 0 < t \le T$$

$$s_0 \ge 0$$

$$z_0 = -p_0 s_0$$

$$c_{tm} = \sum_{s=0}^{t} z_{sm} \qquad 1 \le t \le T$$

$$z_{tm} = \mu_t^T x_{t-1m} - p_{tm} s_{tm} - \sum_{\tau=0}^{t-1} q_{\tau m}^t f_{\tau m}^t + \rho c_{t-1m} \qquad 1 \le t \le T$$

$$s_t,\ e_t,\ f_t^{t+1}, \dots,\ f_t^T \ge 0 \qquad 0 \le t \le T$$

$$W_{Tm} = [\sum_{t=0}^{T} -(1+\rho)^{T-t} z_{tm}]$$

and the nonanticipativity conditions for x_t, s_t, f_τ^t

Where all above must stay for $\forall m \in \{1, \ldots, M\}$, $n \in \{1, \ldots, N\}$, and $\forall m'$: $x_{tm} = x_{tm'}$.

4. Data

For our models which are stated in previous chapter, we need to collect several data for empirical study. We choose to use $T = 4$. For this study we are using publicly available data. As studied company we have chosen one of the main producer of electric energy in Czech republic. This company owns power plants which runs on coal, nuclear power, wind, photovoltaic energy and water and will be indexed based on table 4.1.

energy type	index
coal	1
nuclear	2
hydro	3
photovoltaic	4
wind	5

Table 4.1: Power plants indices.

We are using the sum of maximal production of instances in each category as our maximum allowed production $\boldsymbol{w} = (w_1, \ldots, w_5)^T$ in kWh/year unit. Notice that the highest production limits are in Czech republic for nuclear power and coal. On the contrary the capacity for using wind as the energy source is very limited.

energy source	production limit [kWh]
1	30856320000
2	37606140000
3	16802353224
4	1097132398
5	71004600

Table 4.2: Maximum annual production limit for each power plant type

The coefficients $h = (h_1, \ldots, h_5)$ of released CO_2 for one kWh in tons was collected from ipcc.ch from where we are using median estimation which can be found in table 4.3. It is noticeable that the volume of CO_2 when using coal is significantly higher than the other values. On the contrary, using nuclear power or wind energy is the most efficient way in terms of produced greenhouse gasses which needs to be covered with allowances.

As our selected company has its primary business in Czech republic, for estimation of market demand we use CR electric energy consumption time series from czso.cz to predict future demand. For estimation of margins for each power plant type, we use difference between time series of energy prices and LCOE estimates (for more detailed info see [5]). As the margins are time dependent, we

energy source	CO_2 production [tCO_2/kWh]
1	0.0082
2	0.00012
3	0.00024
4	0.00048
5	0.00012

Table 4.3: Median of tons of CO_2 produced by each power plant type for 1 kWh.

estimate the future values which are in table 4.4 and are considered as deterministic. In general, the margins on nuclear energy are the lowest and those on wind energy are the highest. Also notice that over the course of time, the margins are predicted to be lower.

energy source	time 0	time 1	time 2	time 3	time 4
1	0.098	0.092	0.088	0.084	0.081
2	0.053	0.050	0.047	0.044	0.042
3	0.129	0.123	0.119	0.115	0.113
4	0.093	0.088	0.084	0.080	0.077
5	0.144	0.139	0.134	0.131	0.128

Table 4.4: Predicted margins on energy.

4.1 Prices Simulations

The spot prices of EUA were downloaded from SENDECO2.com, where daily prices from year 2014 till the end of 2020 are available for modelling. When we look on the prices in 4.1 , we can see great growth since 2018. Because of unstable behavior, we decide to model monthly average returns rather than prices itselves. On this dataset we use one of the most common techniques for time series prediction to model future price development which is ARIMA models. ARIMA models in general take a form We have also tried to fit the GARCH(1,1) model but all coefficients are insignificant and so we have decided to use only ARIMA model. ARIMA(p,d,q) in general takes a form:

$$(1 - \sum_{i=1}^{p} \Phi_i L^i)(1 - L)^d X_t = \delta + (1 + \sum_{i=1}^{q} \theta_i L^i)\epsilon_t$$

where L is the lag operator,Φ_i are parameters of autoregressive part, θ_i are parameters of moving average part, ϵ_t is white noise and d is the level of differencing.

Based on Augmented Dickey-Fuller test which results in p-value $< 10^{-5}$ there is no need for differencing of the time series of spot returns. Based on ACF, PACF and AIC which can be seen in figure 4.2 we find out that the simple MA(1) model fits the best. The resulting model for spot returns r_t, from which the prices were calculated, is of form:

$$r_t = 0.028 + \epsilon_t + 0.3099\epsilon_{t-1}$$

After fitting this model, we use Monte Carlo simulations to generate multiple scenarios on which we use clustering to generate nodes for scenario tree.

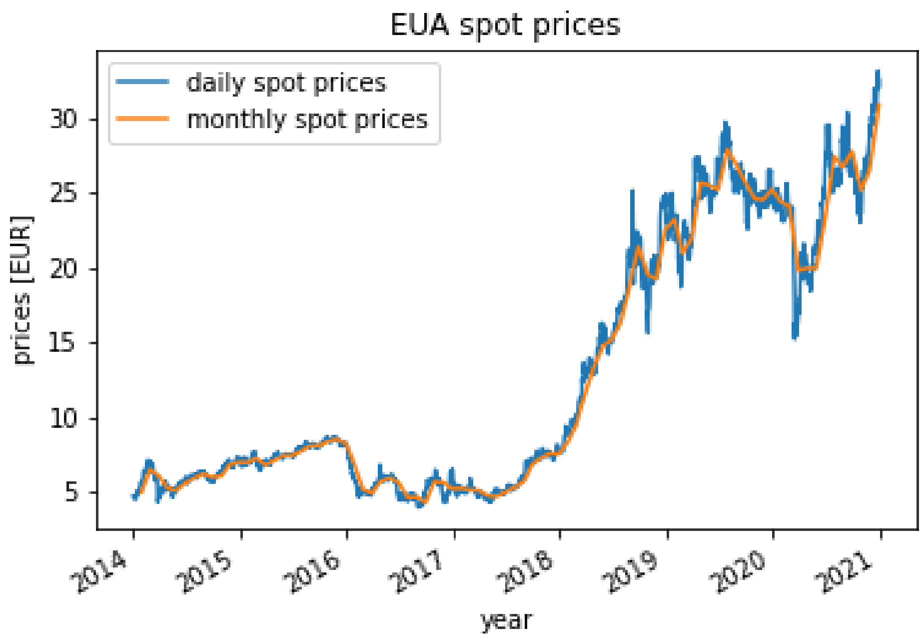

Figure 4.1: Development of EUA spot price over last 6 years.

As we are working with yearly time periods, we use this time series for calculating multiple scenarios of spot prices for the end of following years and generate scenario tree. As our root with known demand and prices we choose the beginning of the year 2021. From each node there is 10 leaves for following 4 years (i.e. 10 nodes for year 2022, 100 for 2023, 1000 for year 2024 and 10 000 for 2025). This give us scenario tree with 10^4 = 10 000 scenarios.The resulting tree can be seen in figure 4.3. One can notice that the predicted prices are volatile which is in compliance with the observed time series.

In table 4.5 we can notice that the mean value is slightly increasing over time but also the volatility grows so the minimal values is actually decreasing with each following time period. The median is not growing and neither decreasing.

For futures prices prediction we use data from Barchart.com with cost-of-carry model in the form of $f_t^\tau = \exp^{a(\tau-t)+\epsilon_t} s_t$ where a is real parameter and $\epsilon_1, \epsilon_2, \ldots$ are centered mutually independent random variables. For easier estimation of the parameter, we can use log-future-spot spread where $y_{t,\tau} = \log(f_t^\tau) - log(s_t)$ which follow linear regression model $\frac{y_{t,s}}{s-t} = a + \nu$ where ν is white noise. After fitting we get the value of $a = 0.00000779$ and for simplification, we assume $\nu = 0$. Then we use scenario tree 4.3 and compute with this equality the future prices. Their descriptive statistics can be found in table 4.6. Notice also that we want

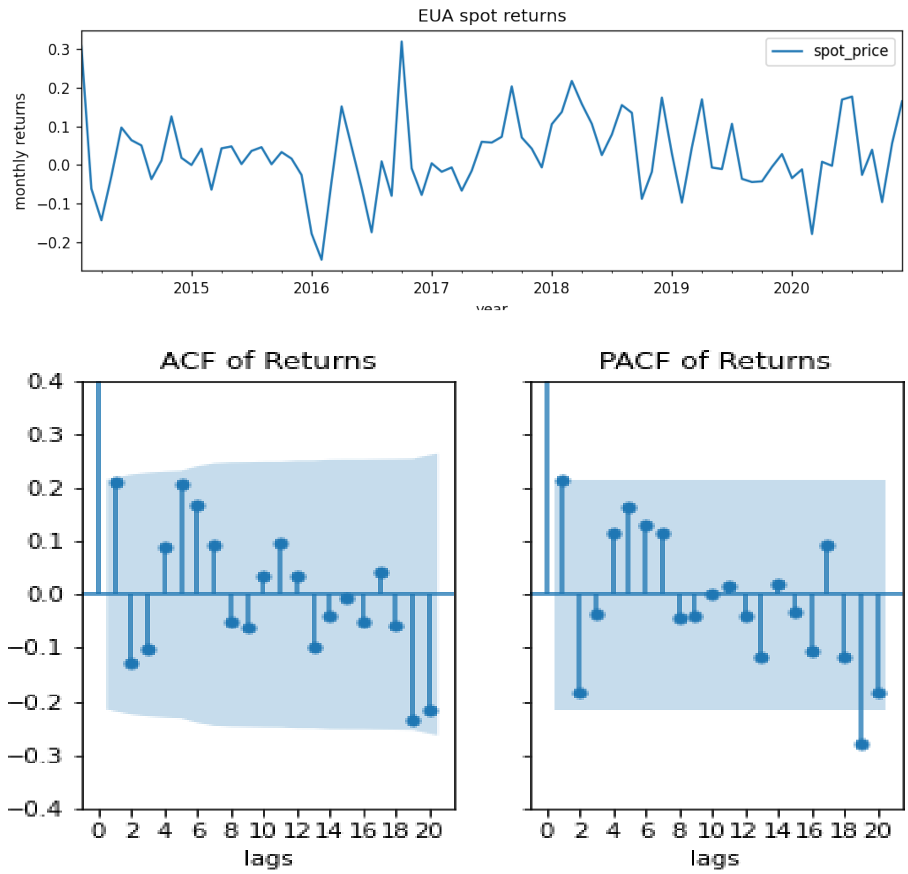

Figure 4.2: spot returns, ACF and PACF

time period	mean	std	min	25%	50%	75%	max
0	32.54	0.00	32.54	32.54	32.54	32.54	32.54
1	32.13	3.33	26.87	29.70	31.81	34.19	38.84
2	33.03	4.62	22.38	29.81	32.83	35.85	45.11
3	33.96	5.89	20.00	29.73	33.59	37.70	53.14
4	34.93	6.95	15.90	29.96	34.32	39.21	66.47

Table 4.5: Mean, standard deviation and quantiles of predicted spot prices for each time period.

to consider only those futures that have an expiration date as the last considered time period.

4.2 Market Demand Simulation

Lastly, we need to create a model for demand prediction. We consider yearly data for electricity demand in the Czech republic as the closest approximation as the considered company majority market is in the Czech republic. The time series which can be seen in figure 4.4 we would say that the series is not stationary.

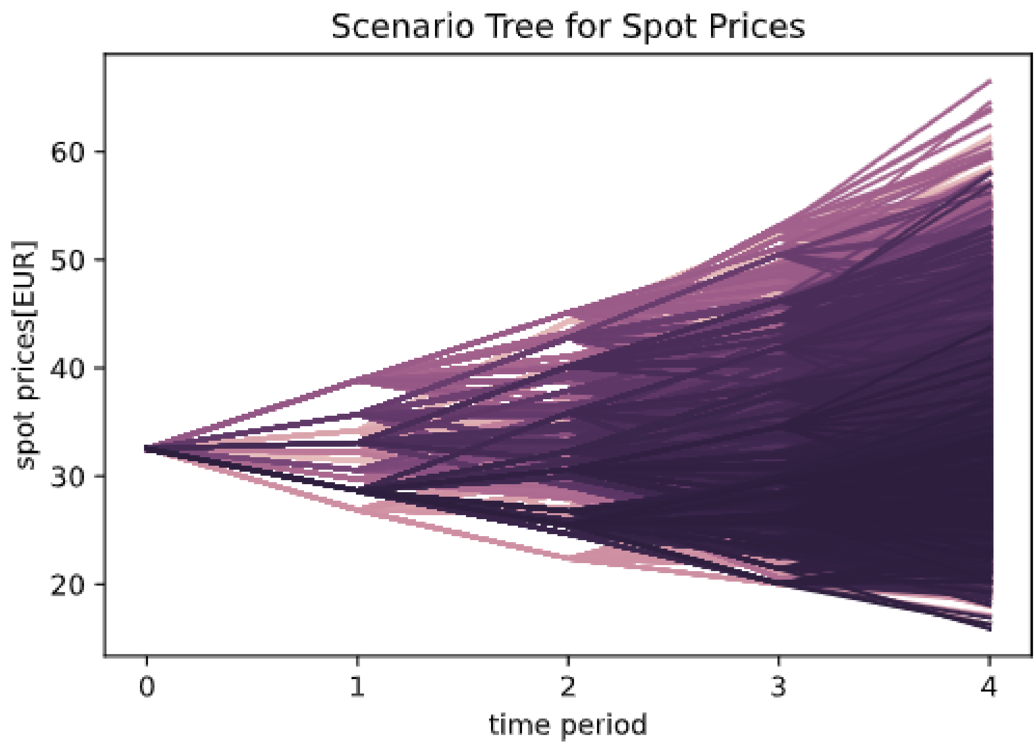

Figure 4.3: Predicted scenarios for EUA spot prices.

time	maturity time	mean	std	min	25%	50%	75%	max
0	1	32.63	0.00	32.63	32.63	32.63	32.63	32.63
0	2	32.73	0.00	32.73	32.73	32.73	32.73	32.73
0	3	32.82	0.00	32.82	32.82	32.82	32.82	32.82
0	4	32.90	0.00	32.91	32.91	32.91	32.91	32.91
1	2	33.61	5.49	18.52	29.79	33.07	36.54	64.12
1	3	33.71	5.51	18.57	29.87	33.17	36.65	64.30
1	4	33.80	5.52	18.62	29.96	33.26	36.75	64.48
2	3	33.60	5.49	16.36	29.79	33.10	36.57	60.76
2	4	33.70	5.51	16.40	29.87	33.19	36.67	60.93
3	4	33.60	5.46	17.23	29.79	33.13	36.54	62.56

Table 4.6: Mean, standard deviation and quantiles
of predicted futures prices for each time period and maturity date.

Similar to the spot returns investigation, we use the augmented Dickey-Fuller test to find out that the recommended differencing is 2, then we use ACF and PACF figures to estimate AR and MA levels which results in ARIMA(2,2,0) model. After fitting we get a model for demand d_t of the form:

$$d_t = 0.0485d_{t-2} - 0.0365d_{t-4} + \epsilon_t$$

Again, we use this model to simulate monte carlo process and clustering to create a prediction for each node of decision tree. We also try to investigate correlation between the demand and yearly returns on emission allowances. We find out that the correlation is 0.09 and so we consider the series as independent

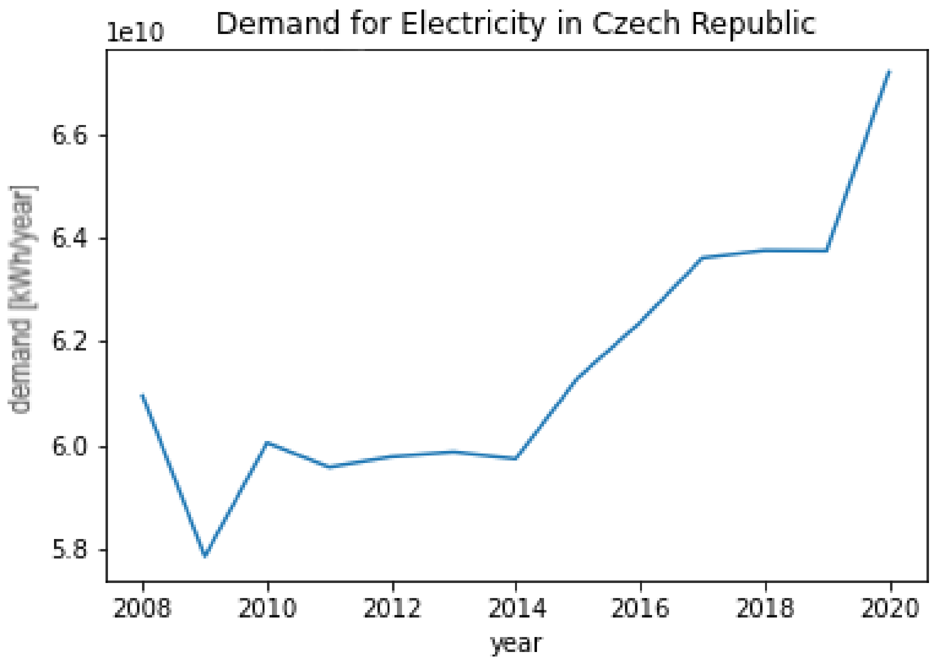

Figure 4.4: Development of electricity demand over last 12 years.

when building the scenario tree.

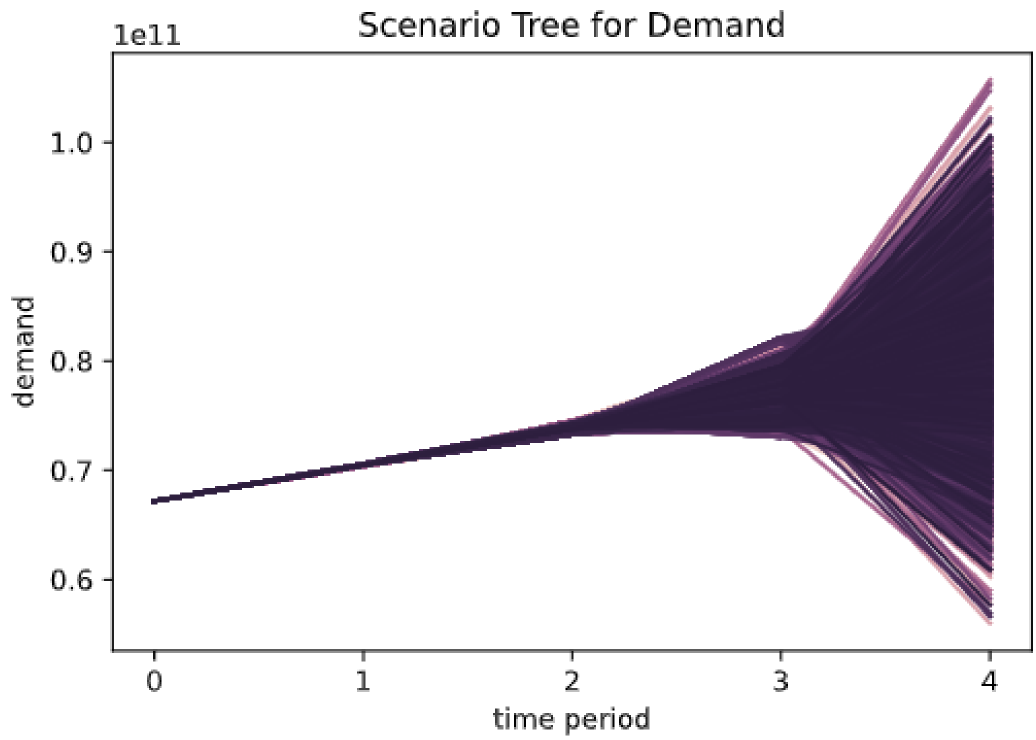

Figure 4.5: Predicted scenarios for market power demand.

In figure 4.5 we can see that the scenarios are less volatile compare to prices some greater volatility is observed for the last time period which again is in compliance with the observed values 4.4. We observe growth in mean value, in the maximum value and in median. Notice also, that the minimum value grows except for last time period where the scenario tree include possibility of downfall.

time	mean	std	min	25%	50%	75%	max
0	67.19	0.00	67.19	67.19	67.19	67.19	67.19
1	70.45	0.05	70.38	70.41	70.45	70.49	70.56
2	73.86	0.28	73.25	73.63	73.86	74.10	74.51
3	77.27	1.53	73.05	76.18	77.29	78.35	82.14
4	80.69	6.78	55.97	76.13	80.64	85.27	105.69

Table 4.7: Mean, standard deviation and quantiles of predicted market demand in [kWh/year 10^{-9}] for each time period.

As the last step we randomly pair the generated prices with demand to get the final scenario tree where each node consists of generated values for demand, price of spots, and futures for next time periods which we are considering.

5. Results

In previous chapters, we have developed multiple models and introduced data that we are working with. Here we would like to compare results for each of them. We want to compare the decisions and final wealth on different risk aversion levels (i.e. different values of parameter λ) in the base model. After this investigation, we want to also compare other models with this one. We use GAMS Python API for running models in Jupyter notebooks. As we are dealing with the count of bought allowances our models are mixed=integer problems. The base model is a linear mixed-integer problem as well as the chance constraint one and the second order stochastic dominance one. They were solved with CPLEX solver. The last problem is nonlinear problem and so it resulted in mixed=integer nonlinear programming. This was solved with KNITRO solver. Our models have a common parameter ρ which is the interest rate. In all models, we choose to use it with value $\rho = 0.03$. All models which use the conditional value at risk were run with $\alpha = 0.05$. In table 5.1 you can see the approximate time to find the optimal solution (note that the second order stochastic dominance problem and utility problem was solved on reduced scenario tree). Note also, that the run time for base model was highly dependent on the λ parameter choice, the higher the λ the longer run time and so in table 5.1 is provided time interval.

model	run time [h]
base model	0.1-0.5
chance constraint model	0.4
second order stoch. dominance model	0.2
utility model	3.1

Table 5.1: Running time to find optimal solution for each model

5.1 Base Model Dependency on Risk Aversion

In the base model, the risk aversion is controlled with the parameter $\lambda \in [0, 1]$ in the equation 3.12. Remember that the risk aversion in the base model is modeled by CVAR_α on random losses $\{-z_t\}_{t=1}^T$ which are given as the sum of expenses on allowances, interests from previous periods (either as loan or profit) and profit from energy sales.We generate the results of base model for $\lambda \in \{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9\}$. With higher values of λ, the company is more risk averse.

First we look into which power sources are preferable in each time period. For better visibility we choose to use logarithmic scale of produced energy for each year in graphs 5.1, 5.2, 5.3, 5.4 and 5.5. We can notice that over the course of time, for all values of λ the model advise to use the full capacity w for all power plants types except for the coal ones. This is due to the controversial behaviour as the margins on coals are high but the amount of CO_2 produced is also much higher as we saw in the table 4.3. At time 0 the higher the λ the more we should use this power source. As we can see from figure 5.1 the company starts to use coal from level $\lambda > 0.2$. On the contrary for the other time periods, for $\lambda = 0.1$ we always choose to use coal as well at least in some scenarios and we choose to use it as much as the demand limitation for the total produced energy allows us for $\lambda = 0.9$. In the figure 5.5 we can see the sum for all decision periods of average values. Here we can get to the conclusion that for levels of risk aversity $\lambda \in \{0.1, 0.2\}$ the company is under some scenarios willing to use more coal power generators than for $\lambda \in \{0.3, \ldots, 0.7\}$ and that if the company is very risk averse than under majority of scenarios, it is willing to use this source which gives it the certain high margin for one unit of produced energy.

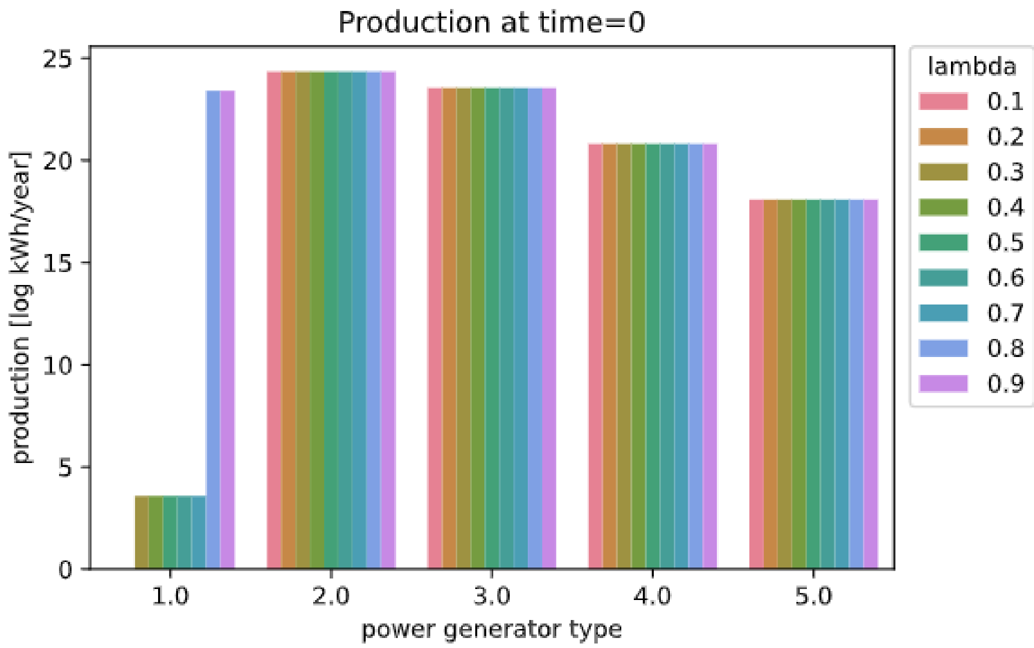

Figure 5.1: Produced energy volume on different levels of risk aversion at time zero

We should also check the behaviour in the terms of bought futures and spots over time based on how risk averse the investor is. In figure 5.6 one can notice that under all values of λ except for $\lambda = 0.1$, the company always decides to bought vast majority of its allowances at time $t = 0$. This is probably due to its concerns about growing spot prices (see table 4.5). In the cases of $\lambda = 0.1$ and $\lambda = 0.2$, on average, it rather invests to futures which we can see in figures 5.7. In the first figure, we see that the company buy most of future allowances at time 0, 2 and 3 for both $\lambda = 0.1$ and $\lambda = 0.2$ and from the second figure, we can see that most of this allowances are with maturity at time period 3 and 4. There we can also notice that for any other values of λ it choose not to bought futures. It is most likely because of the amount of bought spot allowances in the first time period which will cover also the following time periods and so the company is certain that without the allowances price development, it is going to have enough to cover the production. We should also remind that in cases of $\lambda \in [0.3,\ 0.7]$ the company is not using the coal as the energy source and so it is going to need much less allowances than in other cases as the coal power plants produce much more

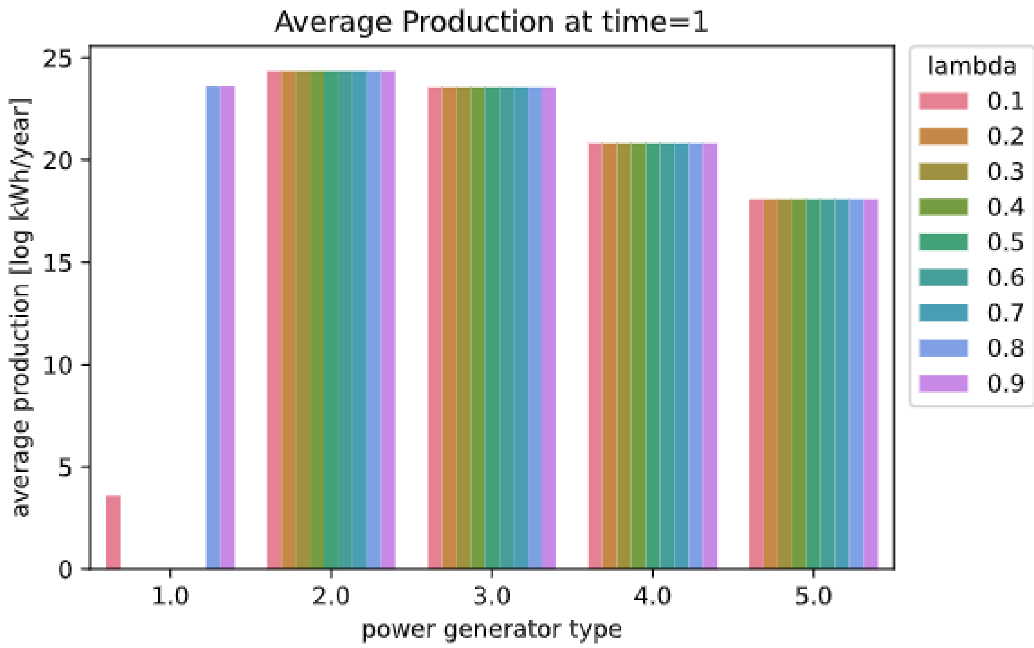

Figure 5.2: Produced energy volume on different levels of risk aversion at time one.

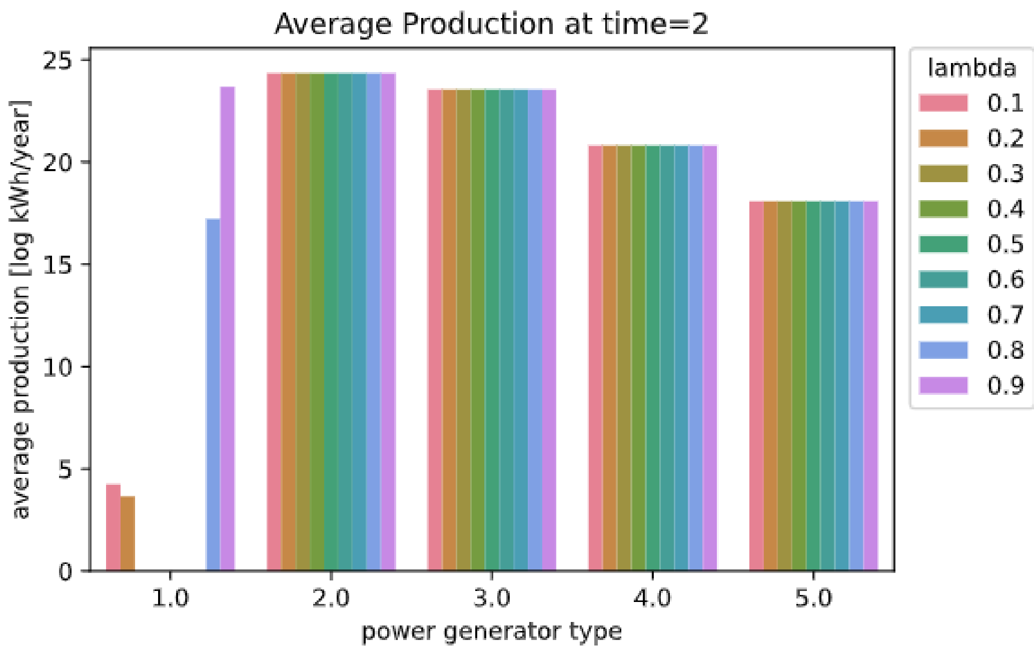

Figure 5.3: Produced energy volume on different levels of risk aversion at time two.

CO_2 than other sources (see table 4.3). That is noticeable from the total amount of allowances which the company bought over the full 4 years. We can see that with only very low use of coal as energy source, it is enough for the company to buy spot allowances in the very beginning to have enough for last 4 years.

Finally we want to know the average value of total wealth gained over all time

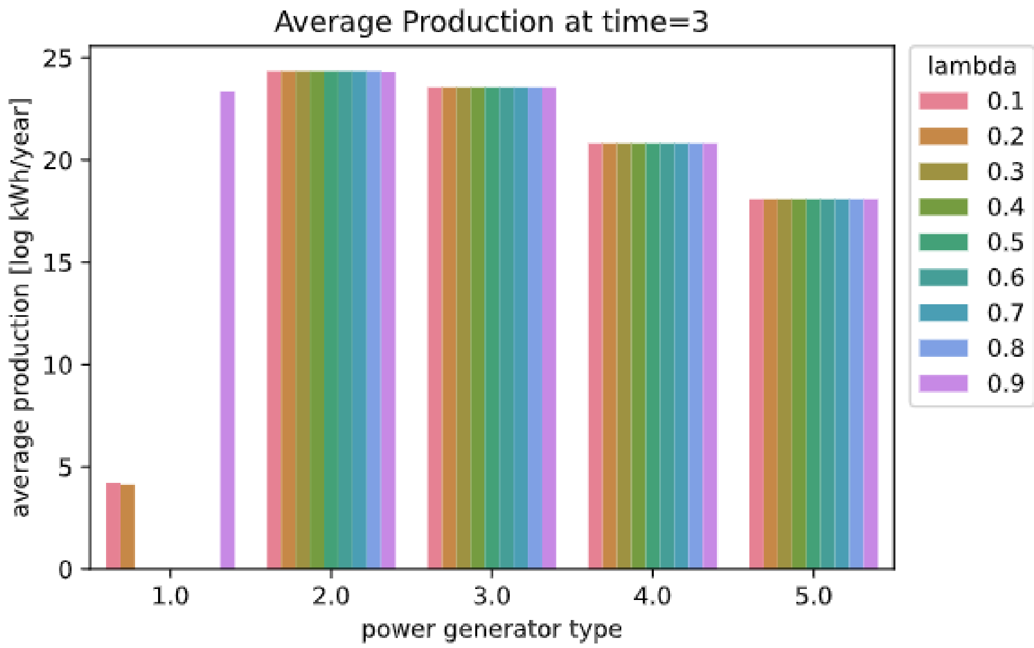

Figure 5.4: Produced Energy Volume on Different Levels of Risk Aversion at Time Three.

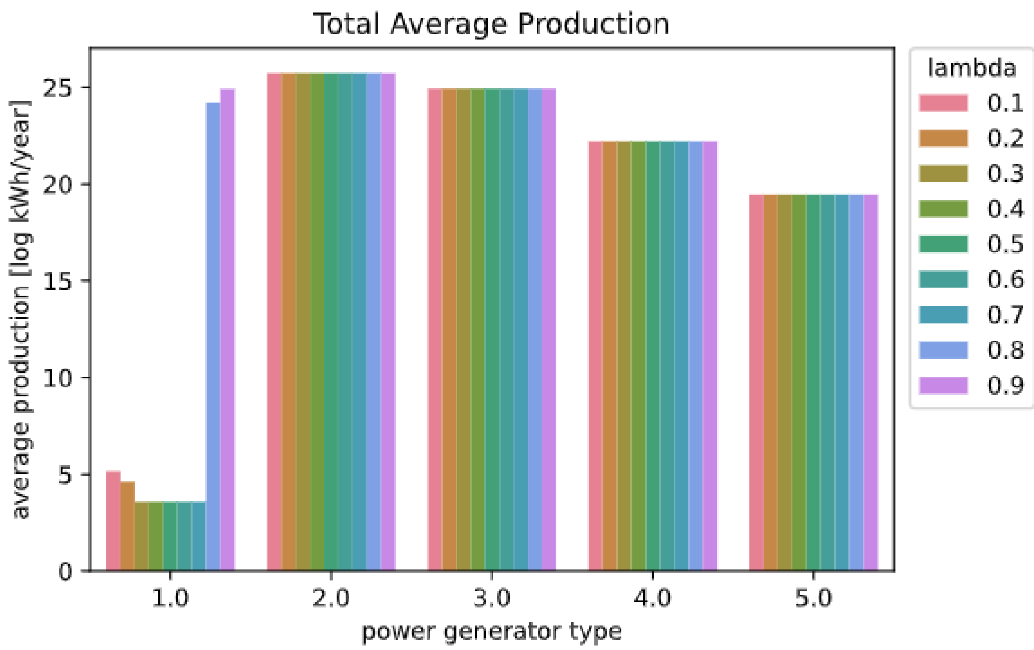

Figure 5.5: Sum of Produced Energy Volume on Different Levels of Risk Aversion over time.

periods. On the first sight in figure 5.8 we can see that the behaviour in terms of cashflows follows what we mentioned above. The company which has very low risk aversion wants to have immediate profit and so for all time periods it decides to behave in a way that on average it has positive cashflows. As the company is more risk averse, it prefers to ensure that it is going to have enough allowances to cover its planned production and so it is willing to go into red numbers which

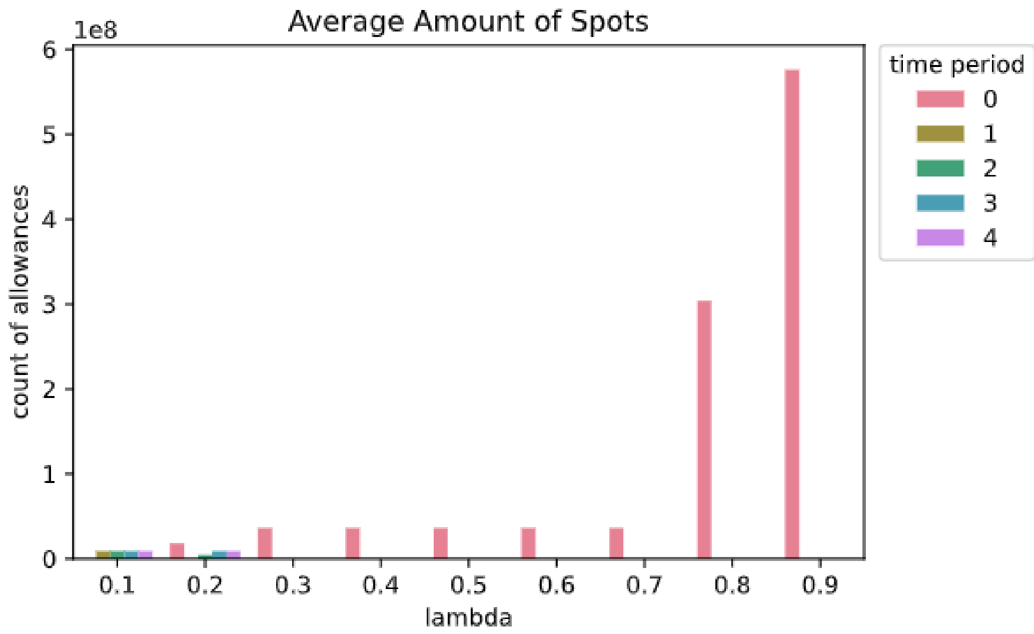

Figure 5.6: Average amount of bought spot allowances based on time period.

might be covered with the loan for very low interest rate $\rho = 0.03$ not to have to risk price development in following years. In conclusion, its expected total wealth can be seen in figure 5.9. From which we can see that for $\lambda < 0.8$ the average wealth is the same and the only difference is whether the company decides to rather not to produce coal and not to have to risk the price development or it takes the risks and on average it seems to be convenient strategy in terms of the final wealth. For much more risk averse companies with $\lambda \geq 0.8$ the average final wealth is lower because the company choose not to take the risk of price volatility and also the uncertainty of market demand development.

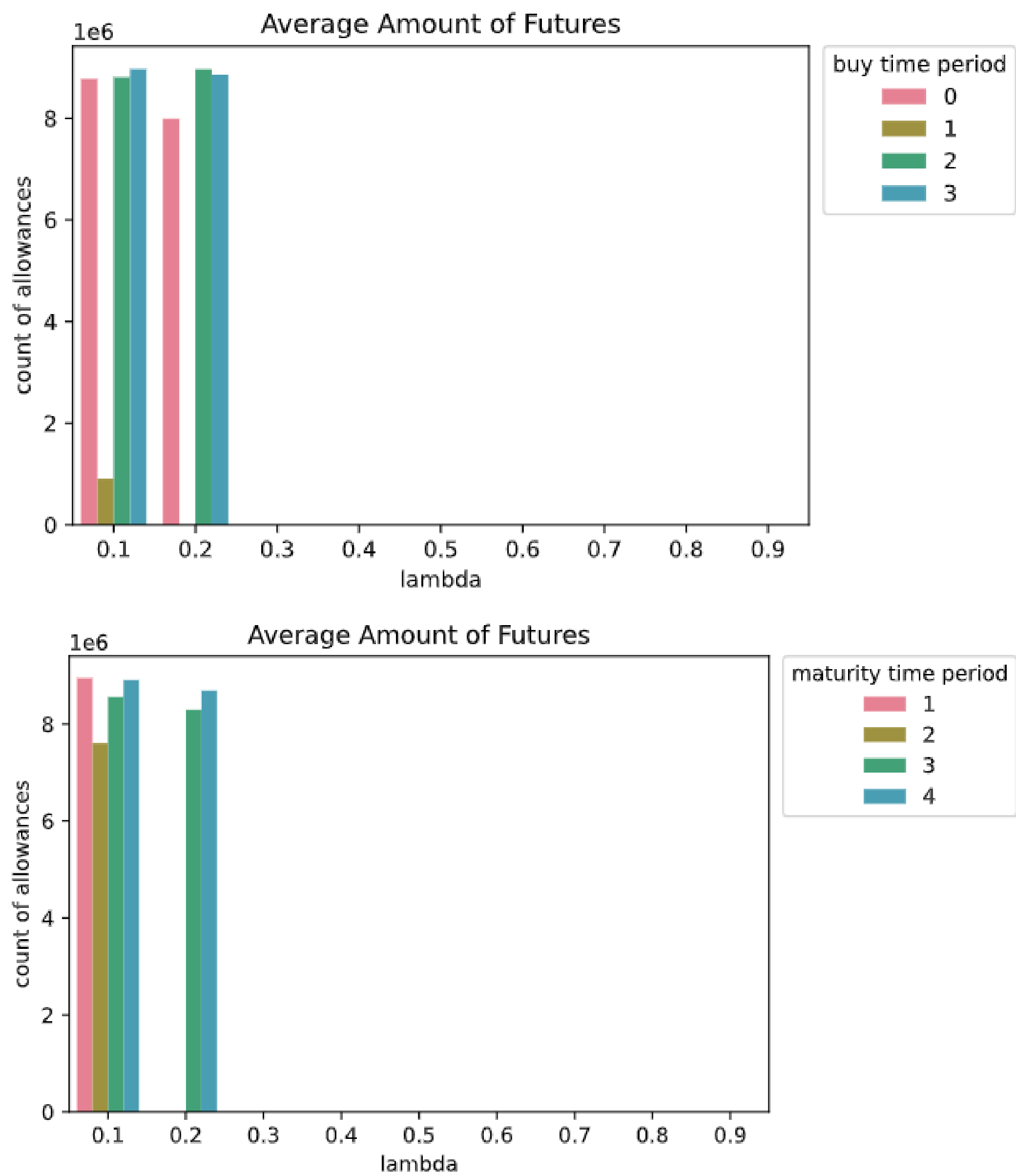

Figure 5.7: Average amount of bought futures allowances based on time period.

5.2 Chance Constraint Model

In this section we want to compare the difference in behaviour of base model and the model with chance constraint for different values of its parameter ϵ. These two models have the same objective function and the only differences are in constraints. The first difference is in equation 3.23 where we suppose that if some produced energy is not sold due to low demand, it is the energy from coal power generator. So we need to subtract the margins on unsold part of produced energy and also the costs for production γ_1. As we previously mentioned, these costs were taken from [5].

Another difference is in the terms of the chance constraint from equation 3.15

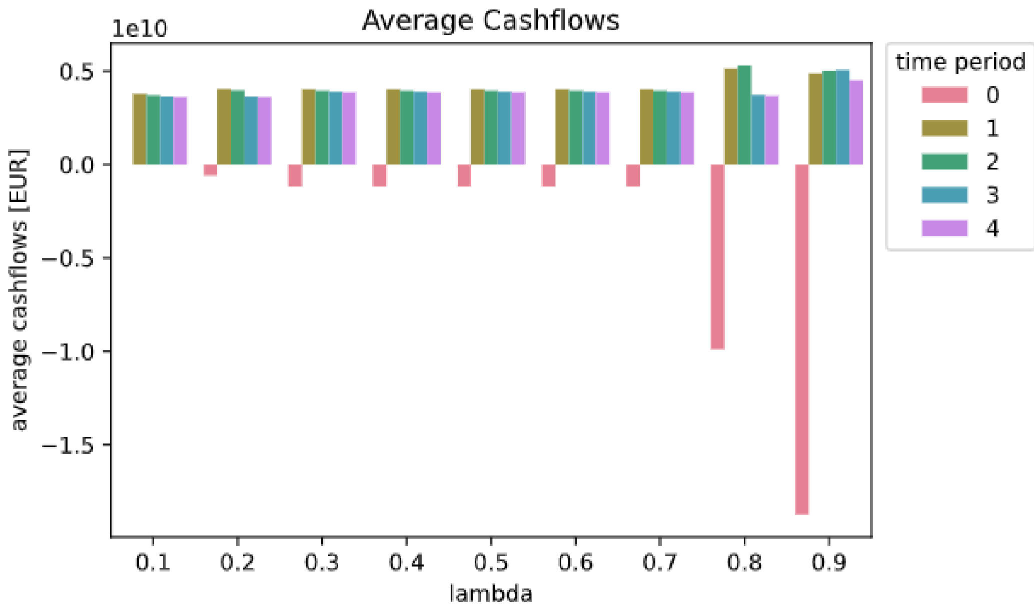

Figure 5.8: Average amount of cashflows over time

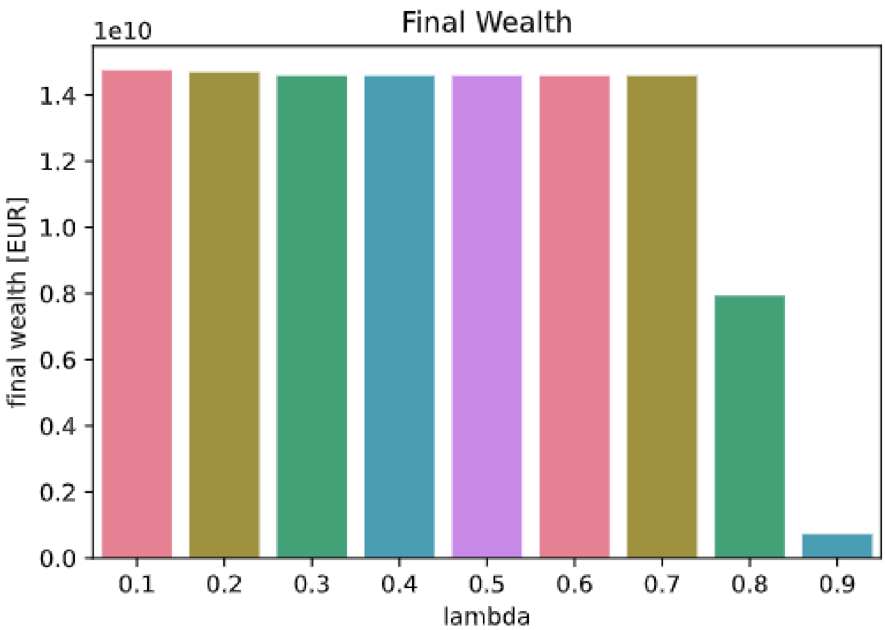

Figure 5.9: Average amount of cashflows in total.

which regulates the count of scenarios in which the demand constraint (i.e. the constraint which ensures that the production cannot be higher than all possible future demand values in following time period) might be violated. The allowed percentage of violated constrained is regulated with parameter ϵ.

We test this model with parameter $\lambda = 0.9$. We test to run the chance constrained model with the parameter $\epsilon \in \{0.1, 0.3, 0.5, 0.7\}$ to see the results in comparison to the base model. Surprisingly, we found out that the effect of limitation on demand has no effect on cash flows for $\lambda = 0.9$. Here you can see

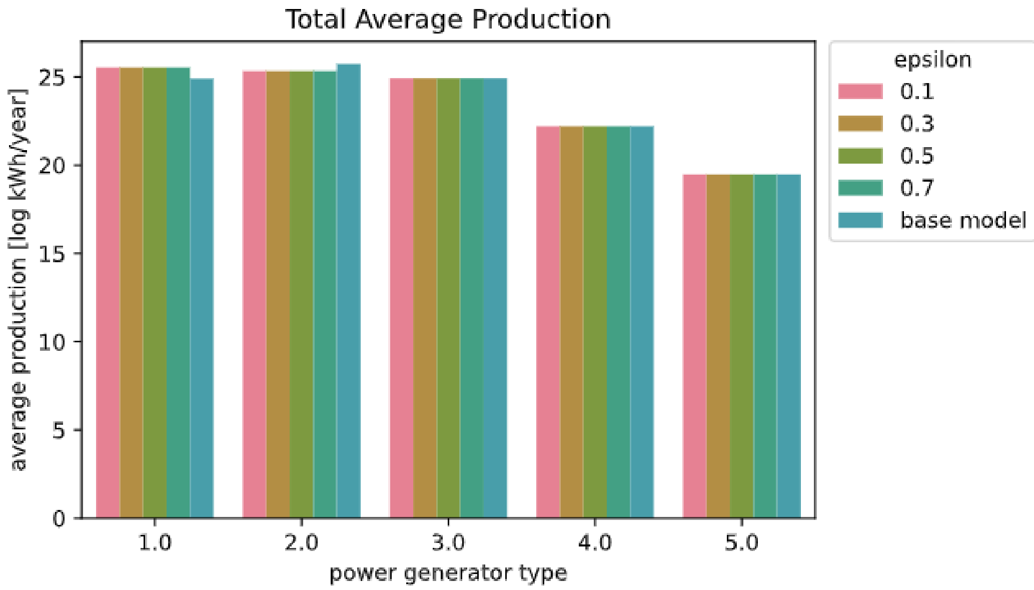

Figure 5.10: Sum of Produced Energy Volume on Different Levels of Risk Aversion over time.

several figures. In the figure 5.10 one can see that the only difference between base model and the model with chance constraint is in terms of production in coal power plants which in base model is slightly lower and in nuclear power which is slightly higher. This is natural consequence of allowing to produce more as the margins on coals are higher. The same results for all tested values of ϵ can be made also observed in figure 5.12.

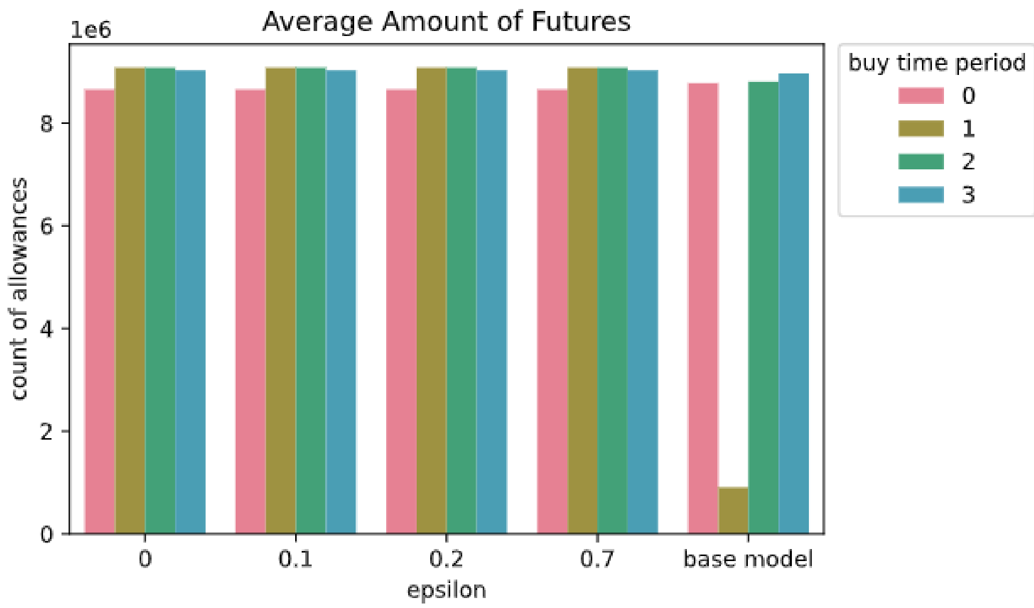

Figure 5.11: Average amount of bought futures allowances based on time period.

In figure 5.11 we can see again that the behaviour is the same independently of setting of parameter ϵ.

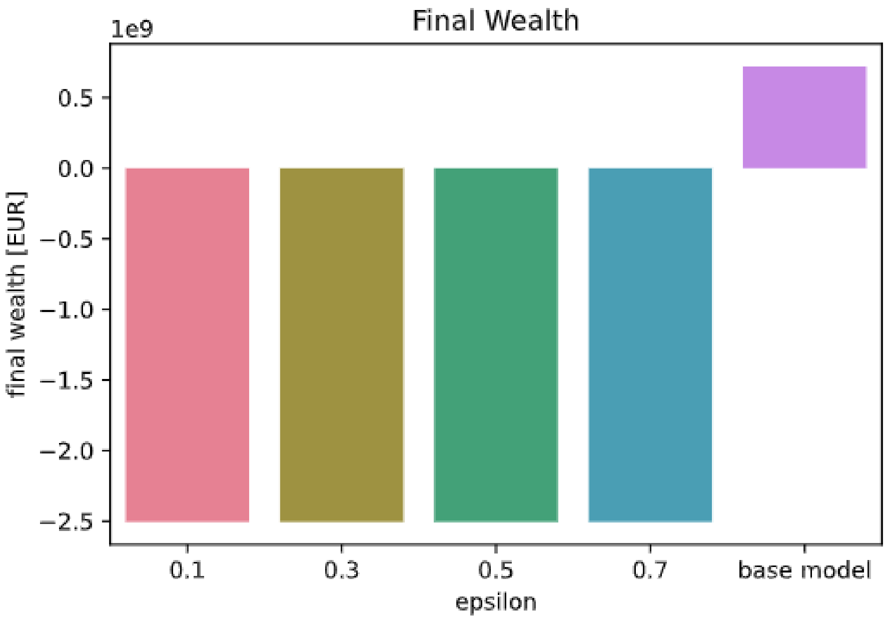

Figure 5.12: Average amount of cashflows in total.

Based on these results we can assume that the demand is very stable and does not play much role in company´s decisions.

5.3 Second Order Stochastic Dominance Model

In this model instead of using conditional value at risk in objective function we want to rather have the results stochastically dominate some benchmark portfolio and the objective function is defined only as the expected value of losses over time. The approach to generate this portfolio has been already explained in chapter 3. For running this model we need to use the tree dimensionality reduction. We use nodal clustering method and for each time period, we want to end up with the half of the original nodes. This means that we get resulting 5000 scenarios. As the clustering method, we use constrained K-means clustering which was introduced in formula 2.8 thanks to which we can easily reach to reduced scenario tree that is still balanced with univariate distribution of prices and demand values. Distribution of prices and demand for this reduced scenario tree can be seen in tables 5.2, 5.3 and 5.4. Comparing to the full scenario tree, we can see that the descriptive statistics for distribution of prices (both spot and futures which was again calculated using cost-of-carry model) is almost the same. In case of demand, we can notice even lower values of volatility.

time period	mean	std	min	25%	50%	75%	max
0	32.54	0.00	32.54	32.54	32.54	32.54	32.54
1	32.13	3.21	27.77	30.14	31.81	33.65	37.26
2	33.03	4.55	24.42	29.76	32.88	35.95	44.21
3	33.96	5.83	20.89	29.76	33.54	37.54	52.56
4	34.93	6.89	17.59	30.02	34.41	39.24	64.43

Table 5.2: Mean, standard deviation and quantiles of predicted spot prices in reduced scenario tree.

time	maturity time	mean	std	min	25%	50%	75%	max
0	1	32.63	0.00	32.63	32.63	32.63	32.63	32.63
0	2	32.73	0.00	32.73	32.73	32.73	32.73	32.73
0	3	32.82	0.00	32.82	32.82	32.82	32.82	32.82
0	4	32.91	0.00	32.91	32.91	32.91	32.91	32.91
1	2	33.64	5.42	19.0	30.23	33.15	37.22	61.26
1	3	33.74	5.44	19.05	30.31	33.25	37.33	61.44
1	4	33.84	5.45	19.1	30.4	33.34	37.43	61.61
2	3	33.62	5.44	17.64	30.10	33.15	37.22	60.12
2	4	33.71	5.46	17.69	30.19	33.25	37.33	60.29
3	4	33.56	5.38	17.68	30.03	33.12	37.22	64.61

Table 5.3: Mean, standard deviation and quantiles of predicted futures prices in reduced scenario tree.

In this model we are using benchmark portfolio which should be stochastically dominated. As we mentioned in section 3.3, the portfolio was generated to reach the overall minimum demand in our scenario tree and the distribution of power generators in it is given as ratio of maximal production for that specific power

time	mean	std	min	25%	50%	75%	max
0	67.19	0.00	67.19	67.19	67.19	67.19	67.19
1	70.45	0.03	70.42	70.44	70.45	70.46	70.51
2	73.86	0.22	73.38	73.70	73.85	74.19	74.32
3	77.27	1.12	74.19	76.45	77.23	78.05	80.43
4	80.69	4.87	63.02	77.37	80.66	83.94	97.16

Table 5.4: Mean, standard deviation and quantiles
of predicted market demand in [TWh/year] in reduced scenario tree.

plant over the total production. At each time, this production needs to be fully covered with allowances. This results in productions which can be seem in table 5.5. Notice that the production of energy using coal is very high in comparison to base model results.

energy source	production
1	23986780864.84
2	30642087128.32
3	14265756258.16
4	941771319.82
5	51771966.62

Table 5.5: Production in each time period for benchmark portfolio.

The expected cash flows under each scenarios of spot prices are given in the table 5.6. See that the strategy is quite similar to highly risk averse company in base strategy as it decides to spend significant amount at the beginning and profit later.

Firstly, we compare the total average energy produced by each type of power

time	cash flow
0	-7149653430.0
1	-3633728160.0
2	692723697.0
3	4510751240.0
4	472626978.0

Table 5.6: Cash flows for benchmark portfolio.

plants under the base strategy with $\lambda = 0.9$, the benchmark strategy and the dominating strategy. In figure 5.13 we can see that the benchmark model produce more energy using coal than the other two models. Stochastically dominating strategy produces the least energy from coal. In the other types of power sources, all models performs pretty much similar.

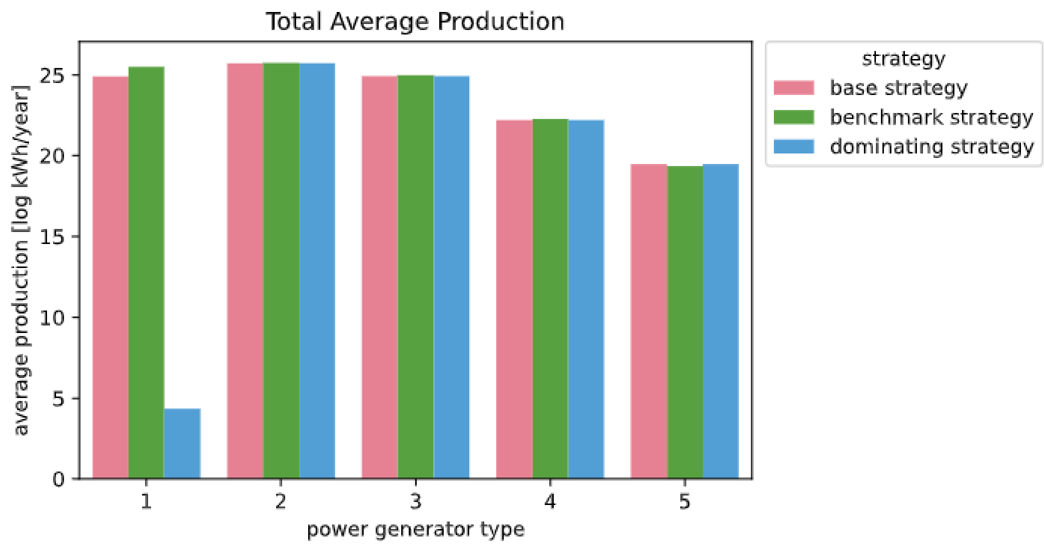

Figure 5.13: Average Energy produced based on model and energy source.

Dominating strategy is able to stochastically dominate the benchmark one thanks to using futures. As we can see in graph 5.14, it is the only strategy that uses them.

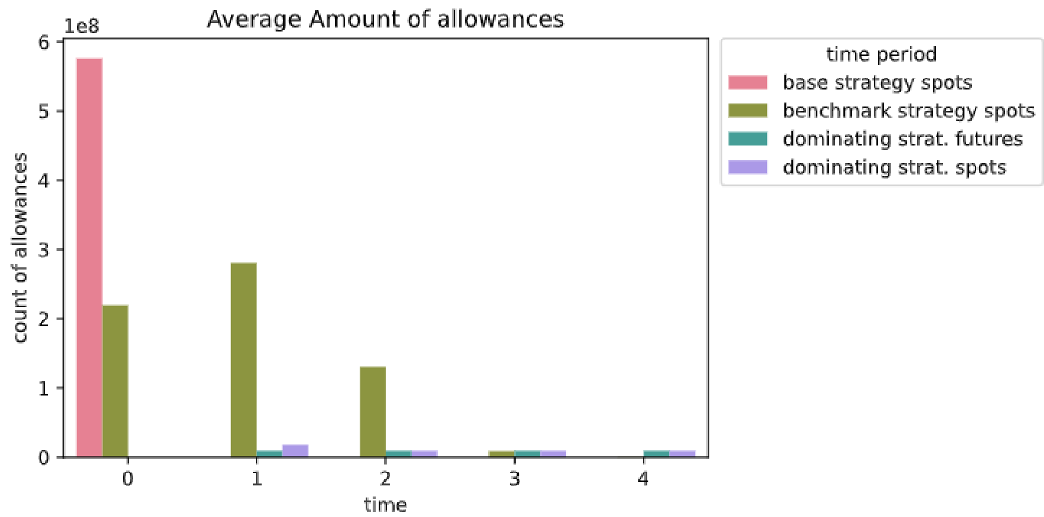

Figure 5.14: Allowances purchase for different strategies.

In terms of wealth, we can see in figure 5.16 that the dominating strategy keeps its cashflows stable over time in contrary to base strategy and benchmark which both are willing to go to red numbers at the beginning.

This behaviour results in significant difference between the dominating model and the other two models.

To sum it up, the dominating strategy takes advantage of avoiding using coal in most of the scenarios which results in lower need of allowances and it also takes

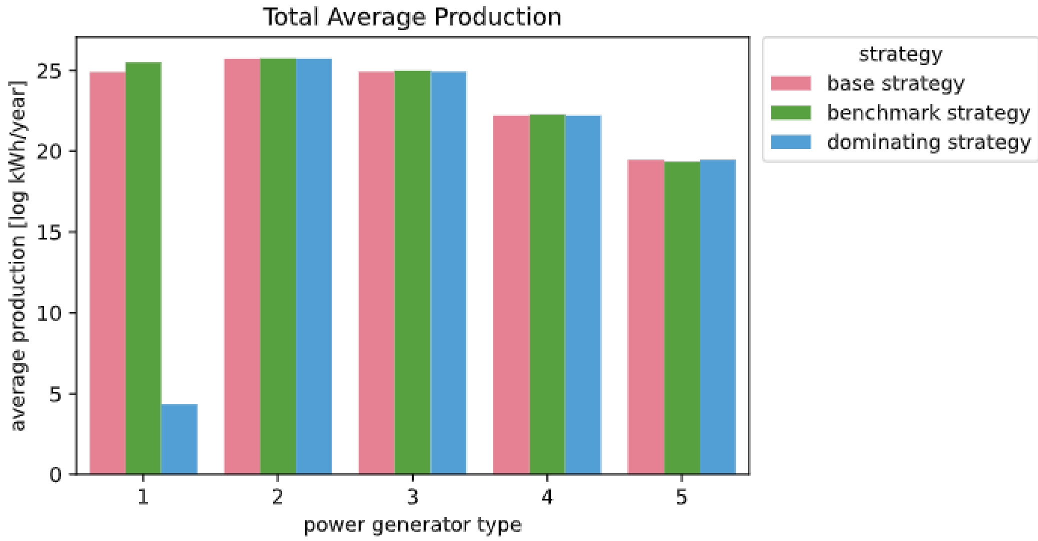

Figure 5.15: Average amount of cashflows over time

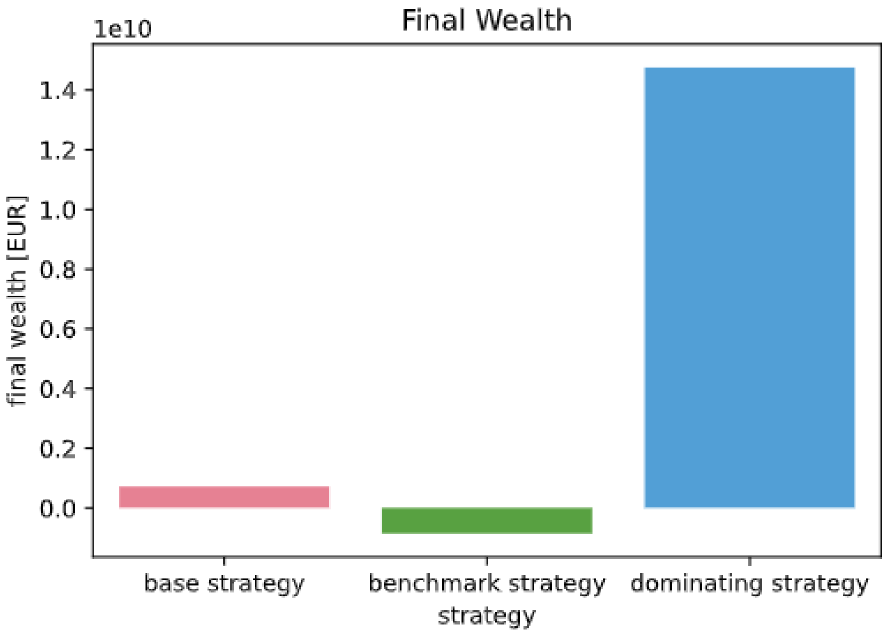

Figure 5.16: Average amount of cashflows in total.

advantage of using futures.

5.4 Model with Utility Function

In this section we want to evaluate the results for model where instead of mean-risk stochastic problem we are dealing with maximization of expected value of risk averse utility function which was stated in section 3.4. The model form was stated in the previous chapter. To evaluate this model, we use the same reduced scenario tree as for second order stochastic dominance model and so all details regarding this reduce tree can be found in section 5.3.

In this model, we can see quite different behaviour in terms of choice of power plants from which we want to produce the energy. In figure 5.17 we can notice that actually the company decides to invest into coal power plant in all considered time periods. Also notice that in comparison to base strategy this results into less using all other sources.

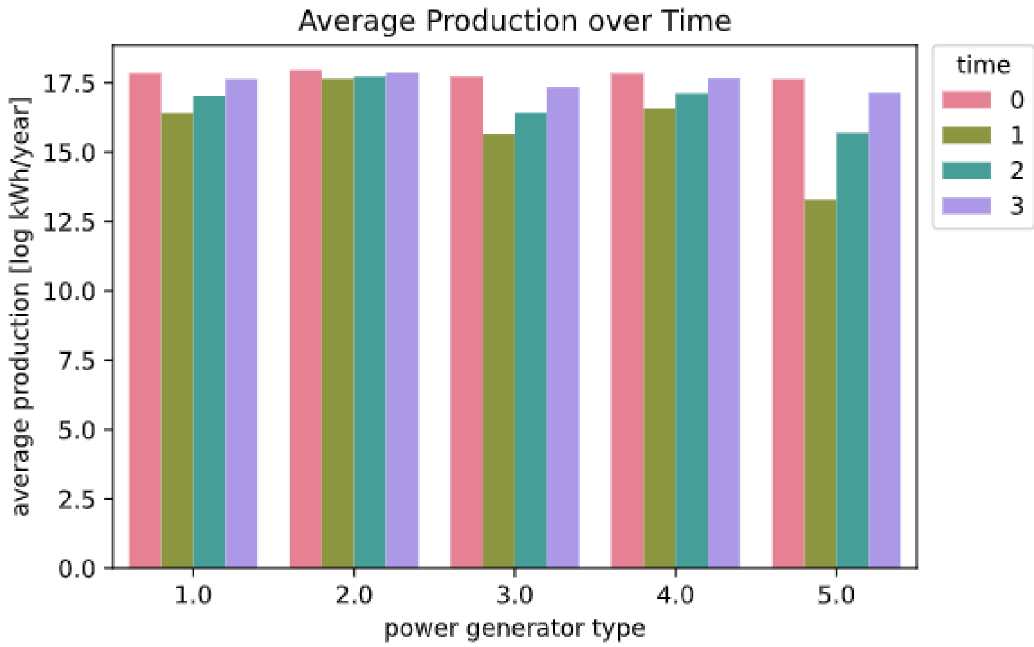

Figure 5.17: Average Energy produced based on energy source.

As we can see from table 4.3 using of coal will result into need of more allowances as it produces the highest amount of CO_2 for 1 kWh production. This we can observe in figures 5.18 and 5.19 where the company needs significantly more allowances both from spots and futures in comparison to all previous strategies. Notice also that the amount of spot allowances that are bought on average in the last time period is very high. The other risk averse strategies from previous models tried to avoid buying spots in the last time period as the volatility and the chance that the price will grow is too risky.

In figure 5.20 we can see that all average cash flows are in red numbers which is only confirmation of what we have mentioned above. To sum it up, it seems that this model is not that risk averse as the previous ones which results to take the risk in allowances price development and on average, it does not payoff.

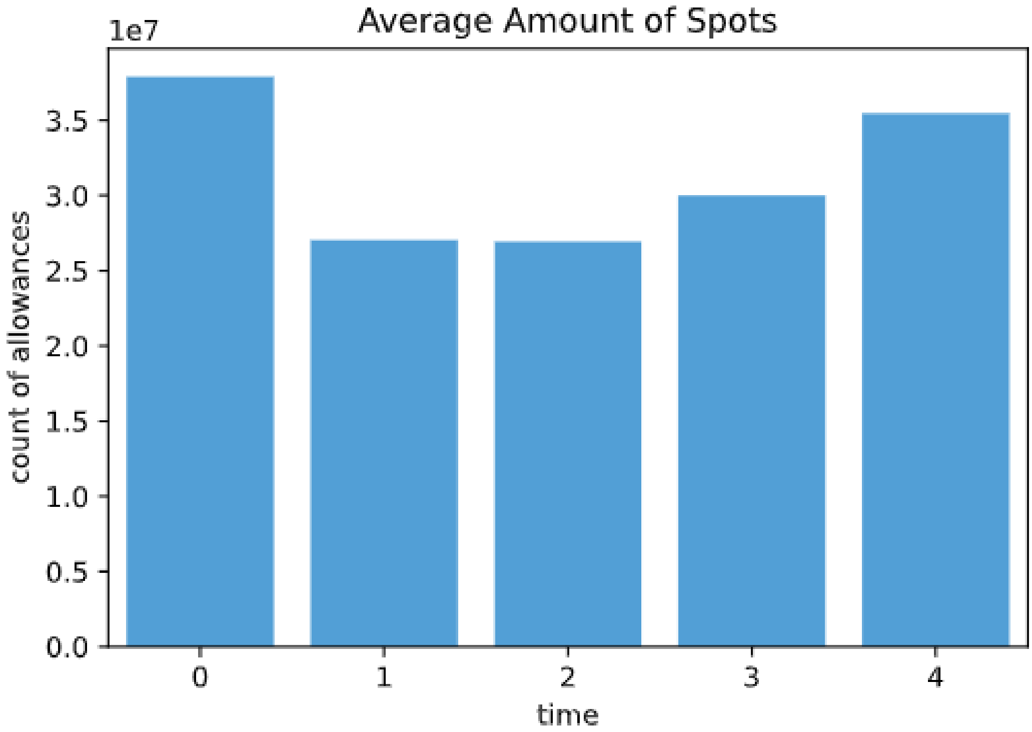

Figure 5.18: Spot Allowances purchased over time.

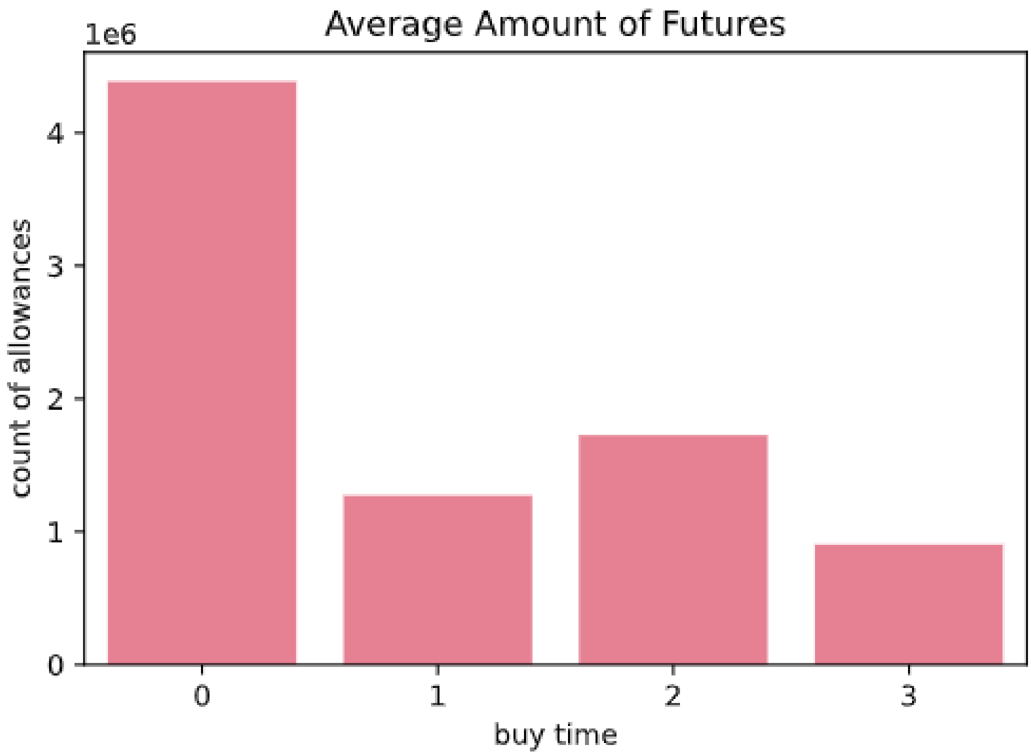

Figure 5.19: Average amount of bought futures allowances based on time period.

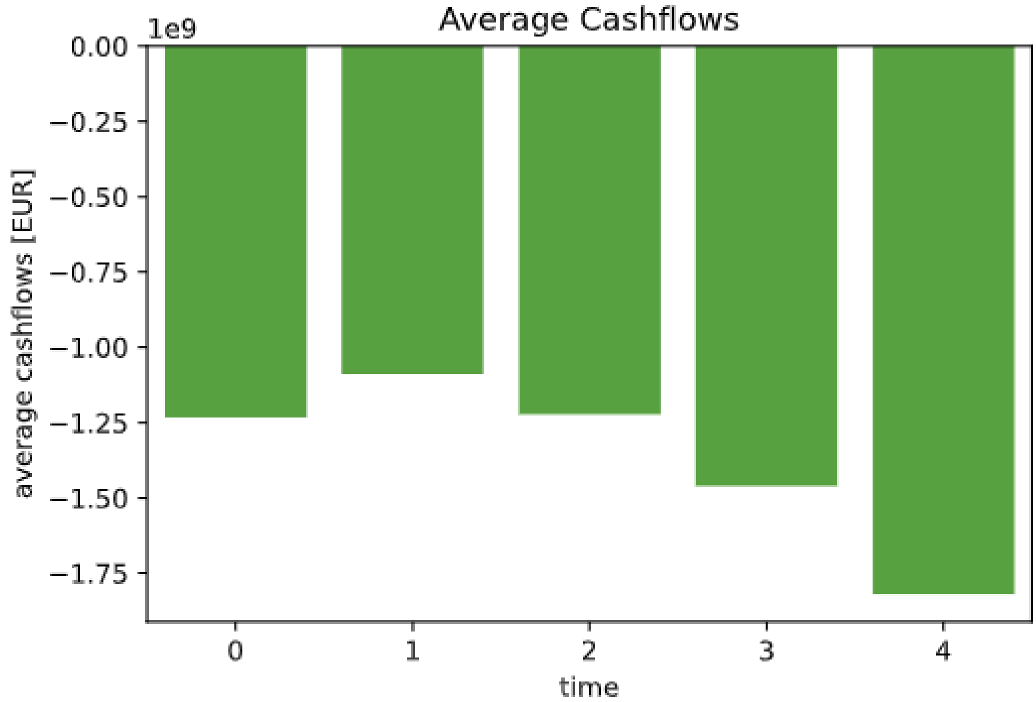

Figure 5.20: Average amount of cashflows over time

Conclusion

This thesis developed multiple models which reflect risk aversion of company and should protect the company from undesired loses while still generating sufficient profit.

We have discussed different approaches to modelling risk aversion, we showed how to generate scenario tree and how to reduce its dimensionality. We also explained how the market with allowances works and what are the current regulation for energy sector.We explained the multistage stochastic models and how to use them when working with scenarios.

Then several models were developed and tested in empirical study. The base model with CVAR limit has shown dependency on the risk aversion level and we found out that most changes are done in terms of using the coal as power source and in using futures. Another significant difference was in terms of time period where the company decides to buy allowances. The chance constraint model shows that when using simplification in terms of average margin on energy, the market demand does not play huge role as the chance constraint does not effect the results. There is possibly space for some further development as we can make some assumption about how much from each generator will be used and from which the energy will be left. For second order stochastic dominance model we had to use scenario reduction. Then we have compared results for similarly behaving base model, the benchmark model and our stochastically dominating model. Here again, further test might be run for different benchmark choices. We might also consider first order stochastic dominance instead of the second order stochastic dominance.

From this work, several other modifications might be derive. One might use different risk measure instead of CVAR, e.g. it might be used VAR or standard deviation or mean absolute deviation. Another modification might be in terms of using options as financial derivative. Several test regarding the scenario tree dimensions and differences between reduction techniques can also be nice enhancement of this work. The power utility function can be replaced with exponential one, discussed the choice of its parameter and the effect on results and compared with power utility function.

In conclusion, our suggestion is to use the base model because of this relatively short run time (when compared to the rest of tested models). and easy risk aversion control with parameter λ.

Bibliography

[1] Bradley P. Bennett K. Demiriz A. Constrained kmeans clustering. online, 2000. URL https://www.microsoft.com//en-us//research//wp-content//uploads//2016//02//tr-2000-65.pdf.

[2] Shapiro A. Dentcheva D. Ruszczynski A. *Lectures on Stochastic Programming : Modeling and Theory*. Society for Industrial and Applied Mathematics and the Mathematical Programming Society Philadelphia, 2009. ISBN 978-0-898716-87-0.

[3] Directorate-General for Communication. An official website of the european union. online, 2020. URL https://ec.europa.eu/clima/policies/ets_en.

[4] Luedtke J. Ahmed S. Nemhauser G. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, (122):247–272, 2010.

[5] LAZARD. Lazard's levelized cost of energy analysis. online, 2020. URL https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2020/.

[6] Kopa M. Chovanec P. A Second–order stochastic dominance portfolio efficiency measure. *Kybernetika*, 44:243–258, 2008.

[7] Dupačová J. Consigli G. Wallace S. Scenarios for multistage stochastic programs. *Annals of Operations Research*, 100:25–53, 2000.

[8] United Nations Climate Change. What is the kyoto protocol? online, 2020. URL https://unfccc.int/kyoto_protocol.

[9] Horejšová M. Vitali S. Kopa M. Moriggia V. Evaluation of scenario reduction algorithms with nested distance. *Computational Management Science*, 17: 241–275, 2020.

[10] Heitsch H. Römisch W. Scenario tree reduction for multistage stochastic programs. *Computational Management Science*, 6:117–133, 2009.

[11] Zapletal F. Šmíd M. Kopa M. Multi-stage emissions management of steel company. *Annals of Operations Research*, 292:735–751, 2020.

[12] Šmíd M. Zapletal F. Hančlová J. Which carbon derivatives are applicable in practice? a case study of a european steel company. *Kybernetika*, 53: 1071–1085, 2017.