



Norwegian University of
Science and Technology

Hedging Price Risk for an Electricity Producer with Multistage Distributionally Robust Optimization

Mikkel Johan Boye Ahlgren
Andreas Aalberg Huse

Industrial Economics and Technology Management

Submission date: June 2018

Supervisor: Sjur Westgaard, IØT

Co-supervisor: Asgeir Tomasgard, IØT

Norwegian University of Science and Technology

Department of Industrial Economics and Technology Management

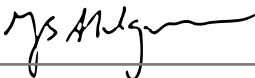
Preface

This thesis concludes our Master of Science in Industrial Economics and Technology Management at the Norwegian University of Science and Technology (NTNU). The thesis is written during the spring of 2018, intersecting between the fields of Applied Economics and Optimization, and Financial Engineering.

Writing on a field of study where NTNU has limited experience has been challenging, but highly motivating. We would like to thank our supervisors Sjur Westgaard and Asgeir Tomasgard at the Department of Industrial Economics and Technology Management for guidance and helpful insights throughout the process. In addition, we are truly grateful to TrønderEnergi AS with Gunnar Aronsen for providing data and insight on the decision problem. Finally, a thankful note is also directed to Professor Alois Pichler from Chemnitz University of Technology and Sara Séguin from Université du Québec à Chicoutimi for valuable inputs.

Mikkel Johan Boye Ahlgren and Andreas Aalberg Huse

Trondheim, June 2018





Abstract

In this thesis, we consider optimal hedging decisions for an electricity producer. In addition to account for uncertain prices and production, we let the underlying probability distribution in itself be subject to uncertainty. Distributional uncertainty is known as ambiguity, and is accounted for by applying a multistage distributionally robust optimization model. By modeling ambiguity as a so-called ambiguity set of possible probability distributions, the distributionally robust model finds the optimal hedging decision for the probability distribution in the ambiguity set that causes the most harm. We extend an existing framework for multistage distributionally robust optimization to incorporate risk aversion, using time consistent conditional value at risk as a risk measure. The input scenario tree is generated from a forecast fan with application of stochastic approximation.

We find that the hedging strategy from the distributionally robust model outperforms a stochastic model under the worst case distribution from the ambiguity set, while it suffers only a slight reduction in performance when the distribution is correctly estimated. The risk is therefore reduced by applying a distributionally robust hedging model under ambiguity. Backtesting the hedging strategy on historical data from 2014 to 2017 shows that the distributionally robust model outperforms both the strategy from a stochastic model and a strategy where hedging is absent, in terms of mean profits. Our findings therefore suggest that distributional uncertainty should be accounted for when developing optimal hedging strategies for electricity producers.

Sammendrag

I denne oppgaven studerer vi optimale sikringsstrategier for en kraftprodusent. I tillegg til å ta høyde for usikkerhet i priser og produksjon, tar vi høyde for at den underliggende sannsynlighetsfordelingen i seg selv kan være usikker. Slik fordelingsmessig usikkerhet omtales i litteraturen som *ambiguity*. Vi tar høyde for fordelingsmessig usikkerhet ved å formulere en *distributionally robust* optimeringsmodell, som modellerer den fordelingsmessige usikkerheten gjennom et såkalt *ambiguity set* av mulige sannsynlighetsfordelinger. Optimeringsmodellen finner beslutningen som er optimal for den sannsynlighetsfordelingen fra *ambiguity* settet som gjør mest mulig skade for beslutningstakeren. Vi utvider et eksisterende rammeverk for multistage *distributionally robust* optimering til å kunne ta høyde for risikoaversjon, med bruk av tidskonsistent *conditional value at risk* som risikomål. Scenariotreet som optimeringsmodellen tar inn genereres fra en vifte av simulerte prognoser ved å benytte stokastisk approximering.

Funnene våre viser at sikringsstrategien fra en *distributionally robust* modell er signifikant bedre enn strategien fra en stokastisk modell hvis den verst tenkelige sannsynlighetsfordelingen i *ambiguity* settet skulle vise seg å være riktige. I tillegg presterer denne modellen kun svakt dårligere hvis den estimerte sannsynlighetsfordelingen viser seg å være riktig. Å ta høyde for fordelingsmessig usikkerhet resulterer derfor i beslutninger med lavere risiko. Ved å studere historiske data fra 2014 til 2017 finner vi at *distributionally robust* modellen oppnår høyere gjennomsnittlig profitt enn både tilfellet der man ikke hedger og der man bruker en stokastisk modell. Funnene våre viser derfor at fordelingsmessig usikkerhet bør tas høyde for i optimeringsmodeller som søker optimale sikringsstrategier for kraftprodusenter.

Contents

Preface	i
Abstract	ii
1 Introduction	1
2 Background	5
2.1 Hedging	5
2.2 The Norwegian electricity market	8
3 Literature Review	13
3.1 Hedging based on the original mean-variance analysis	13
3.2 Hedging with stochastic optimization	14
3.3 Accounting for ambiguity	18
4 Problem Description	23
5 Solution Method	25
5.1 Introduction to distributionally robust optimization	26
5.1.1 The Wasserstein distance	28
5.1.2 The nested distance	31
5.2 Creating a solvable problem	36
5.2.1 Scoping down and convexifying the problem	38
5.3 The solution algorithm	44
5.3.1 Transportation kernels	46

5.3.2	A multistage distributionally robust problem with transportation kernels	53
5.3.3	The CVaR optimization problem	57
5.4	Scenario tree generation	61
5.4.1	Scenario tree generation by stochastic approximation	64
6	Electricity price hedging model	75
6.1	Assumptions	75
6.2	The multistage stochastic hedging model	77
6.2.1	Formulating the objective function	78
6.2.2	Formulating the optimization model with a time consistent risk measure	84
6.3	The distributionally robust hedging model	89
6.3.1	Modifying the outer problem	90
6.3.2	Formulating the inner problem	91
7	Computational Study	105
7.1	Problem instance	105
7.1.1	Scenario tree and input data	105
7.1.2	Test data and the rolling-horizon procedure	111
7.1.3	Computational performance	112
7.2	Numerical results	112
7.2.1	In Sample Stability	113
7.2.2	Presenting and evaluating the hedging decisions	114
7.2.3	The expected results of implementing the models	125
7.2.4	out-of-sample performance	130
8	Concluding Remarks	149
9	Future Research	153
	Bibliography	155
	Appendices	165
A	Distributionally robust optimization	167

A.1	Approaches to generate the ambiguity set	167
A.2	An incorrect convex combination	168
A.3	Results from the Inventory Control Problem	
	Example	170
A.4	An accumulated formulation with Nested CVaR	173
B	Numerical results	175
B.1	Hedging strategies from the optimization models	175
B.2	Gain for distributional robustness and Price of ambiguity	179
B.3	Simulations from alternative probability distribution II	182
B.4	Details from backtesting the hedging models	185

List of Tables

3.1	A comparison of this this thesis and the literature on distributionally robust optimization.	21
5.1	Optimal transportation subplans the example in Figure 5.6, where $(k, l) \in \mathcal{N}_1$. $k+_i$ and $l+_j$ are the direct successors of k and l , indexed from the left.	49
5.2	Transportation Kernels created from the optimal transportation subplans of Figure 5.1.	50
5.3	Kernels $K(i, 0)$, $i \in \mathcal{N}_T$ made by the composition of subkernel paths from the root down to the leaf nodes.	50
5.4	Subkernels and optimal transportation subplan conditional on nodes 0 at stage $t = 1$ in Figure 5.6.	52
7.1	Alternative structures for the scenario tree, including the number of scenarios per stage and average nested distance to the scenario fan. The chosen structure is in bold	109
7.2	In-sample stability test for the stochastic hedging model (SO) and the distributionally robust model with $\varepsilon = 5$ and $\varepsilon = 10$. The difference is the represented as the percent-wise absolute deviation from the mean.	114
B.1	<i>2014</i> : Detailed hedging strategies from the stochastic- and distributionally robust hedging models.	176
B.2	<i>2015</i> : Detailed hedging strategies from the stochastic- and distributionally robust hedging models.	177
B.3	<i>2016</i> : Detailed hedging strategies from the stochastic- and distributionally robust hedging models.	178

B.4 2017: Detailed hedging strategies from the stochastic- and distributionally robust hedging models. 179

B.5 Expected utility for the distributionally robust (DRO) and stochastic model (SO) if the realized probability model is the worst case model. This is presented for a set of ε . The difference, gain for distributional robustness (GFDR), is here represented in percent. 180

B.6 Expected utility for the distributionally robust (DRO) and stochastic (SO) models if the realized probability model is the baseline model. The price of ambiguity (POA) is the difference between the stochastic and the distributionally robust expected utility, here presented in percent. 181

List of Figures

5.1	The optimal transportation plans π_{i_3j} between probability distributions \hat{P} and P with $\hat{\mathcal{N}} = \mathcal{N} = 3$	30
5.2	Two processes that have identical leaf node probabilities and states. The tree structures are on the other hand different, causing the nested distance to be different between the models.	32
5.3	<i>Left:</i> The compound tree $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$, <i>Right:</i> The augmented tree $\mathbb{P}_+^{(0)}$	40
5.4	If the nested distance $\text{dl}_r(\mathbb{P}_+^{(0)}, \mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda))$ is less than or equal to the convex combination of $\text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})$ and $\text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(2)})$ then we know that also $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ is within the ambiguity set.	41
5.5	The convex hull of the ambiguity set is a compound, here consisting of k sampled probability models and the convex combinations between them.	43
5.6	Two trees with same structure \mathbb{T} and different probabilities. Unconditional probabilities are shown below the leaf nodes. Conditional probabilities are shown along arcs.	49
5.7	The inner problem solution algorithm for solving the distributionally robust optimization problem with CVaR.	60
5.8	The scenario tree generation method transforms a scenario fan(left) to a scenario tree(right).	62
7.1	A scenario tree (right) is generated from a scenario fan (left) based on stochastic approximation. In the hedging models, the scenario fan consists of 220 forecast paths from the EMPS model. The generated scenario tree contains 5 stages and 192 scenarios at the last stage, where each scenario contains information about the spot price and production.	108

7.2	A scenario tree for the price of a forward contract with maturity in period 2. The forward price at a stage is consistent with the conditional expectation of the spot price at the maturity period	110
7.3	The forward curve for monthly contracts with maturities from January to September. The forward price is consistent with the expectation of the spot price at maturity.	110
7.4	Average yearly hedge ratios over the period 2014-2017, achieved by the stochastic and distributionally robust hedging model with varying ambiguity radius.	115
7.5	When the ambiguity radius increases, the worst-case scenario trees narrow down to single scenarios. Note that the scenarios are not presented in the order of increasing profits.	117
7.6	Expected production for increasing ambiguity radius (left), and expected spot price for increasing ambiguity radius (right). The figure is based on the optimizations conducted on an arbitrary year within the considered time horizon.	118
7.7	Expected hedge ratios for increasing ambiguity, illustrated with uncertain production (dark columns) and with deterministic production (light columns).	119
7.8	Hedge ratios for the considered months in 2017, split on entrance date of the contracts. The strategies from the stochastic model (left) and distributionally robust model with $\varepsilon = 5$ (right) are shown	121
7.9	The monthly hedge ratios for 2017 for various levels of α	122
7.10	The monthly hedge ratios for 2017 for various levels of risk aversion, λ . The stochastic model (left) and distributionally robust model with $\varepsilon = 10$ (right) are shown.	123
7.11	Hedge ratios in February 2017 for different types of contracts from the stochastic model.	124
7.12	The price of ambiguity, marked as difference between the <i>solid</i> and the <i>dashed</i> line, and the gain for distributional robustness, marked as the difference between the <i>dotted</i> and the <i>dash-dotted</i> line. The gain for distributional robustness is higher than the price of ambiguity.	127
7.13	Reduction in objective function value per outer problem iteration for problems with and without the CVaR problem iterations.	129

7.14	Average hedge ratios 2015 for the stochastic- and distributionally robust hedging model with varying ambiguity radius.	132
7.15	Illustration figure: The scenarios from the EMPS model and their corresponding probabilities under the assumed baseline probability distribution. All the scenarios have equal probabilities of being drawn in the simulation.	133
7.16	Mean profits and tail risk under the simulations from the baseline probability distribution. The performance is shown for the naive-, stochastic- and distributionally robust hedging model	134
7.17	Deviations in expected profits and tail risk under the simulations from the baseline probability distribution. Shown for the stochastic- and distributionally robust hedging model, relative to the naive strategy	135
7.18	Utilities for the naive-, stochastic- and distributionally robust hedging strategies under the assumed baseline distribution, measured on the primary axis. The price of ambiguity for the distributionally robust strategies, measured on the secondary axis.	137
7.19	Illustration figure: The potential scenarios from the EMPS model and their corresponding probabilities under the alternative probability distribution. Compared to the assumed distributions, the downside revenue scenarios are weighted with higher probabilities.	138
7.20	Mean profits and tail risk under the simulations from the alternative probability distribution. The performance is shown for the naive-, stochastic- and distributionally robust hedging model	139
7.21	Deviations in expected profits and tail risk under the simulations from the alternative probability distribution. Shown for the stochastic- and distributionally robust hedging model, relative to the naive strategy	140
7.22	The profit density functions of different hedging strategies under the alternative probability distribution	141
7.23	Utilities for the naive-, stochastic- and distributionally robust hedging strategies under the alternative distribution, measured on the primary axis. The gain for distributional robustness is measured on the secondary axis. .	142
7.24	Average yearly profits from backtesting the naive-, stochastic- and distributionally robust hedging strategies for 2014-2017 on historical data. The relative deviation in profits from the naive strategy is shown on the secondary axis.	144

7.25	The standard deviation of the average yearly profits from backtesting the respective hedging models, measured on the secondary axis. Average yearly profits are shown for the hedging models are shown on the primary axis.	145
7.26	Yearly profits for the naive-, stochastic- and distributionally robust strategy with $\varepsilon = 5$ and $\varepsilon = 50$. Backtested on historical data from 2014-2017	146
A.1	The arc values are conditional probabilities, the nodes values are the values in the state space.	169
A.2	When the ambiguity radius increases the scenario tree narrows down to focus on only a few scenarios.	172
A.3	Decline in objective function value when the ambiguity radius increases. . .	173
B.1	Illustration figure: The potential scenarios from the EMPS model and their corresponding probabilities under the probability distribution the simulations are conducted from. Compared to the baseline distributions, the downside revenue scenarios are weighted with higher probabilities.	182
B.2	Mean profits and tail risk under the simulations from the worst case probability distribution. The performance is shown for the naive-, stochastic- and distributionally robust hedging models. We observe that the mean profits and the tail risk is respectively the highest and lowest for the distributionally robust model with $\varepsilon = 10$	183
B.3	Deviations in mean profits and tail risk under the simulations from the worst case probability distribution. Illustrated for the stochastic- and distributionally robust hedging models, relative to the naive strategy	183
B.4	The utilities for the naive-, stochastic- and distributionally robust hedging strategies under the worst case distribution are measured on the primary axis. The gain of distributionally robustness is measured on the secondary axis. We observe that the gain of distributionally robustness is highest for the distributionally robust model with $\varepsilon = 10$. Further, this gives a significantly higher gain of distributionally robustness than price of ambiguity, as is shown in Chapter 7.	184

- B.5 *2014*: We observe that the realized profits tend to increase with higher degrees of ε . This is because the achieved forward prices exceed the spot price for the first four months. The realized hedge ratios are over 100% in March following the negative deviation in realized production relative to the expectation for that month. The latter emphasizes the risk of high hedge ratios under production uncertainty. 185
- B.6 *2015*: We observe that high hedge ratios was favourable in terms of profits because the achieved forward price exceeded the spotprice for all months. We further observe low hedging for $\varepsilon = 50$ in May, in line with the discussion in Chapter 7. In turn this gives slightly lower profits for this model. . 186
- B.7 *2016*: We observe decreasing realized profits for strategies up to $\varepsilon = 10$. For higher ambiguity radii however, profits are higher. This follows as $\varepsilon = 25$ and $\varepsilon = 50$ gave close to zero hedging in May, in line with the discussion in Chapter 7, and that the spotprice in this month was significantly higher than the achieved forward price. 186
- B.8 *2017*: We observe strictly decreasing revenues for the models with higher ambiguity radii as forward prices were exceeded by the spotprice in every month. Further, we observe a high deviation between the expected and realized production in April. 187

List of Algorithms

5.1	Successive convex programming algorithm	45
5.2	Stepwise linearization of the inner problem (5.27)	54
5.3	Scenario tree generation by stochastic approximation	74

Chapter 1

Introduction

Electricity producers are exposed to a considerable risk from uncertain spot prices and production quantities. As an example to illustrate this, the spot prices in 2016 varied from 15 EUR/MWh to 81 EUR/MWh. In order to reduce the risk exposure, hedging has become an important part of most electricity producers risk management. While reducing production risk through hedging is difficult, price risk can be reduced by using financial derivatives, which in the Nordic electricity market primarily involves application of forward contracts. When entering into a forward contract, an electricity producer agrees on selling electricity for a specified price at a future date. Consequently, the risk related to decreasing spot prices is reduced, but so is also the potential upside. Hedging therefore involves a trade-off between risk and potential rewards. Following this, an important and complex decision for electricity producers is to determine the hedging positions that ensure as high potential profits as possible, while satisfying the company's level of risk aversion.

TrønderEnergi is a Norwegian electricity producer who lacks a hedging strategy for reducing the risk exposure. This thesis therefore suggests an optimization based decision support tool for determining an optimal hedging strategy for TrønderEnergi. In doing this, a distributionally robust optimization model is proposed and compared against a stochastic optimization model.

Stochastic optimization is a widely recognized optimization paradigm for decision making under uncertainty, where the realization of a random variables is assumed to be governed

by a probability distribution. A fundamental limitation with stochastic models, however, is the assumption that the probability distribution is known. Limited and inaccurate distributional information, as well as changing statistical characteristics over time, are both factors that makes it difficult to accurately estimate probability distributions. It is further recognized that estimation errors highly deteriorate the decision quality of stochastic models. As a consequence of distributional uncertainty, Stochastic optimization models are thus exposed to unwanted model risk, which in the literature is known as the concept of *ambiguity*.

Distributionally robust optimization is on the other hand a modelling paradigm that approaches ambiguity by letting the observed distributional information itself be subject to uncertainty. This is achieved by considering a set of distributions in the proximity of the one that is observed, and find the best decision for the one that would cause the most harm. Even though this robustification on one hand comes at the cost of lower expected profits, it on the other assures a considerable reduction in downside losses if a more unfortunate distribution than the expected is the true.

The distributionally robust hedging model we develop and test in this thesis is applied on the decision problem of an electricity producer, using historical data. This has to our knowledge never been done. Because the literature on distributionally robust optimization is highly theoretical, this thesis is an important contribution. In addition, while the literature commonly regards single- or two stage distributionally robust optimization problems, the model we develop can handle multistage decisions. Our model is based on an existing framework for multistage distributionally robust optimization, which assumes risk neutral decision making. We extend this framework to manage risk aversion and incorporate conditional value at risk (CVaR) as a risk measure. CVaR is also made time consistent so that it can be applied to a multistage problem. The performance of the distributionally robust hedging model is simulated and tested for different degrees of ambiguity aversion on historical data, and compared to a stochastic model and a naive strategy without hedging. In generating the scenario tree for the distributionally robust hedging model, we apply a recently developed clustering algorithm for dynamic generation of multistage scenario trees. The method bases on the conditional distribution of a set of forecasts paths, and the tree is generated using random draws from the conditional distribution.

The outline of the report is organized as follows: Chapter 2 gives the necessary background

information about hedging and the Norwegian electricity market. Chapter 3 is a literature review, discussing how optimal hedging decisions have historically been approached. In Chapter 4 a description of the hedging problem TrønderEnergi faces is elaborated. Chapter 5 gives a detailed description of the solution methods applied to approach the hedging problem. This involves the general approach for solving the multistage distributionally robust optimization model, as well as the algorithm for generating the scenario tree. Further, Chapter 6 presents the mathematical formulation of both a distributionally robust and a stochastic hedging model. In Chapter 7 we present and discuss the results from the hedging models, focusing on the differences between the distributionally robust model and its stochastic counterpart in a risk-reward perspective. The concluding remarks are found in Chapter 8, followed by a discussion of future research in Chapter 9.

Chapter 2

Background

The purpose of this chapter is to give the necessary background information for this thesis. The chapter is organized in two, where the first section covers hedging, while we in the second section present relevant information about the Norwegian energy market.

In this thesis we expect the reader to be familiar with the fundamental concepts and terminologies within statistics, finance and optimization.

2.1 Hedging

The following section is structured in three. The concept of hedging is first introduced, followed by a presentation of the most commonly applied hedging instruments. Lastly, a discussion on literature investigating empirical findings from hedging among companies is conducted, focusing on the motivation and implication of hedging.

Hedging is a risk reducing investment, widely applied for companies and investors that are seeking to reduce the likelihood of large losses. A hedge normally involves investing in a derivative, a financial instrument whose value is determined by an underlying asset such as the electricity price (McDonald, 2014). When applying derivatives for hedging, an off-setting position in the derivative is taken, meaning that the value of the instrument increases when the value of the underlying asset decreases (Alexander, 2008). Hence, by investing in derivatives written on the electricity price, an energy provider can reduce its

sensitivity to volatile spot prices, and hence the risk of losses through periods of falling prices. However, although hedging reduces potential losses, the potential profits are often reduced as well. Hedging is thus a risk-reward trade-off (McDonald, 2014), where the size of the hedging positions depends on the company's willingness to take risk, i.e. its risk aversion.

Examples of derivatives that are widely applied for hedging includes forwards, futures and options. A forward contract is an agreement between two parties to sell or buy a quantum of the asset, for a specified price at a future date (McDonald, 2014). Consider the following example on how an energy provider can apply forwards to hedge its exposure to price risk. The provider can enter into a short position on forwards written on the electricity price, which means agreeing on selling a defined quantum of electricity for a specified price at a future date. This future date is known as the time of maturity. Then, if the spot price at the time of maturity is below the predetermined forward price, the company gains from the contract, while the opposite happens in cases where the spot price is above the forward price. Consequently, the risk for large losses and the volatility in the cash flow is reduced. Following this, an advantage of hedging is greater predictability (McDonald, 2014). Further, by varying the quantity of electricity that is sold through forward contracts, the company can adjust its risk exposure.

Future contracts are similar to forwards, but standardized with respect to the quantities, maturities and the underlying assets. In addition, futures are traded on an exchange. This is in opposition to forwards, which are traded over the counter (McDonald, 2014). Following the ability to customize the contracts, forwards are suitable for hedging. However, a drawback with forwards compared to futures is that the liquidity often is lower, and that the default risk often is higher for such contracts (Alexander, 2008).

An option gives the investor the right, but not the obligation, to buy (a call option) or sell (a put option) an asset for a specified price (a strike price), at a future date. Hence, oppositely from forwards and futures, the option is only exercised by the investor if the strike price is favourable compared to the spot price at maturity. Put options are suitable for hedging, as the downside risk of the spot price is eliminated, while the upside potential of the spot price is maintained. On the other hand, options requires a premium to be payed upfront of the investment, hence being more expensive than futures and forwards.

As indicated in this section, Tanlapco and Liu (2002) point out that the main motivation of

hedging among companies is reducing the risk of unfortunate price movements, rather than to profit from it. This is further elaborated in Stulz (1996), where it is argued that hedging is about reducing the probability of large losses, while preserving the upside potential as far as possible. In Sanda et al. (2013), the risk management among 12 Norwegian hydropower companies is studied. Here, the stakeholders risk aversion is pointed out as one of the major reasons why the considered companies enter into derivative positions. The authors argue that this mainly is motivated by the companies' ability of providing stable dividends to the owners. Sanda et al. (2013) further show that hedging successfully has reduced the downside risk for a majority of the studied hydropower companies. As hedging not is expected to yield any profits, a more surprising result is that a majority of the companies increased profits and did not reduce the cash flow volatility by entering into derivative positions. An explanation for this is that these companies incorporate market views into their hedging decisions. Including speculative considerations into hedging decisions conflicts with Stulz (1996) and Tanlapco and Liu (2002), but is in line with Deng and Oren (2006), where profit maximization is mentioned as a motivation for hedging.

Although hedging is proven as a useful tool for reducing a firm's exposure to risk, it is debated in the literature whether or not it increases the value of a firm. In a world where investors are able to reduce risk on their own at no cost, Modigliani and Miller (1958) argue that hedging does not increase firm value. However, this is under strict assumptions of market efficiency. Smithson and Simkins (2005) argue that following reduced cash flow volatility, the value of a firm can be increased due to a reduced probability of financial distress, and the ability to take advantage of attractive investment opportunities during bad industry cycles. Froot et al. (1993) support the latter, arguing that this follows from the additional cash flow that hedging positions create for a firm under tough market conditions. Carter et al. (2006) investigate the relation between hedging and firm value in the US airline industry. They find that the airlines that hedged against higher fuel prices are higher valued than the airlines that did not hedge. The main reason is that less risk exposed firms can take advantage of investment opportunities through cycles of high fuel costs to increase their market shares. Further, Smith and Stulz (1985) find that an important argument for why firms should hedge is that the expected costs of financial distress are reduced. The reason for this is that hedging reduces the risk for bankruptcy, which further should increase the value of a firm. Graham and Rogers (2002) also emphasize this, and further argue that another important reason for firms to hedge is

to increase the debt capacity. Following higher debt, additional tax benefits increases the firm value. Jin and Jorion (2006) study the hedging activity among 119 US oil and gas companies from 1998 to 2001. They verify that hedging reduces the sensitivity to volatile oil prices. However, they find that it does not tend to increase the value of the involved companies.

Although hedging is a useful tool for risk reduction, making the correct hedging decisions can be challenging as it often requires expertise and application of decision supporting tools. In Chapter 3 we discuss how the literature approaches optimal hedging strategies. The next section presents information about the Norwegian energy market that is relevant for this thesis.

2.2 The Norwegian electricity market

In the following section we discuss the Norwegian electricity market, focusing on aspects that are relevant for hedging. In brief, the section covers the characteristics of the most significant sources of risk in the market; the electricity production and the electricity price. In addition, the types derivatives available in the market for handling this risk are discussed. We lastly discuss the tax system that Norwegian hydropower producers are exposed to, a system that influences the hydropower producers' hedging strategies.

Nearly all electricity production in Norway steams from hydropower, where the major source of uncertainty is variations in hydro inflow. Further, a large share of the production capacity is flexible, meaning that water can be stored as potential energy in reservoirs (NVE, 2016). Non-storability of electricity is however an important characteristic of the electricity market, meaning that it is consumed at the same time as it is produced (Ek and Thorbjørnsen, 2014).

The main area for trading power is the Nord-Pool day-ahead market, which is the electricity spot market. Even though a spot market technically should handle immediate sales, the electricity traded in the Nord-Pool day-ahead market is sold the next day (Nord-Pool, 2017c). The Norwegian electricity market is divided into five price regions, NO1 to NO5, where the price of each area is known as the *area price*. Area prices are determined by the balance between supply and demand, where a major factor impacting supply and demand in an area is changing weather conditions. The *system price* on the other hand

is the unconstrained market clearing reference price for the entire Nord Pool exchange, including all the orders from the Nordic and Baltic areas (Nord-Pool, 2017a).

Compared to traditional financial markets, electricity prices are known to be highly volatile, emphasizing the relevance for risk management (Bystrøm, 2003). According to Geman (2008), the nonstorability of electricity is a major reason for the high volatility. In addition, there are seasonal variations in the price, where the prices normally are higher during the winter than during the summer. Further, Ek and Thorbjørnsen (2014) find a positive linear relationship between prices and production in the electricity market during the winter, and a weaker relationship during the summer. Following this, the cash flow sensitivity for electricity producers is higher during the winter, which Ek and Thorbjørnsen (2014) further use to argue that the need for hedging is higher through these periods. However, Sanda et al. (2013) find that there is no trend among Norwegian electricity producers to increase hedging during the winter.

The most liquid exchange traded derivatives available in the electricity market are traded on NASDAQ OMX Commodities Europe. This exchange includes regular futures, deferred settlement (DS) futures, options and Electricity Price Area Differentials (EPADs). Both futures and DS futures have identical characteristics, except of DS futures having settlements during the whole delivery period (NASDAQ, 2017b). Despite of being exchange traded, DS futures are therefore often seen as forward contracts according to Sanda et al. (2013). For futures, DS futures and options, the reference price is the Nordic system price, while the difference between the area price and the system price is the reference for EPADs. In theory, futures in combination with EPADs are therefore suitable for producers seeking to hedge against the area price risk in the price region they are operating, and not only the system price risk (NASDAQ, 2017a).

All derivatives on NASDAQ are settled in cash, meaning that the difference between the spot price and the derivative price is paid at the settlement, instead of electricity physically being exchanged. Further, there exists derivatives on NASDAQ with time horizons up to ten years, with both daily, weekly, monthly, quarterly and annual delivery periods (NASDAQ, 2017b).

The contracts traded on NASDAQ has delivery over a period, oppositely to the standard for storable commodities, where delivery is at a specific time (Ek and Thorbjørnsen, 2014). Following this, the contracts are also commonly referred to as swaps, which is a

derivative where a fixed price is switched with a floating price over a period with multiple settlements (Benth and Koekebakker, 2008). However, irrespective of differences in the settlement, both forwards, futures and DS futures are referred to as forwards through the rest of this thesis.

A problem discussed in Tanlapco et al. (2002) is limited liquidity for forwards with longer maturities, where lack of potential buyers and sellers can be problematic. Sanda et al. (2013) also point out that the low liquidity of EPADs make them less suitable to be applied as hedging instruments. EPADs are therefor not considered as hedging instruments in this thesis.

Note that none of the exchange traded financial derivatives cover the production risk that mainly follows from uncertainty in hydro inflow. Hedging production uncertainty is hence difficult, although there exist alternatives such as weather derivatives where the relationship between temperatures and demand for electricity can be utilized (Oum and Deng, 2005). For examples of papers where the volumetric risk is approached, see Keppo (2002), Oum and Oren (2010) and Yumi et al. (2006). In addition, under volumetric uncertainty, Mo et al. (2001) argue that production planning and risk management should be integrated to optimize the decision making.

Norwegian electricity producers are exposed to tax regulations that affects the risk management. In addition to the corporate tax of 23%, electricity producers are also obligated to pay a natural resource tax of 35,7% for hydropower production to compensate for the use of natural resources (Finansdepartementet, 2018). While the basis for the corporate tax is the attained income, the natural resource tax is based on the spot price when the electricity is produced, unconditional on whether the electricity is sold in the spot market or through forward contracts. Hence, producers risk to pay the natural resource tax on non-earned revenues in cases where the spot price exceeds the forward price at maturity, and vica versa in cases where the spot price is exceeded by the forward price. In understanding why the presence of the natural resource tax is decision relevant, the following paragraph considers an example of a producer seeking to fully eliminate the price risk exposure of a deterministic production quantum

For a producer seeking to fully hedge the price risk, the whole deterministic quantum of electricity should be sold through forward contracts in the case where there is no natural resource tax. That is, a hedge ratio of 1. When the natural resource tax is present we prove

that the hedge ratio eliminating all price risk is 53.6%. A full hedge is a position such that any change in price has no effect on the profits. This can be expressed mathematically as setting the differentiated profit function with respect to price to zero. The hedging position is found by solving this equation, which is written in equation (2.1) below. Here, x is the hedge ratio and the produced quantum is set to one.

$$\frac{d}{dSpot} \left(x \cdot Forward + (1 - x) \cdot Spot - Tax_{res} \cdot Spot - Tax_{corp} (x \cdot Forward + (1 - x)) \right) = 0 \quad (2.1)$$

Solving (2.1) gives a hedge ratio x of 0.536. Hence, the optimal hedging strategy depends on the presence of natural resource taxes. Specifically, the tax system Norwegian hydropower producers are exposed to gives incentives to less hedging (Ek and Thorbjørnsen, 2014) as there exists a risk that natural resource taxes must be paid on non-earned revenues in cases where the spot price exceeds the forward price. These calculations are also in line with Sanda et al. (2013).

Chapter 3

Literature Review

This chapter contains an overview of how optimal hedging decisions have been approached in the literature, focusing on applications in the electricity market. As Norwegian electricity producers rely on hydropower-production, the literature we present on hedging has a focus on studies with hydropower applications. Among the variety of optimization frameworks that historically have been applied to hedging decisions, some are better at handling uncertainty than others. The importance of precisely capturing the underlying uncertainty is emphasized by findings showing that failing to do so highly affects the performance of financial application optimization models (see e.g Chopra (1993), Chopra and Ziemba (1993)). The literature presented in this chapter is therefore ordered according to an increasing ability towards handling uncertainty.

3.1 Hedging based on the original mean-variance analysis

The foundation of modern portfolio optimization was laid with the mean-variance analysis in Markowitz (1952). With the aim of maximizing return under risk constraints, or minimizing risk given a required expected return, the analysis illustrates the trade-off between reward and risk in a financial optimization problem. Markowitz's modern portfolio theory is applicable to hedging decisions as well, where the objective is to find the optimal allocation between derivatives and spot price exposure, given the decision maker's

risk preferences. This makes the mean-variance theory a fundamental starting point for obtaining optimal hedging decision.

Several papers approaching optimal hedging strategies for electricity producers rely directly on Markowitz' modern portfolio theory. In Woo et al. (2004), an efficient frontier, which gives the optimal risk-reward tradeoffs for different degrees of risk aversion, is constructed by combining different fractions of spot exposure to sales through forward contracts. Further, while Woo et al. (2004) considers the UK and US market, a similar study is conducted with focus on the Turkish electricity market in Gökgöz and Atmaca (2012). In Ek and Thorbjørnsen (2014), optimal hedge ratios with forwards are found for hydropower producers in the Norwegian electricity market. Specifically, the effect of seasonality on optimal hedging decisions is investigated. The optimal hedge ratios are found to be higher during the winter following a stronger positive dependency between spot prices and production quantities through these seasons. In approaching the optimal policies, different risk measures were optimized, hence disregarding the tradeoff to expected revenues.

A major drawback with the original Markowitz model is that it is deterministic, meaning that the uncertainty of its input parameters is disregarded. The approach is for this reason inaccurate in every other scenario than when the estimates are correct. Consequently, an important issue related to this approach is to estimate input data with sufficient precision (Luenberger, 1980; Merton, 1961). The severity of this drawback is emphasized since it is shown that erroneous estimates tend to strongly deteriorate the performance of classical Markowitz models (Best and Grauer, 1991; Chopra, 1993; Michaud, 2001).

3.2 Hedging with stochastic optimization

Stochastic optimization is an optimization framework that addresses uncertainty better than deterministic optimization. There are two primary benefits of this framework:

- Stochastic optimization accounts for uncertainty by incorporating possible outcomes of the random variables, and allowing the realization of the random variables to follow some probability distribution (Higle, 2005).
- In a model with multiple decision stages, stochastic optimization values flexibility,

which allows the decision maker to adapt to new situations once the outcome of a random variable is observed (Pflug and Pichler, 2014a).

As pointed out by Bertocchi et al. (2011), these two properties makes stochastic optimization highly attractive for financial decision making. For this reason there exists a substantial amount of literature approaching optimal hedging strategies for hydropower producers with stochastic optimization. We now provide an overview of the similarities and differences between these approaches, and then show how this relates to the decision problem covered in this thesis.

Overview of hydropower hedging problems approached with stochastic optimization

We first present the similarities between the hydropower hedging problems approached with stochastic optimization, and then categorize the differences between them. We lastly position the decision problem considered in this thesis.

A first common factor of the hydropower hedging problems approached with stochastic optimization is the decision problem, which generally is formulated as a trade-off between expected profits and the level of risk related to a decision. In the overview of Wallace and Fleten (2003) on literature concerning optimization for hydropower producers, the following general decision problem for risk management through hedging is proposed. For each stage of a planning period, the decision maker needs to find the optimal quantity or proportion of forward contracts of a given type to enter into, so that expected profits are maximized while satisfying some acceptable level of risk. This formulation is an instance of the Asset-Liability-Management model of Kusy and Ziemba (1986), which is a general framework for financial decision problems under uncertainty (Consigli et al., 2016).

A second common factor in modern literature concerning optimal hedging strategies for electricity producers is the use of multistage stochastic optimization. Examples that show beneficial results of forming hedging strategies for hydropower producers with multistage stochastic optimization are Conejo et al. (2008), Pineda and Conejo (2012) and Fleten et al. (2002). In particular, the latter of these papers compares the strategies made by a multistage model against those of a static model. It is shown that the flexibility of multistage models improves the decision quality in terms of a better risk reward trade-off,

where one can sacrifice less expected profits for a higher reduction in risk.

Several aspects from which one can analyze the differences between hydropower hedging problems exist. The most important are discussed below with references to the literature.

- **With or without production planning** – Works such as Mo et al. (2001) and Fleten et al. (2002) show an improved risk-reward trade-off by *integrating* production planning in the hedging decision. There are on the other hand instances where hedging is *separated* from production. A common reason for separating hedging and production decisions is that these decisions are made in different organizational departments (Wallace and Fleten, 2003).
- **Types of risk considered** – Price risk is naturally the risk factor forming the common ground in the literature. Another source to risk that often is considered for hydropower producers is production risk, which largely is caused by uncertainty in inflow to hydropower plants. Risk factors that are less common in the literature includes the risk related to differences in the area price and the system price, covered in Woodard and Garcia (2008) and Broll et al. (2015), and *Currency risk* which for instance is covered in (Wu and Sen, 2000).
- **Choice of risk measure** – One of the most accepted risk metrics today is the conditional value at risk (CVaR), which considers the downside tail risk of a portfolio. Applying a risk measure that consider the downside tailrisk is further in line with Stulz (1996) who argues that risk management is more about reducing the probability of large losses, than to reduce the volatility in the cash flow. CVaR is measured as the expected loss, given that the loss exceeds the value at risk (VaR). Further, VaR is defined as the threshold to which losses shall not exceed with a given confidence level (Zenios and Markowitz, 2008). In addition to its financial relevance, Rockafellar and Uryasev (2000) show that CVaR can easily be integrated in stochastic problems. For this reason, CVaR is widely applied in the literature on hedging with stochastic optimization (see e.g. Conejo et al. (2008), Pineda and Conejo (2012), Shütz and Westgaard (2018)).

A problem with CVaR in multistage optimization models is that the risk measure is time inconsistent, potentially leading to erroneous estimations of the risk at intermediate states (Rudloff et al., 2014). The reason for this is that risk is not considered

according to the conditional probabilities at intermediate stages, or as stated in Rudloff et al. (2014): "at every state of the system, our optimal decisions should not depend on scenarios which we already know cannot happen in the future". A time consistent formulation of CVaR was first formulated by Shapiro (2009). In contrast to the standard formulation of CVaR, the time consistent formulation assures that the risk is regarded throughout the whole stochastic process, by ensuring that only scenarios that are reachable from the current state (Shapiro, 2009). Rudloff et al. (2014) show that applying a time consistent CVaR gives more conservative decisions by using a time inconsistent CVaR formulation. Our model formulation in Chapter 6 is based on the approach of Rudloff et al. (2014), who ensure time consistency with a recursive CVaR formulation. Applications of this formulation within hedging and power production are presented in Shütz and Westgaard (2018) and Pisciella et al. (2016).

Applications of alternative risk measures than CVaR are for instance presented in Fleten et al. (2002), who consider the *target shortfall* exceeding a predetermined target profit. Other risk measures such as total absolute deviation, value at risk and variance are evaluated in Gómez-Villalva and Ramos (2004) and Ek and Thorbjørnsen (2014).

- **Types of derivatives** – As emphasized in Chapter 2 lack of liquidity commonly reduces the number of available derivatives. While this makes forward contracts the most common type in the Nordic electricity market, applications using options are for instance found in Conejo et al. (2008) and Pineda and Conejo (2012).

The decision problem we base on in this thesis considers a portfolio of hydropower plants, where production planning is done separately from the hedging decision we approach. Both production and price risk are considered, and a time consistent CVaR is used as risk measure. As hedging instrument, forward contracts are regarded.

A fundamental challenge of stochastic optimization models is the assumption of the probability distribution of the random variable to be known, while in reality often being subject to uncertainty. This is problematic because stochastic optimization problems tend to be highly sensitive to wrongly estimated distributional parameters (Consigli et al., 2016). Stochastic models thus often give rise to an overfitting problem referred to as the "Optimizer's curse", causing poor performance on out of sample tests, despite of being perfectly

optimized for in sample data (Smith and Winkler, 2006). In the literature, distributional uncertainty is known as the concept of ambiguity (Ellsberg, 1961). Ambiguity aversion therefore represents a decision maker's preference for known probability distributions over uncertain distributions. Risk aversion on the other hand describes a decision maker's preference for less risky portfolios with lower expected payoffs, under a known probability distribution of the uncertain assets (Ellsberg, 1961).

The shortcoming of stochastic optimization when it comes to handling distributionally uncertainty is approached by an optimization paradigm called distributionally robust optimization. Relevant literature on this field of optimization is presented in the next section. There are however no literature on distributionally robust optimization applied on hedging problems.

3.3 Accounting for ambiguity

The paradigm of distributionally robust optimization approaches the problem of ambiguity by letting the probability distribution be subject to uncertainty itself. This is achieved by assuming that the true distribution of the random variable is contained in a set of potential distributions, an *ambiguity set*, rather than assuming one specific distribution (Delage, 2017). As neatly illustrated by Consigli et al. (2016), distributionally robust optimization can be interpreted as a game against 'nature'. The decision maker first maximizes expected returns, and as a response, 'nature' selects the distribution, from a set of potential probability distributions, that inflicts maximum harm to the decision maker. Consequently, the decision maker accounts for the distributional uncertainty by making the decision that is optimal under the worst-case conditions.

The earliest approach in the direction of distributionally robust optimization is the work of Žáčková (1966), who attempted to solve stochastic programs under limited distributional information. Distributionally robust optimization did, however, not come to exist until after robust optimization became popular in the late 1990s, as an alternative to stochastic optimization (Postek et al., 2014).

Based on the work of Ben-Tal and Nemirovski (1997), robust optimization addresses distributional uncertainty by assuming the uncertain input parameter to be contained within some uncertainty set, and then by optimizing for the worst-case. Robust optimization has

proven to have a multitude of successful implementations in a wide range of industries (Ben-Tal et al., 2009). Among these there are also examples applied to hedging (Shen et al., 2013; Fonseca et al., 2009). The principle findings from these papers are that robust models tend to perform better when expected return is overestimated, and that the worst-case approach of robust optimization is useful for risk averse investors. However, robust optimization is criticized for being too conservative, as all distributional information is disregarded (Consigli et al., 2016).

Distributionally robust optimization came to rise during the 2000s, as a generalization of robust and stochastic optimization. The different methods for formulating distributionally robust optimization problems are often distinguished by how they model ambiguity. A second factor that distinguishes the literature on distributionally robust optimization is the amount of distributional information that is assumed available.

The differences in the methods for modelling ambiguity is determined by how the ambiguity set is formulated. The ambiguity set is a set of possible probability distributions, which is large enough to contain the unknown true probability distribution. There are three common methods for formulating the ambiguity set.

The first method creates ambiguity sets based on *statistical moments*. Statistical moments are quantitative measures of the shape of a probability distribution (Papoulis, 1965). A statistical moment ambiguity set is formulated by constraining the statistical moments to be within some interval around the moments of the baseline distribution (Delage and Ye, 2010; Goh and Sim, 2010).

The second method creates ambiguity sets from *goodness of fit confidence regions*, where the ambiguity set contains distributions that have passed a statistical hypothesis test relative to the baseline distribution. This is for example done in Bertsimas et al. (2014).

The third method, which we rely on in this thesis, creates the ambiguity set based on a *distance metric*. This type of ambiguity set can be considered as a ball around the empirically observed baseline distribution, with radius according to some statistical distance measure or probability metric. The ambiguity set then contains all probability distributions within some distance metric from the empirically observed distribution. Typical statistical distance measures are ϕ -divergence (Bayraksan and Love, 2015), or the nested distance, which is the basis for the model applied in this thesis in Chapter 5. For examples on papers applying distance metrics in creating the ambiguity set, see Pflug and Wozabal

(2007), Esfahani and Kuhn (2015) and Analui and Pflug (2014). A further elaboration of the pros and cons with the different approaches is found in Appendix A.1.

The second distinction between the distributionally robust problem formulations in the literature, is the assumptions about available information. The following examples emphasize the diversity in these assumptions. The early work of Shapiro and Kleywegt (2002) assumes a finite set of distributions to be known, to which probabilities are assigned such that distributionally robust problems can be solved as stochastic problems over the set of probability distributions. Goh and Sim (2010) assume that only the support and covariance matrix are known, while the mean is subject to uncertainty. Similarly Xin et al. (2013) assume the support of a distribution, that is the area where a probability distribution is not zero, and the two first statistical moments to be known, while Esfahani and Kuhn (2015) assume that the decision maker only has a set of observations and knows the support of the distribution. Finally, Analui and Pflug (2014) require a scenario tree for the baseline distribution, similar to what is required for a stochastic problem.

A problem with the field of distributionally robust optimization is that it is highly theoretical, with lack of application on real data. Hence, only a few papers have investigated the gain of applying distributional robust optimization on real decision problems. Among the applications, Pflug and Wozabal (2007) consider a portfolio optimization problem consisting of six stocks, where the tradeoff between return, risk and robustness to ambiguity successfully is illustrated. Pflug and Wozabal (2007) find that by increasing the robustness against ambiguity, a relatively small reduction in expected returns is sacrificed for a significant reduction in the risk, measured by CVaR. Hence, they find it advisable to apply distributionally robust optimization in portfolio optimization for risk averse investors. Esfahani and Kuhn (2015) also approach a mean-risk portfolio optimization problem of stocks, where similar results are obtained. In addition, the authors point towards a better post decision disappointment for distributionally robust models, in contrast to stochastic models, when comparing realized to expected revenues out of sample. Among other case applications of distributionally robust optimization, Analui and Pflug (2014) approach a simple inventory control problem. They show that the gain of applying a distributionally robust model is found to be higher than the cost of ambiguity. That is, in cases where the true probability distribution is worse than assumed, the gain of applying a distributionally robust model is greater than the loss in cases where the true distribution actually is correctly estimated.

Despite of promising results on portfolio optimization examples, there are still no applications of distributionally robust optimization on hedging problems. We therefore develop a multistage distributionally robust optimization model when approaching the optimal hedging strategy for an electricity producer. The distributionally robust approach used in this thesis bases on the work of Analui and Pflug (2014), but we extend it to incorporate a risk measure. Specifically, a time consistent version of CVaR is applied to measure the risk. As the distributionally robust optimization lack of applications and testing on real-life data, we backtest and benchmark the performance of the model on historical realizations of spotprices and production quantities.

The table below compares this thesis against central literature on distributionally robust optimization. A clear distinction is the application on a practical case with historical data and the use of multiple decision stages.

Literature	Type of ambiguity set	Assumptions about information	Recourse decisions	Practical application
This thesis	Distance metric - nested distance	A complete scenario tree.	Yes (multistage)	Yes
Shapiro and Kleywegt (2002)	Non-existing	Distributions are assigned probabilities.	No	No
Pflug and Wozabal (2007)	Distance metric - Wasserstein distance	A set of observed sample points.	No	No
Delage and Ye (2010)	Statistical moments	Support and first two moments.	No	No
Goh and Sim (2010)	Statistical moments	Support and covariance matrix.	Yes (two-stage)	No
Xin et al. (2013)	Statistical moments	Support and first two moments.	Yes (multistage)	No
Bertsimas et al. (2014)	Goodness of fit test	Support, mean and covariance matrix.	Yes (two-stage)	No
Analui and Pflug (2014)	Distance metric - nested distance	A complete scenario tree.	Yes (multistage)	No
Bayraksan and Love (2015)	Distance metric - Phi divergences	Observed sample points or discrete distributions.	Yes (two-stage)	No
Esfahani and Kuhn (2015)	Distance metric - Wasserstein distance	A set of observed sample points.	Yes (two-stage)	No
Hanasusanto and Kuhn (2016)	Distance metric - Wasserstein distance	A set of observed sample points.	Yes (two-stage)	No

Table 3.1: A comparison of this this thesis and the literature on distributionally robust optimization.

A detailed elaboration of the approach towards formulating and solving the general multistage distributionally robust problem is found in Chapter 5, while the hedging specific model is elaborated in Chapter 6.

Chapter 4

Problem Description

We consider TrønderEnergi AS, a small Norwegian energy producer and provider, operating in the NO3 market. Its portfolio of power plants primarily consists of hydropower plants. Due to the high volatility in electricity prices, TrønderEnergi's revenue is exposed to a significant price risk. In addition, as the hydro inflow to power plants is uncertain, there is a substantial production risk related to TrønderEnergi's operations as well. Hence, fluctuating electricity prices and production quantities may result in a volatile and unpredictable cash flow for TrønderEnergi, potentially leading to large losses.

Managing production risk is difficult in practice, but by hedging with financial instruments, such as forward contracts, TrønderEnergi's price risk exposure can be reduced. TrønderEnergi therefore seeks a model for finding the optimal use of forward contracts, so that expected profits are maximized and the company's level of risk aversion is satisfied.

In addition to determining the optimal quantity to buy of forward contracts with different maturity periods, TrønderEnergi has to consider at what time the respective forwards should be entered. For the latter decision there is trade-off between entering into a position for an agreed price today and waiting for new information. TrønderEnergi considers a planning horizon of 4-6 months as relevant for the decision problem, as this is the time it normally takes to empty the water from a reservoir. Further, the company considers the hedging decision of top on their production decision. Hence, no production decisions are to be considered in this decision problem.

TrønderEnergi considers the following revenue and cost components as decision relevant when developing the optimal hedging strategy:

Revenue components

- The uncertain *spot price* and the *quantity* sold through the spot market for different periods.
- The production quantity sold through *forward contracts* and the price of the respective contracts, for different periods. Following lack of liquidity, options written on the Nordic system price are not applied, according to TrønderEnergi.

Cost components

- *Transaction costs* occur when entering into derivatives, giving TrønderEnergi incentives for entering into a lower quantity of contracts. In addition, as transaction cost are paid when the contracts are entered, this gives TrønderEnergi incentives to wait longer before entering into contracts, as the time value of the transaction costs is decreasing.
- *Tax costs* that have to be considered are the natural resource tax, and the corporate income tax. The natural resource tax is, as opposed to the corporate tax, solely based on the spot price during the production hours, independent on whether the production is sold on contract or in the spot market. A consequence of this, as explained in Section 2.2, is that TrønderEnergi risks to pay the natural resource tax on non earned revenues in cases where the spot price exceeds the forward price through maturity.

As of today, TrønderEnergi has no structured hedging strategy for approaching this decision problem. This thesis suggests a decision support tool for determining an optimal hedging strategy for the TrønderEnergi, given company's level of risk aversion. In order to develop a dynamic tool that properly captures the underlying risk of the electricity prices and production quantities, a multistage distributionally robust optimization model is developed. The details of the model follows is Chapter 5 and Chapter 6. In the remainder of this thesis we use *decision maker* and TrønderEnergi interchangeably.

Chapter 5

Solution Method

The purpose of this chapter is to present the general solution method used to solve the hedging problem described in Chapter 4.

One of the most challenging aspects of formulating a hedging model for an electricity producer, is to address uncertainty in price and production quantity. As presented in Section 3.2, a limitation with stochastic optimization is the assumption of full distributional information, which is problematic since accurate uncertain information in many instances is limited. In turn, this can lead to badly estimated input parameters, which further can deteriorate the quality of decisions given by these tools. Stochastic models hence struggle with an undesired model risk. Recall that Distributionally robust optimization on the other hand is an optimization paradigm that seeks to reduce the model risk of the stochastic models by making the available distributional information in itself subject to uncertainty.

To determine the optimal hedging strategy for TrønderEnergi, we create a multistage distributionally robust optimization model. We also create a stochastic model in order to benchmark the performance of the multistage distributionally robust model.

In Section 5.1 to Section 5.3 we elaborate on how a general multistage distributionally robust problem is formulated and solved. Section 5.1 gives an introduction to the applied formulation and solution method, and provides understanding of how ambiguity is accounted for. In Section 5.2 we use these insights to formulate the multistage distributionally robust optimization problem, starting from a familiar stochastic formulation. In

Section 5.3 we present a solution algorithm for approximating the optimal solution of the formulated problem.

In Section 5.4 we explain the method used to generate the scenario tree which is used as input for the optimization models. An iterative clustering algorithm applying stochastic approximation is used, where the tree is generated from a set of forecast paths.

5.1 Introduction to distributionally robust optimization

Recall that a distributionally robust optimization problem can be understood as a game between two actors. The first actor makes the optimal decision while knowing that the other actor, who is in possession of a set of probability distributions, chooses the probability distribution that causes the most harm to the optimal decision. In the instance where the objective is to minimize a cost function, this analogy can be elaborated as follows: The decision problem consists of two optimization problems that are solved simultaneously. The outer problem sets the decision variables such that the cost function is minimized. The inner problem finds the probability distribution from an *ambiguity set* of probability distributions, with the purpose of maximizing expected costs. Because of these two counteracting problems, distributionally robust optimization problems are called minimax problems.

The purpose of this minimax structure is that it avoids making any assumptions about the probability of choosing one probability distribution over another, which is achieved by simply using the worst-case distribution from the ambiguity set (Conigli et al., 2016). This underlines the importance of modelling ambiguity set properly, which usually is a trade-off between covering the true distribution with sufficient likelihood, while not making the optimal decisions too conservative (Esfahani and Kuhn, 2015).

In this thesis the distributionally robust optimization problem formulation and solution method are based on the work of Analui and Pflug (2014). Recall from Section 3.3 that the different methods for modelling ambiguity can be regarded as a threefold, depending on how the ambiguity set is modelled. Analui and Pflug (2014) formulate the ambiguity set based on a probability metric. A probability metric is in Section 3.3 described

as a measure of the extent to which two probability distributions differ. It can be regarded as the radius of a ball that circumfers the empirically observed *baseline probability distribution*. Increasing this so-called *ambiguity radius* accounts for more and more probability distributions. If the ambiguity radius on the other hand is zero, no ambiguity is accounted for. The distributionally robust optimization problem is then reduced to a stochastic problem (Pflug and Wozabal, 2007).

Analui and Pflug (2014) consider multistage distributionally robust optimization problems. Rather than capturing random outcomes at a single point in time, the applied probability metric measures the difference between two stochastic processes, each representing random outcomes over multiple points in time. This probability metric is called the *nested distance*. We represent stochastic processes with probability models, which is in the hedging model instance of Chapter 6 are in the form of scenario trees. We distinguish probability models from probability distributions, which only represent random outcomes at a single point in time.

A first advantage with the approach of Analui and Pflug (2014) is that it allows multistage formulations, which has been proven to be valuable due to the ability of incorporating flexibility in the decision strategies.

A second advantage is the support for linear formulations with considerable modelling leeway. As a consequence, non-anticipativity constraints and suitable risk measures such as CVaR can be modeled and solved with commercially available optimization software.

The primary disadvantage with this method is that it is an approximation method, that cannot guarantee that the worst-case distribution in the ambiguity set is found. The optimal solution can therefore neither be guaranteed, even though the solution method is proven to always converge towards a stable solution.

Further, the problem uses a scenario tree to model uncertainty. According to Shapiro (2018), there are two central drawbacks with this. First, as we describe in detail in Section 5.4, we often observe data as realizations, i.e. sample paths. This is also the situation in our instance. A scenario tree requires that we evaluate the conditional distributions of the stochastic process, and the only way to do so is to make assumptions about the structure of this process. A second problem is that scenario trees heavily deteriorate the computational performance.

Another limitation is that the structure and values in the scenario tree are assumed to be fixed, while only the probabilities of the scenarios themselves are allowed to vary. Without this assumption the problem would be very difficult to solve due to its resulting size and non-convexity.

Before formulating the distributionally robust optimization, we need to establish an understanding of how the ambiguity set is formulated. In order to do so, we need to understand the concept of nested distance. This concept is a multistage generalization of another probability metric called the Wasserstein distance. An initial step is therefore to comprehend the theory behind this probability metric.

5.1.1 The Wasserstein distance

The Wasserstein distance is a metric used to measure how different two probability distributions are. The Wasserstein distance is the optimal solution of a *mass transportation problem*, which can be explained with the following analogy. Consider the probability distributions as piles of sand. The objective is to move as little sand as possible in the first pile of sand, such that it becomes identical to the second. More precisely, the objective is to transport the mass of one distribution, such that it becomes identical to the other, by transporting the total mass the shortest possible distance.

The mass transportation problem is a linear optimization problem. Consider two distributions \hat{P} and P , discretized as \hat{P}_i and P_j , where the points i and j belong to the sets $\hat{\mathcal{N}}$ and \mathcal{N} respectively. Consider also a distance measure d_{ij} between the points on the probability spaces. A typical distance metric is the r -order distance, which we describe in detail in the paragraphs below. We write the distance measure as d_{ij}^r , and denote the decision variables as π_{ij} . The Wasserstein distance \mathbf{d}_r^W is found by solving

$$\mathbf{d}_r^W = \min_{\pi} \sum_{i \in \hat{\mathcal{N}}} \sum_{j \in \mathcal{N}} d_{ij}^r \pi_{ij} \quad (5.1)$$

s.t.

$$\sum_{j \in \mathcal{N}} \pi_{ij} = \hat{P}_i \quad i \in \hat{\mathcal{N}} \quad (5.2)$$

$$\sum_{i \in \hat{\mathcal{N}}} \pi_{ij} = P_j \quad j \in \mathcal{N} \quad (5.3)$$

$$\pi_{ij} \geq 0 \quad i \in \hat{\mathcal{N}}, j \in \mathcal{N} \quad (5.4)$$

In the following paragraphs we first explain the objective function components, and then explain the constraints of this problem.

The objective function (5.1) consists of two components. The first is the distance measure d_{ij}^r , a matrix of dimensions $\hat{\mathcal{N}} \times \mathcal{N}$ that denotes the distance associated with transporting a unit of mass from point i in \hat{P} to j in P . We hereafter refer to d_{ij}^r with the more general term, transportation cost. A typical distance metric is the r -order distance. For the random variables $\hat{\xi}_i$ and ξ_j , which are vectors of length m , i.e. on \mathbb{R}^m , with weights w^m , the r -order distance is

$$d_r(\hat{\xi}_i, \xi_j) = \left(\sum_{m=1}^M w^m \left| \hat{\xi}_i^m - \xi_j^m \right|^r \right)^{1/r} \quad (5.5)$$

The second component of the objective function (5.1) is the problem's decision variables π_{ij} . When solved to optimality π_{ij} form the optimal transportation plan, which is the amount of mass transported from each point i to j with the lowest possible total transportation distance. Mass transportation with an optimal transportation plan between two distributions is illustrated in Figure 5.1.

The transportation plan π_{ij} is a bivariate probability measure consisting of marginal distributions \hat{P} and P (Pflug and Pichler, 2014b). Intuitively, π_{ij} contains information about the mass transported from any point i in \hat{P} to j in P and the total mass to be

transported hence sums up to one. The bivariate probability distribution π_{ij} therefore has the following properties

$$\begin{aligned} \sum_{i \in \hat{\mathcal{N}}_T} \sum_{j \in \mathcal{N}_T} \pi_{ij} &= 1 \\ \pi_{ij} &\geq 0 \end{aligned} \quad i \in \hat{\mathcal{N}}, j \in \mathcal{N} \quad (5.6)$$

Constraints (5.2) and (5.3) are balance constraints that ensure consistency in the transported mass. (5.2) ensure consistency in the mass transported and the available mass at each point in \hat{P} . Similarly (5.3) ensure that the mass received at P is neither more nor less than what is demanded. Constraints (5.4) prevent negative probabilities.

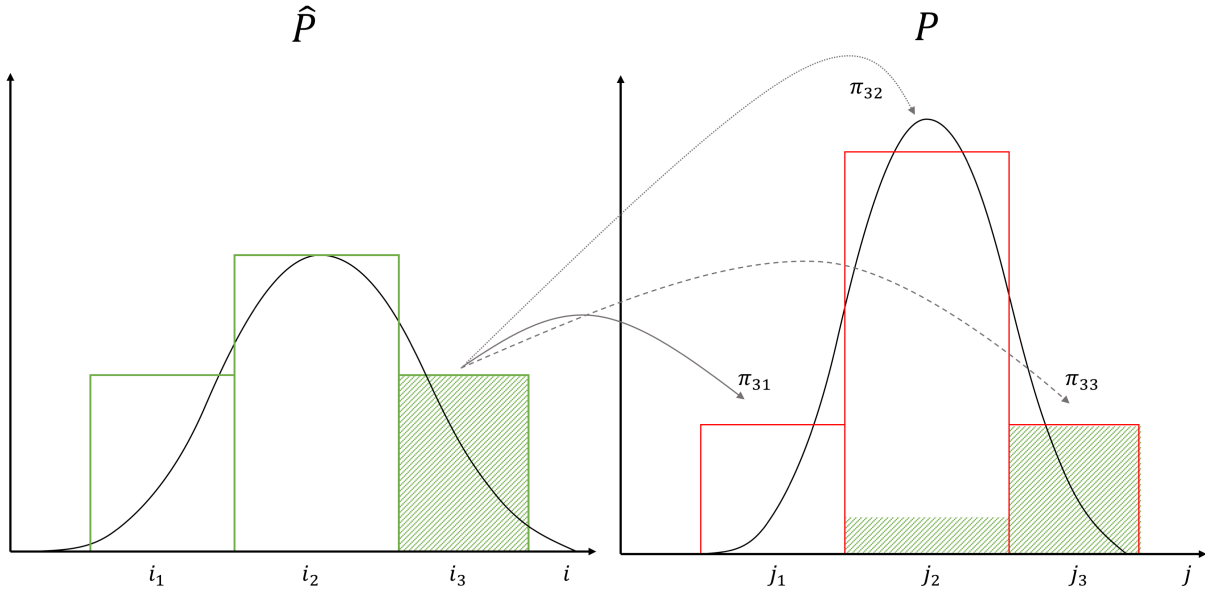


Figure 5.1: The optimal transportation plans π_{i_3j} between probability distributions \hat{P} and P with $\hat{\mathcal{N}} = \mathcal{N} = 3$.

The Wasserstein distance is appropriate for measuring distance between probability distributions. However, as we elaborate in the following section, this distance measure is insufficient when measuring the distances between scenario trees in a multistage setting, since it does not consider the stage at which information is received (Pflug and Pichler, 2014b).

A distance measure that considers whole stochastic processes is a concept introduced by Pflug and Pichler (2012) known as the nested distance. As we explain in the following subsection, this is a multistage generalization of the Wasserstein distance. It is therefore important to keep in mind that despite of increased complexity, the nested distance problem is also a mass transportation problem, where the goal is to transport the total mass at minimal costs.

5.1.2 The nested distance

The nested distance is a metric for describing the extent to which two stochastic processes, for instance scenario trees, differ. In contrast to the Wasserstein distance, nested distance considers more than just the single random variables, but a collection of these in entire stochastic processes. This allows nested distance to capture how the stochastic information evolves over time (Pflug and Pichler, 2014b). The need for using another distance measure than the Wasserstein distance to measure the distance between stochastic processes is demonstrated in the following example. First, let \mathbb{P} denote a probability model, which is a representation of a random process. To express that \mathbb{P} is composed of a probability distribution P of unconditional probabilities at the leaf stage, and a tree structure \mathbb{T} , we write $\mathbb{P} = \mathbb{P}(\mathbb{T}, P)$.

Consider the scenario trees $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$, illustrated in Figure 5.2. Denote their respective structures as $\mathbb{T}^{(1)}$ and $\mathbb{T}^{(2)}$. Since P is identical for both trees, we can express the trees as $\mathbb{P}^{(1)} = \mathbb{P}(\mathbb{T}^{(1)}, P)$ and $\mathbb{P}^{(2)} = \mathbb{P}(\mathbb{T}^{(2)}, P)$.

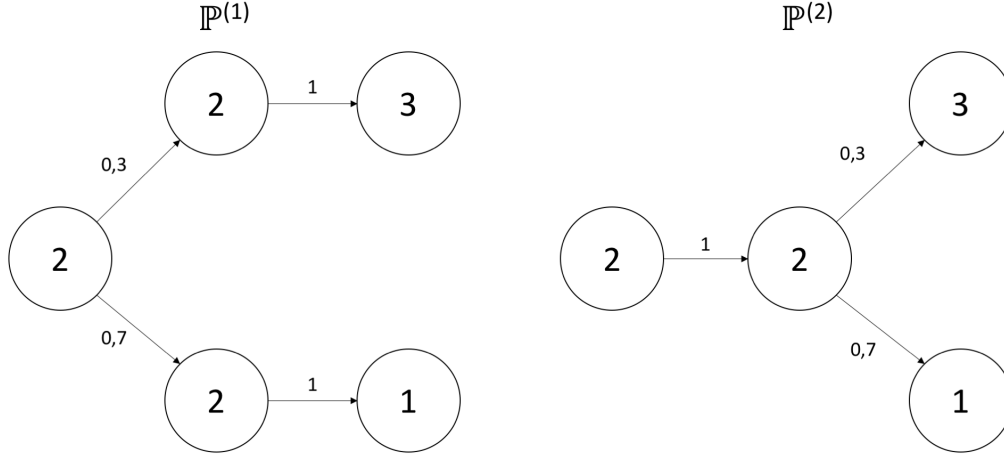


Figure 5.2: Two processes that have identical leaf node probabilities and states. The tree structures are on the other hand different, causing the nested distance to be different between the models.

Since the final stage also has identical node values, and since the Wasserstein distance does not care about the evolution of information up to this point, the Wasserstein distance between these trees is 0. On the other hand, the nested distance takes into account that in $\mathbb{P}^{(1)}$ the final observation is revealed earlier than in $\mathbb{P}^{(2)}$, hence making the nested distance non-zero. By implementing the nested distance optimization problem (5.8) to (5.12), which we explain in detail in the following paragraphs, we confirm this by showing that the nested distance $\text{dl}_1(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}) = 0.84$.

In order to establish a thorough understanding of the nested distance, we first need to define some notation. Consider the baseline probability model $\hat{\mathbb{P}}$ and the alternative model \mathbb{P} , which are trees of the same height. We denote stage t as a stage in the set of stages $T = \{0, \dots, T\}$. Let i be a leaf node in the set of leaf nodes $\hat{\mathcal{N}}_T$ in $\hat{\mathbb{P}}$, and $j \in \mathcal{N}_T$ a leaf node in \mathbb{P} . For a stage t , let k be a node in the set of nodes at this stage $\hat{\mathcal{N}}_t$ if $t \in T \setminus \{T\}$, and let k be a direct predecessor of node $i' \in \hat{\mathcal{N}}_{t+1}$. We write the direct predecessor relationship as $k = i' -$. Similarly, let $l \in \mathcal{N}_t$ and for node $j' \in \mathcal{N}_{t+1}$ let $l = j' -$. If i' is a child in the set of children of k we write $i' \in k+$. We further write $j' \in l+$ if j' is a child in the set of children of l . For any predecessor k and l that not necessarily are direct predecessors of i' and j' , we write the predecessor relationship as $k \prec i'$ and $l \prec j'$. Keep in mind that we distinguish between the leaf nodes i and j , and the successor nodes i' and j' . The two latter nodes may be intermediate nodes. When representing

paths of predecessors we index i and the preceding nodes as i as $i, i_{T-1}, i_{T-2}, \dots, i_0$. We write the scenario probabilities \hat{P}_i and P_j for i and j in $\hat{\mathcal{N}}_T$ and \mathcal{N}_T respectively.

According to Pflug and Pichler (2014b), the reason why the Wasserstein distance is not suitable for measuring scenario trees, is that it does not consider the conditional probabilities that reside within the tree structure. The nested distance problem addresses this by ensuring consistency between conditional probabilities $\hat{P}(i | k)$ and $P(j | l)$ and conditional transportation plans, or subplans $\pi(i, j | k, l)$. Following the mass transportation analogy from Section 5.1.1, the conditional probabilities can be interpreted as the available or required mass for leaf nodes i and j respectively, under the condition of being in the subtrees where k and l are the respective roots. Similarly the subplans can be interpreted as the transportation plans in these subtrees. The full transportation plan π_{ij} is an aggregation of subplans, which can be interpreted as the transportation plan between entire paths in each tree. π_{ij} is therefore defined over the leaf nodes i and j . $\pi(i, j | k, l)$ can be defined from π_{ij} as the proportion of mass transported between the paths from each respective root node down to i and j , relative to the total mass transported between all paths that end up at leaf node successors of k and l . Or more specifically this can be formulated by equation (5.7)

$$\pi(i, j | k, l) = \frac{\pi_{ij}}{\sum_{\bar{i} \succ k} \sum_{\bar{j} \succ l} \pi_{\bar{i}, \bar{j}}} \quad (5.7)$$

The nested distance problem can now be formulated by (5.8) to (5.12)

$$dl_r = \min_{\pi} \sum_{i \in \hat{\mathcal{N}}_T} \sum_{j \in \mathcal{N}_T} \pi_{ij} d_{ij}^r \quad (5.8)$$

s.t.

$$\sum_{j \succ l} \pi(i, j | k, l) = \hat{P}(i | k) \quad k \prec i, k \in \hat{\mathcal{N}}_t, l \in \mathcal{N}_t, t \in T \setminus \{|T|\} \quad (5.9)$$

$$\sum_{i \succ k} \pi(i, j | k, l) = P(j | l) \quad l \prec j, k \in \hat{\mathcal{N}}_t, l \in \mathcal{N}_t, t \in T \setminus \{|T|\} \quad (5.10)$$

$$\sum_{i \in \hat{\mathcal{N}}_T} \sum_{j \in \mathcal{N}_T} \pi_{ij} = 1 \quad (5.11)$$

$$\pi_{ij} \geq 0 \quad i \in \hat{\mathcal{N}}_T, j \in \mathcal{N}_T \quad (5.12)$$

Note that the distance measure d_{ij}^r has to be slightly modified in the nested distance objective function (5.8), such that the distance measure now considers the entire paths from the root down to leaf nodes i and j . The commonly used r -order distance, when adapted to a stochastic process, is written

$$d_r(\hat{\xi}_i, \xi_j) = \left(\sum_{t=1}^T \sum_{m=1}^M w_t^m \left| \hat{\xi}_{it}^m - \xi_{jt}^m \right|^r \right)^{1/r} \quad (5.13)$$

Even though being more complex, this problem highly resembles the Wasserstein distance problem (5.1) to (5.4). As for the Wasserstein distance constraints, constraints (5.9) to (5.11) ensure consistence in the total mass transported from the baseline distribution, and consistence in the total mass received by the alternative distribution.

As an example, consider constraints (5.9) that require the mass transported from i , conditional on k , to all leaf node successors of l to be neither more nor less than the available mass $\hat{P}(i | k)$. The similar consistence in what is received at j , conditional on l , is ensured by constraints (5.10). To further exemplify the similarities to the Wasserstein distance, consider the instance where the scenario tree only has two stages. Nodes k and l then become root nodes. Consequently, $\hat{P}(i | k) = \hat{P}(i)$ and $P(j | l) = P(j)$, and

$\pi(i, j \mid k, l) = \pi(i, j)$. Since the total mass of the distribution resides within \hat{P} and P , constraint (5.11) becomes redundant. The instance of the nested distance problem has then become equivalent to the Wasserstein distance problem.

Pflug and Pichler (2014b) also present an equivalent linear formulation of problem (5.8) to (5.12) that considers subplans only at successive stages. Even though the formulation in itself is not of importance other than being equivalent to the one presented above, it is important to understand that full transportation plans, as shown by Pflug and Pichler (2014b), can be formulated as the product of the path of direct successor subplans by relation (5.14)

$$\pi(i, j) = \pi(i, j \mid i_{T-1}, j_{T-1}) \cdot \dots \cdot \pi(i_1, j_1 \mid i_0, j_0) \quad (5.14)$$

Remark how the product of direct successor subplans, from the leaf nodes along predecessor path up to the root, resembles how the product of conditional probabilities along a similar path forms the scenario probability.

This relation is particularly useful in this thesis since central components in the multistage distributionally robust optimization problem, called transportation subkernels, closely relate to direct successor subplans. We elaborate this relation in Section 5.3.2.

For now, recall that a distributionally robust optimization problem finds the worst-case probability model from a set of probability models. Suppose then that we have a variant of the nested distance problem above, with the purpose of changing $P(j \mid l)$ in constraints (5.10), while being constrained by the nested distance. Subkernels would then, in a similar manner as subplans in the nested distance problem, be the decision variables in this problem.

In the following section we show how a distributionally robust optimization problem can be formulated with an ambiguity set that only contains scenario trees, or probability models, within a certain nested distance from the baseline probability model.

5.2 Creating a solvable problem

As introduced in Section 5.1, distributionally robust optimization problems are highly complex. Our approach towards the final goal of solving a multistage distributionally robust optimization problem therefore starts at the common ground that this problem type shares with a more familiar problem type, multistage stochastic optimization. Once the similarities and differences between multistage stochastic and distributionally robust optimization are properly explained, we proceed by formulating a general multistage distributionally robust problem. We then follow several steps for transforming the complex general distributionally robust problem into one that is manageable. Finally, we elaborate on the approach for solving the manageable distributionally robust optimization problem.

Consider first the linear multistage stochastic problem

$$\min_x \{ \mathbb{E}_{\mathbb{P}} [H(x, \xi)] : x \in \mathbb{X}, x \triangleleft \mathfrak{F}; \mathbb{P} \sim (\Omega, \mathfrak{F}, P, \xi) \} \quad (5.15)$$

$H(x, \xi)$ denotes a cost function consisting of the decision variables x , which in (5.15) is a vector with components x_t , and the random variable vector ξ with components ξ_t for each stage $t \in T$ of the stochastic process. ξ describes the values of possible outcomes. Both x_t and ξ_t are vectors on \mathbb{R}^m . The constraint $x \in \mathbb{X}$ requires all x to be within the feasible linear vector space \mathbb{X} . The uncertainty is represented by a scenario tree, which we represent with the probability model \mathbb{P} . In problem (5.15), \mathbb{P} is defined over the probability space $(\Omega, \mathfrak{F}, P, \xi)$. In brief, this formulation provides the same principal information as the definition $\mathbb{P} = \mathbb{P}(\mathbb{T}, P)$ from Section 5.1.2, stating that \mathbb{P} in addition to contain information about the uncertain values, also has information about the structure of when information is available. The sample space Ω is the set of possible outcomes, i.e. the set of paths through the tree, and the filtration $\mathfrak{F} = (\mathcal{F}_1 \dots \mathcal{F}_T)$ is the set of events, representing the possible outcomes at each stage of the decision tree. $x \triangleleft \mathfrak{F}$ denotes the non-anticipativity constraints, ensuring that x_t only depends on information available before or at stage t . P is the distribution of unconditional probabilities over the nodes in \mathcal{N}_T .

Distributionally robust optimization problems share the same basis as stochastic optimization problems in terms of seeking to find the optimal expected value of the objective function $\mathbb{E}_{\mathbb{P}}[H(x, \xi)]$. This means that the optimal decision is based according to the uncertainty provided by the probability model \mathbb{P} . As introduced in Section 5.1, the fundamental principle of distributionally robust optimization is to assume that we do not know \mathbb{P} for certain, but rather assume that it almost certainly resides within some ambiguity set \mathcal{P} of possible probability models. Since we cannot say anything about the likelihood of choosing one probability model over another, we find the decision that is optimal under the worst-case probability model from \mathcal{P} .

A factor that increases the complexity of a distributionally robust optimization problem is that it performs two counteracting actions at once. While x on one hand are decided such that the objective function value is minimized, \mathbb{P} is on the other simultaneously decided with the goal of maximizing the objective function. Any solution to this problem can therefore be regarded as the equilibrium of a game between two opposing actors, making the distributionally robust optimization problem a so-called *min-max* problem. On its general form, the distributionally robust counterpart of (5.15) is formulated

$$\min_x \max_{\mathbb{P} \in \mathcal{P}} \{\mathbb{E}_{\mathbb{P}}[H(x, \xi)] : x \in \mathbb{X}, x \triangleleft \mathfrak{F}\} \quad (5.16)$$

Following the work of Analui and Pflug (2014) and according to the discussion in Section 5.1, we in this thesis formulate a multistage distributionally robust problem that considers \mathcal{P} as a ball of possible probability models. The radius of this ball is known as the ambiguity radius ε , and is measured in nested distance around the baseline model $\hat{\mathbb{P}}$

$$\mathcal{P} := \{\mathbb{P} : \text{dl}_r(\mathbb{P}, \hat{\mathbb{P}}) \leq \varepsilon\} \quad (5.17)$$

This problem is typically non-convex, which makes us unable to guarantee that a solution exists. The problem does additionally have a size that makes it difficult to solve. In the first part of this section we therefore present the assumptions and structure that need to be in place in order to create a solvable problem.

5.2.1 Scoping down and convexifying the problem

In order to reduce the problem size, we assume the tree-structure \mathbb{T} to be fixed, such that only the scenario probabilities $P = (P_i)_{i \in \mathcal{N}_T}$ are allowed to vary. Analui and Pflug (2014) argue that the opposite case would require non-convex optimization even in the single-stage case. Hence, we from now denote the set of nodes at stage t in the baseline model or any alternative model as \mathcal{N}_t . We can then write \mathbb{P} as $\mathbb{P}(\mathbb{T}, P)$. We further rewrite (5.17) as two components. The first is the ball \mathcal{B}_ε that contains all probability distributions that satisfy the following condition: When the probability distributions are inserted into the probability model $\mathbb{P}(\mathbb{T}, \cdot)$, this probability model has a nested distance no more than ε away from the baseline probability model $\hat{\mathbb{P}}$

$$\mathcal{B}_\varepsilon := \{P : \text{dl}_r(\mathbb{P}(\mathbb{T}, P), \hat{\mathbb{P}}(\mathbb{T}, \hat{P})) \leq \varepsilon\} \quad (5.18)$$

The second component is the ambiguity set

$$\mathcal{P}_\varepsilon := \{\mathbb{P}(\mathbb{T}, P) : P \in \mathcal{B}_\varepsilon\} \quad (5.19)$$

consisting of probability models $\mathbb{P}(\mathbb{T}, P)$ where every P satisfies (5.18). This reformulation allows us to rewrite problem (5.16) as

$$\min_x \max_{\mathbb{P}(\mathbb{T}, P) \in \mathcal{P}_\varepsilon} \{\mathbb{E}_{\mathbb{P}(\mathbb{T}, P)}[H(x, \xi)] : x \in \mathbb{X}, x \triangleleft \mathfrak{F}\} \quad (5.20)$$

Even though the fixed tree structure has reduced the problem size, a remaining problem is the non-convexity of the distributionally robust problem. Recall from the introduction to Section 5.2 that solutions to distributionally robust optimization problems can be interpreted as an equilibrium strategy in a game between two actors. Whether or not we can guarantee existence of such an equilibrium relies on whether we can satisfy the min-max theorem (von Neumann, 1928) or not. In order to hold, this theorem requires

the decision space to be convex, which is complicating in our case since neither \mathcal{B}_ε nor \mathcal{P}_ε satisfy this criterion (Analui and Pflug, 2014). We proceed by presenting an approach for convexifying the decision space, which in turn allows us to guarantee the existence of an equilibrium decision.

The goal of this approach is to formulate the convex hull $\bar{\mathcal{P}}_\varepsilon$ of \mathcal{P}_ε , which is the smallest convex set containing all probability models in \mathcal{P}_ε . Under the assumption of having access to all probability models within \mathcal{P}_ε , the set of these probability models and their convex combinations form the convex relaxation of \mathcal{P}_ε , which is defined as the convex hull $\bar{\mathcal{P}}_\varepsilon$. The set of all probability models in \mathcal{P}_ε is approximated by sampling P from \mathcal{B}_ε and then adding the corresponding probability model $\mathbb{P}(\mathbb{T}, P)$ to the finite set of other sampled probability models, whose convex combinations form the approximation of $\bar{\mathcal{P}}_\varepsilon$.

Since all sampled probability models are within \mathcal{B}_ε , they satisfy (5.18). Since we want any convex combination of sampled probability models to also be within \mathcal{P}_ε , we need to ensure that any convex combination between two sampled probability models also satisfy (5.18). However, ensuring this is not a trivial task. In the following paragraphs we present an approach for formulating convex combinations of probability models, such that this criterion holds. We then rewrite the ambiguity set as its convex hull, and finally formulate the convex distributionally robust problem.

Formulating convex combinations of probability models

We want to formulate a convex combination between two probability models, such that the following criterion is satisfied: The nested distance between a baseline probability model and the convex combination has to have a nested distance no greater than the maximal nested distance between the baseline model and either of the two probability models.

Pflug and Pichler (2014a) argue that for a fixed tree structure, there are at least two approaches towards formulating such convex combinations. A first is to create a probability model from the convex combination of the scenario probabilities from two probability models within the ambiguity set. However as we show in Appendix A.2, this formulation does not satisfy the above mentioned criterion.

On the other hand, Pflug and Pichler (2014a) show that a correct way of forming these

convex combinations is by the use of compounding. Compounding can be understood as adding a root on top of the probability models in the ambiguity set and then assigning probabilities λ for each probability model. For two probability models $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$, the compound \mathcal{C} is given by

$$\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda) = \begin{cases} \mathbb{P}^{(1)} & \text{with probability } \lambda, \\ \mathbb{P}^{(2)} & \text{with probability } 1 - \lambda. \end{cases} \quad (5.21)$$

$\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ is then a tree of height $T + 1$. Since the nested distance compares trees of the same height, we need to denote $\mathbb{P}_+^{(0)}$ as a third probability model, which is the baseline model $\mathbb{P}^{(0)}$ with an additional root node. These two trees are illustrated in Figure 5.3

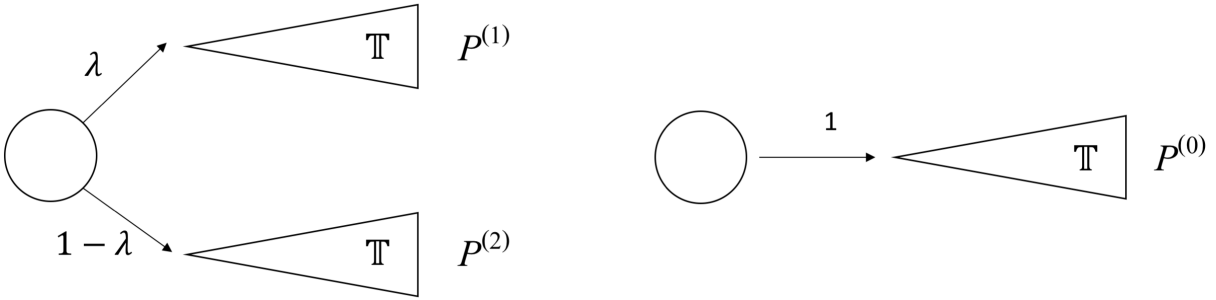


Figure 5.3: *Left:* The compound tree $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$, *Right:* The augmented tree $\mathbb{P}_+^{(0)}$

Following (A.1), we now have to show that the nested distance between $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ and $\mathbb{P}_+^{(0)}$ is no more than the convex combination of $\text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})$ and $\text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(2)})$. To exemplify why this is relevant in the case of the ambiguity set, consider the convex combination of two nested distances between the baseline model $\mathbb{P}^{(0)}$ and any two probability models $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$ in the ambiguity set. If this convex combination is no greater than the nested distance between the baseline model $\mathbb{P}_+^{(0)}$ and $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$, then $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ is also in the ambiguity set. This example is illustrated this in the Figure 5.4.

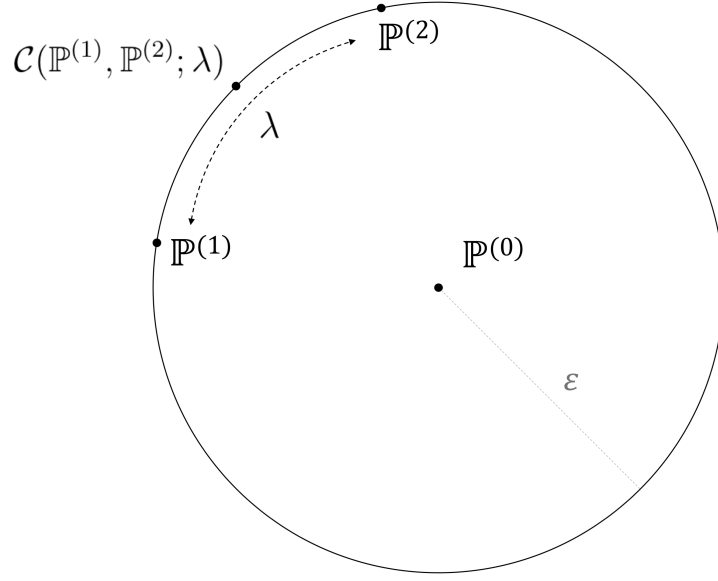


Figure 5.4: If the nested distance $\mathbf{dl}_r(\mathbb{P}_+^{(0)}, \mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda))$ is less than or equal to the convex combination of $\mathbf{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})$ and $\mathbf{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(2)})$ then we know that also $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ is within the ambiguity set.

More specifically, we need to satisfy

$$\mathbf{dl}_r(\mathbb{P}_+^{(0)}, \mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda))^r \leq \lambda \mathbf{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})^r + (1 - \lambda) \mathbf{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(2)})^r \quad (5.22)$$

Pflug and Pichler (2014a) prove that equation (5.22) holds by showing that there exists a valid transportation plan such that $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ equals the right hand side. In the proof, π_1 is denoted as the optimal transportation plan for $\mathbf{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})$, and π_2 is the optimal transportation plan for $\mathbf{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(2)})$. Recall that a full transportation plan can, by equation (5.14), be formulated as the product of subplans. Since the subtrees of $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ are identical to $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$, then the optimal transportation subplans between the subtrees are equal to π_1 and π_2 . By taking the product of root probabilities λ and $1 - \lambda$, and π_1 and π_2 respectively, we get the valid transportation plan $\lambda \pi_1$ for transportation between $\mathbb{P}^{(0)}$ and the $\mathbb{P}^{(1)}$ -subtree of $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$, and $(1 - \lambda) \pi_2$ for transportation between $\mathbb{P}^{(0)}$ and the $\mathbb{P}^{(2)}$ -subtree of $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$.

Since the values of the subtree root nodes in $\mathcal{C}(\mathbb{P}^{(1)}, \mathbb{P}^{(2)}; \lambda)$ are equal, the only path distances that impact the nested distance (5.8) are those between the subtrees. Further, since the transportation plans are identical to the weighted transportation plans on the right hand side of equation (5.22), the two terms are equal. Since this is a valid transportation plan, the optimal transportation plan can only be better. This proves that (5.22) holds.

We have shown that compounding can be used to formulate convex combinations between probability models. We now show how we can formulate the convex hull of the ambiguity set as a compound tree made of a finite set of probability models, such that we can create a distributionally robust problem where the existence of an equilibrium solution can be guaranteed.

Formulating the convex hull of the ambiguity set

We convexify the ambiguity set \mathcal{P}_ε by approximating the convex hull $\bar{\mathcal{P}}_\varepsilon$ with a finite set of sampled probability models from \mathcal{B}_ε . This is achieved by formulating $\bar{\mathcal{P}}_\varepsilon$ as a compound tree, to which each sampled probability model is added. Since we now know that convex combinations of probability models within the compound tree also are within \mathcal{B}_ε , we can regard the compound as the continuous image of all probability models within $\bar{\mathcal{P}}_\varepsilon$.

Since we consider a fixed tree structure, our sampled probability models are actually sampled scenario probability distributions P from the family of all probability distributions on \mathcal{N}_T . In addition, all sampled P have to be valid in terms of satisfying (5.18), such that they reside in \mathcal{B}_ε . We denote the probability distribution from which we can sample valid probability distributions as Λ . We then denote the compound $\mathcal{C}(\mathbb{P}(\mathbb{T}, \cdot), \Lambda)$ as the tree of probability models formed by all sampled P

$$\mathcal{C}(\mathbb{P}(\mathbb{T}, \cdot), \Lambda) = \mathbb{P}(\mathbb{T}, P), \quad \text{where } P \text{ is distributed according to } \Lambda \quad (5.23)$$

The convex hull $\bar{\mathcal{P}}_\varepsilon$ of \mathcal{P}_ε contains all probability models within \mathcal{P}_ε . The compound structure ensures that any convex combination of probability models is within \mathcal{P}_ε , making $\bar{\mathcal{P}}_\varepsilon$ of \mathcal{P}_ε the convex relaxation of \mathcal{P}_ε . This is defined as the hull $\bar{\mathcal{P}}_\varepsilon$. If we consider a

finite set of k probability distributions, the compound tree of (5.23) can be illustrated as in Figure 5.5.

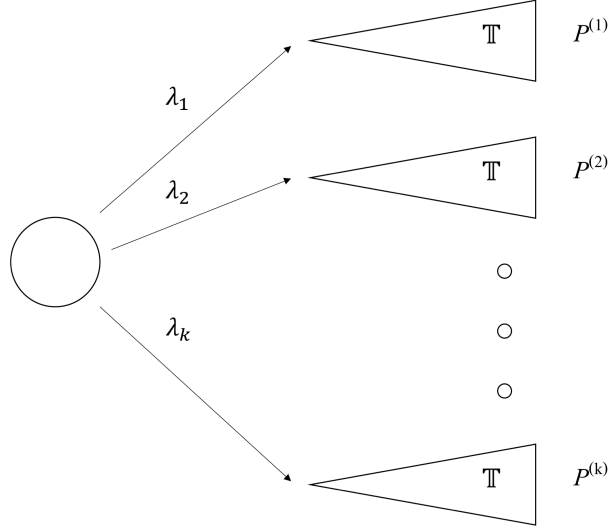


Figure 5.5: The convex hull of the ambiguity set is a compound, here consisting of k sampled probability models and the convex combinations between them.

Since we know that $\bar{\mathcal{P}}_\varepsilon$ is constructed by sampling probability distributions from \mathcal{B}_ε , and since the samples from \mathcal{B}_ε form elements in the convex decision space of $\bar{\mathcal{P}}_\varepsilon$, the non-convexity of \mathcal{B}_ε does not affect the decision space.

We rewrite the ambiguity set as the convex hull of (5.19) that samples distributions from

$$\mathcal{B}_\varepsilon := \{P : \text{dl}_r(\mathbb{P}(\mathbb{T}, P), \hat{\mathbb{P}}(\mathbb{T}, \hat{P})) \leq \varepsilon\}$$

as

$$\bar{\mathcal{P}}_\varepsilon = \{\mathcal{C}(\mathbb{P}(\mathbb{T}, \cdot), \Lambda) : \Lambda \text{ is a probability distribution on } \mathcal{B}_\varepsilon\} \quad (5.24)$$

Since (5.24) is convex, we now have a decision space that satisfies the min-max theorem.

We can therefore reformulate the distributionally robust problem (5.20) such that we can guarantee the existence of an equilibrium solution. We can then write the convexified distributionally robust problem as

$$\min_{x \in \mathbb{X}} \max_{\mathbb{P}(\mathbb{T}, \mathcal{P}) \in \tilde{\mathcal{P}}_\varepsilon} \{ \mathbb{E}_{\mathbb{P}(\mathbb{T}, \mathcal{P})} [H(x, \xi)] : x \triangleleft \mathfrak{F} \} \quad (5.25)$$

5.3 The solution algorithm

As we elaborate in detail in this section, we approximate problem (5.25) by using a successive convex programming algorithm, which is a heuristic for approximating a complex optimization problem by iteratively solving a series of smaller optimization problems. Our approach is based on the work of Analui and Pflug (2014), whose proposed solution algorithm is best understood as an almost direct implementation of the *game versus nature*-analogy from Section 3.3, which is as a game between of two actors. This procedure is thoroughly explained in Algorithm 5.1. The first actor is the decision maker, who with the outer optimization problem (5.26) aims to find the $x \in \mathbb{X}$ that minimizes the expected value of the objective function with respect to the worst-case distribution in the ambiguity set. The optimal outer problem decision is then handed over to the other actor, 'nature'. Based on this optimal decision, 'nature' does in the inner problem (5.27) find a new probability model that maximizes the expected value of the objective function. This probability model is then added to the ambiguity set.

By iterating between the outer and inner problem, new worst-case probability distributions are gradually added to the ambiguity set, such that the actual ambiguity set (5.24) is approximated by a finite set.

Optimality of the algorithm In contrast to Algorithm 5.1, the distributionally robust problem (5.25) solves the outer and inner problem simultaneously. A necessary and sufficient condition for finding the joint optimum decision is that x and \mathbb{P} have to be chosen simultaneously (Analui and Pflug, 2014). As a consequence, even though we can guarantee the existence of an equilibrium solution based on a worst-case distribution within \mathcal{P}_ε , we cannot guarantee that this is the joint optimum solution.

An intuitive interpretation of this condition can be made by considering an analogy to economic game theory. Recall that the optimal solution in a Cournot duopoly is based on the assumption that the two actors make their decisions simultaneously. This optimal solution is different from that of a sequential game, such as the Stackleberg model, where one actor benefits from moving either before or after the other, depending on the type of game (Pindyck and Rubinfeld, 2014).

Presenting the algorithm We now present the algorithm, and elaborate on how we can formulate its outer and inner problem as linear programs.

Algorithm 5.1 Successive convex programming algorithm

- **INITIALIZATION.** Set iteration $k = 1$ and determine the value of ε . Let the initial ambiguity set be the baseline model, $\mathcal{P}_\varepsilon^{(k)} = \{\mathbb{P}\}$
- **OUTER OPTIMIZATION.** Solve the outer problem

$$\begin{aligned}
 & \min_{u,x} u \\
 & \text{s.t.} \\
 & \mathbb{E}_{\mathbb{P}}[H(x, \xi)] \leq u \qquad \mathbb{P} \in \mathcal{P}_\varepsilon^{(k)} \\
 & x \in \mathbb{X} \\
 & x \triangleleft \tilde{\mathcal{F}}
 \end{aligned} \tag{5.26}$$

resulting in the solution $(x^{(k)}, u^{(k)})$. If the solution is not unique, choose any solution in the set of feasible solutions.

- **INNER OPTIMIZATION.** Fix $x^{(k)}$ and solve the inner problem

$$\begin{aligned}
 & \max_P \mathbb{E}_{\mathbb{P}(\mathbb{T}, P)}[H(x^{(k)}, \xi)] \\
 & \text{s.t.} \\
 & \mathbb{P}(\mathbb{T}, P) \in \mathcal{P}_\varepsilon^{(k)}
 \end{aligned} \tag{5.27}$$

Call the solution $\mathbb{P}^{(k)}$ and include it in the ambiguity set for the next iteration, $\mathcal{P}_\varepsilon^{(k+1)} = \mathcal{P}_\varepsilon^{(k)} \cup \{\mathbb{P}^{(k)}\}$. If the solution is not unique, choose any solution in the set of feasible solutions.

- **STOPPING CRITERION.** A stopping criterion can either be chosen to be when there is no improvement in the maximin solution, $u^{(k+1)} - u^{(k)} \leq \theta$, or by in advance defining a number of iterations k . Otherwise, set $k := k + 1$ and go to **OUTER OPTIMIZATION**.
-

The outer problem (5.26) is a slightly modified version of the multistage stochastic problem (5.15), where u represents the minimum solution for the worst-case probability model.

As opposed to the outer problem (5.26), which is quite easy to formulate, two modifications have to be made to the inner problem (5.27) in order to make it solvable. The first modification is to formulate P as a set of variables, such that a worst-case probability model $\mathbb{P}(\mathbb{T}, P)$ within the nested distance ball (5.18), can be found by solving an optimization problem. In the following Section 5.3.1 we show how this can be done by the use of transportation kernels. However, since the resulting formulation is non-linear, a second modification has to be made to ensure that this problem can be approximated linearly. In section 5.3.2 we therefore present an iterative algorithm for approximating the solution of (5.27) by solving a series of linear programs. In Section 5.3.3 we make a set of adaptations to the solution algorithm, so that it can be applied to a financial decision making context.

Since these modifications rely on transportation kernels, we first establish a clear understanding of this concept.

5.3.1 Transportation kernels

A transportation kernel $K(i, j)$ from node i to j in \mathcal{N}_T can be interpreted as a modified version of the optimal transportation plan π_{ij} in the nested distance problem (5.8) to (5.12). The reason for using transportation kernels, or kernels for short, is that they allow us to create a new probability distribution P from the baseline probability distribution \hat{P} with the following relation

$$P(j) = \sum_{i \in \mathcal{N}_T} K(i, j) \cdot \hat{P}(i) \quad (5.28)$$

If we now consider kernels as decision variables, then any alternative distribution P can be formulated from \hat{P} . This principle is the basis for the kernel reformulation of inner problem (5.27). This requires us to formulate the entire scenario tree structure with kernels. We elaborate how this is done in the remainder of this section.

Similar to transportation plans, kernels contain information about conditional probabilities as they also are composed of sub-elements. These subelements are known as subkernels $K_t(j'|i'; k, l)$, and are defined for $t \in T \setminus \{|T|\}$. Recall that a transportation subplan can be viewed as the amount of mass transported from one subtree to another. Similarly, a direct successor subplan can be interpreted as the mass transported between any subtrees consisting of only two stages. As mentioned in Section 5.1.2 direct successor subplans closely relate to subkernels. An intuitive definition of subkernels is that they represent the proportion of mass transported from node $i' \in k+$ to $j' \in l+$, relative to what is transported from i' to the other children in $l+$. Thus, where a direct successor subplan is an absolute measure of transported mass, the subkernel measures a proportion. Subkernels are defined as

$$K_t(j'|i'; k, l) = \frac{\pi(i', j'|k, l)}{\sum_{j \in l+} \pi(i', j|k, l)} \quad (5.29)$$

Since a single subkernel is a proportion, all subkernels $K_t(\cdot|i'; k, l)$ form a probability distribution on the set $l+$, measuring the proportions of the mass at i' , conditional on k and l , that is transported to each successor of l . Hence, all $K_t(j'|i'; k, l)$ are positive values that collectively sum to 1

$$\begin{aligned} \sum_{j \in l+} K_t(j|i'; k, l) &= 1 && i' \in k+, (k, l) \in \mathcal{N}_t, t \in T \setminus \{|T|\} \\ K_t(j'|i'; k, l) &\geq 0 && i \in k+, j \in l+, k, l \in \mathcal{N}_t, t \in T \setminus \{|T|\} \end{aligned} \quad (5.30)$$

We stated above that $K_t(\cdot|i'; k, l)$ are relative measures of the mass transported from i' , conditional on the direct predecessors k and l . Hence, we find the actual transported mass by taking the product of the transported proportion and the available mass, $K_t(j|i'; k, l) \cdot \hat{P}(i'|k)$. By inserting this into the relation between the full transportation plan and its subplan path (5.14), we get the relation between subkernels and the full transportation plan $\pi(i, j)$

$$\begin{aligned} \pi(i, j) = & K_0(j_1|i_1; i_0, j_0) \cdot K_1(j_2|i_2; i_1, j_1) \cdot \dots \cdot K_{T-1}(j|i; i_{T-1}, j_{T-1}) \\ & \cdot \hat{P}(i_1|i_0) \cdot \hat{P}(i_2|i_1) \cdot \dots \cdot \hat{P}(i|i_{T-1}) \end{aligned} \quad (5.31)$$

The product of subkernels in (5.31) can be interpreted as a measure of the proportions of transported mass along the paths from the roots down to the leaf nodes i and j . This composition of subkernels is what defines a transportation kernel for i and j .

$$K(i, j) = K_0(j_1|i_1; i_0, j_0) \cdot K_1(j_2|i_2; i_1, j_1) \cdot \dots \cdot K_{T-1}(j|i; i_{T-1}, j_{T-1}) \quad (5.32)$$

Substituting (5.32) into (5.31) and setting the product of the path of conditional probabilities as the unconditional probability $\hat{P}(i)$, we get the relation between transportation kernels and optimal transportation plans

$$\pi(i, j) = K(i, j) \cdot \hat{P}(i) \quad (5.33)$$

The relationships above are illustrated thoroughly in the following example, where we show how the probability distribution $P^{(2)}$ from the leaf nodes of Figure 5.6 can be constructed from $P^{(1)}$ and the optimal transportation subplan.

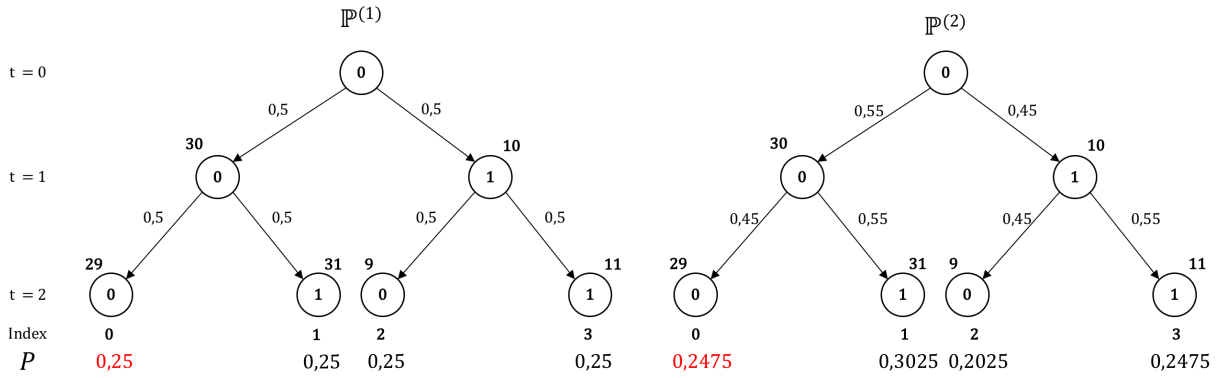


Figure 5.6: Two trees with same structure \mathbb{T} and different probabilities. Unconditional probabilities are shown below the leaf nodes. Conditional probabilities are shown along arcs.

By using problem (5.8) - (5.12) to calculate $dl_1(\mathbb{P}(\mathbb{T}, P^{(1)}), \mathbb{P}(\mathbb{T}, P^{(2)}))$ for the trees illustrated in Figure 5.6, we first find the optimal transportation plans. We then use the transportation plan-subplan relation (5.7) to calculate the optimal subplans. The transportation plans and subplans are presented in Table 5.1.

$t = 1$		$t = 2$		$\pi(k+i, l+j k, l)$			
k, l	$\pi(k, l 0, 0)$	$k+i, l+j \setminus k, l$	00	01	10	11	
00	0,5	00	0,45	0,45	0,45	0,45	
01	0	01	0,05	0,05	0,05	0,05	
10	0,05	10	0	0	0	0	
11	0,45	11	0,5	0,5	0,5	0,5	

Table 5.1: Optimal transportation subplans the example in Figure 5.6, where $(k, l) \in \mathcal{N}_1$. $k+i$ and $l+j$ are the direct successors of k and l , indexed from the left.

By using the relation between kernels and subplans (5.29) we can construct the corresponding subkernels, presented in the following Table 5.2

$\mathbf{t} = 1$		$\mathbf{t} = 2$		$K_1(l +_j k+_i; k, l)$			
l, k	$K_0(l k; 0, 0)$	$l+_j, k+_i \setminus k, l$	00	01	10	11	
00	1	00	0,9	0,9	0,9	0,9	
10	0	10	0,1	0,1	0,1	0,1	
01	0,1	01	0	0	0	0	
11	0,9	11	1	1	1	1	

Table 5.2: Transportation Kernels created from the optimal transportation subplans of Figure 5.1.

We can now use the kernel-subkernel relation (5.32) to form kernels $K(i, 0)$ for $i \in \mathcal{N}_T$. The results are presented in the Table 5.3 below

Kernel index	Kernel Value	Subkernel path values	Subkernel path
$K(0, 0)$	0,9	$1 \cdot 0,9$	$K_0(l_0 k_0; 0, 0) \cdot K_1(l+_0 k+_0; k_0, l_0)$
$K(1, 0)$	0	$1 \cdot 0$	$K_0(l_0 k_0; 0, 0) \cdot K_1(l+_0 k+_1; k_0, l_0)$
$K(2, 0)$	0,09	$0,1 \cdot 0,9$	$K_0(l_0 k_1; 0, 0) \cdot K_1(l+_0 k+_0; k_1, l_0)$
$K(3, 0)$	0	$0,1 \cdot 0$	$K_0(l_0 k_1; 0, 0) \cdot K_1(l+_0 k+_1; k_1, l_0)$

Table 5.3: Kernels $K(i, 0)$, $i \in \mathcal{N}_T$ made by the composition of subkernel paths from the root down to the leaf nodes.

Finally, we can use the relation between transportation kernels and probability distributions (5.28) to compute $P^{(2)}(0)$

$$P^{(2)}(0) = \sum_{i \in \mathcal{N}_T} K(i, 0) \cdot P^{(1)}(i)$$

$$0,2475 = 0,25 \cdot 0,9 + 0,25 \cdot 0 + 0,25 \cdot 0,09 + 0,25 \cdot 0$$

Which shows that the leftmost leaf node probability in tree $\mathbb{P}^{(2)}$ can be created from the leaf node probabilities in tree $\mathbb{P}^{(1)}$ and a set of transportation kernels. The process can then be repeated for remaining $j \in \mathcal{N}_T \setminus \{0\}$.

Using kernels to find conditional probabilities

The kernel formulations by Analui and Pflug (2014) described in the section above only show formulations of the unconditional scenario probabilities, we propose a formulation of conditional probabilities with transportation kernels. As we elaborate in Section 6, this allows us to formulate an objective function that makes use of time consistent CVaR.

We therefore propose a formulation of conditional probabilities with transportation kernels. The idea is to construct the conditional probability $P(j | l)$ in the same manner as the scenario probability $P(j)$ is done in the whole tree, but here to consider the subtree with node l as root node.

This implies that when we create a transportation kernel as the product of a subkernel path with relation (5.32), we only consider the subkernels at stages succeeding the stage of node l . We denote this transportation kernel $K'(i, j)$. For node k in the initial tree and l in the new tree, both at \mathcal{N}_t , we now present an example showing that conditional probabilities can be formulated by subkernels with the following relation

$$P(j | l) = \sum_{i \in \mathcal{N}_T} K'(i, j) \cdot \hat{P}(i | k) \quad (5.34)$$

which expressed with subkernels is formulated according to (5.32) as

$$K'(i, j) = K_t(j_{t+1} | i_{t+1}; k, l) \cdot \dots \cdot K_{T-1}(j | i; i_{T-1}, j_{T-1})$$

Consider again Figure 5.6. By using the subtree of $\mathbb{P}^{(1)}$ with node 0 at stage $t = 1$ as root, we can find the conditional probability of reaching leaf node 0 node in tree $\mathbb{P}^{(2)}$, which is 0,45, by using kernels and the conditional probability of reaching node 0 in $\mathbb{P}^{(1)}$.

Since we now consider the a subtree with only two stages, the kernel-subkernel path relation (5.32) shows that the subkernel path only consists of a single subkernel. We can thus write the subtree transportation kernels as $K'(i, j) = K_1(j | i; k, l)$.

In the same manner as in the previous example, we calculate the optimal transportation subplan and subkernels, conditional on node k and l at stage $t = 1$, by applying (5.7) and

(5.29) respectively.

t = 2			
k, l	$\pi(k, l 0, 0)$	j, i	$K_1(j i; 0, 0) = K'(i, j)$
00	0,45	00	0,9
01	0,05	10	0
10	0	01	0,1
11	0,5	11	1

Table 5.4: Subkernels and optimal transportation subplan conditional on nodes 0 at stage $t = 1$ in Figure 5.6.

By then using the relation between kernels and probability distributions (5.28) we can show that for node k_0 and l_0 at stage $t = 1$, and for i_0 and j_0 at stage $t = 2$, that

$$\begin{aligned}
 P^{(2)}(j_0 | l_0) &= \sum_{i \in \mathcal{N}_T} K_1(j_0 | i_0; k_0, l_0) \cdot P^{(1)}(i_0 | k_0) \\
 0,45 &= 0,5 \cdot 0,9 + 0,5 \cdot 0
 \end{aligned} \tag{5.35}$$

Similarly for i_1 and j_1 conditional k_0 and l_0 we get

$$0,55 = 0,5 \cdot 1 + 0,5 \cdot 0,1$$

which can be confirmed in Figure 5.6 as the correct conditional probabilities.

With an understanding of transportation kernels, we can proceed to reformulate the distributionally robust problem (5.25).

5.3.2 A multistage distributionally robust problem with transportation kernels

In this section we show how the multistage distributionally robust problem (5.25) can be solved, by using transportation kernels to find the worst-case probability distribution in the inner problem (5.27). We then present an algorithm for solving the inner problem as a sequence of linear programs.

The purpose of the inner problem is to find the probability distribution that maximizes the expected cost function $\mathbb{E}_{\mathbb{P}(\mathbb{T}, P)}[H(x^{(k)}, \xi)]$. This is done by setting transportation kernels and subkernels as decision variables, and by using the relations from Section 5.3.1 as constraints, which ensure that the inner problem outputs a valid worst-case probability model.

When formulating the inner problem with kernels, Analui and Pflug (2014) write the relations from Section 5.3.1 on a more compact format. They denote the kernel-probability distribution relation (5.28) as $P = K \circ \hat{P}$, where K is the $\mathcal{N}_T \times \mathcal{N}_T$ matrix of transportation kernels $K(i, j)$. They also set K_t as the $\mathcal{N}_T \times \mathcal{N}_T$ matrix of subkernels $K_t(j'|i'; k, l)$ at stage t , that precede leaf nodes i and j . The kernel-subkernel path relation (5.32) is written $K = K_0 \circ \dots \circ K_{T-1}$.

Recall relation (5.33), which is written

$$\pi(i, j) = K(i, j) \cdot \hat{P}(i)$$

An important observation is that we can reformulate the nested distance objective function (5.8)

$$dl_r = \sum_{i \in \mathcal{N}_T} \sum_{j \in \mathcal{N}_T} \pi_{ij} d_{ij}^r$$

to include kernels by substituting π_{ij} with the right hand side of (5.33) and get

$$dl_r = \sum_{i,j \in \mathcal{N}_T} d_{ij}^r K(i,j) \hat{P}(i) \quad (5.36)$$

The multistage multistage distributionally robust problem (5.25) can then be rewritten with transportation kernels as

$$\min_{x \in \mathbb{X}} \max_{K, K_t} \{ \mathbb{E}_{\mathbb{P}(\mathbb{T}, K \circ \hat{P})} [H(x, \xi)] \text{ s.t. } K = K_0 \circ \dots \circ K_{T-1}, \sum_{i,j \in \mathcal{N}_T} d_{ij}^r K(i,j) \hat{P}(i) \leq \varepsilon \} \quad (5.37)$$

It should be noted that the relation $K = K_0 \circ \dots \circ K_{T-1}$ is non-linear due to the product of subkernels, which are all variables. Analui and Pflug (2014) propose the following Algorithm 5.2 to linearly approximate the optimal solution of the inner problem (5.27).

Algorithm 5.2 Stepwise linearization of the inner problem (5.27)

At iteration k in Algorithm 5.1 the inner problem (5.27) has to be solved. Suppose that $P^{(k-1)}$, the worst-case probability distribution from the previous step, is on the form $P^{(k-1)} = K \circ \hat{P}_0^{(old)} \circ \dots \circ K_{T-1}^{(old)}$.

If $\mathbf{k} = \mathbf{0}$: Let $K_t^{(old)}$ be the identity matrix, i.e. $K_t^{(old)}(i'|j'; k, l) = 1$ if $i' = j'$, otherwise 0.

STAGEWISE ITERATION:

- For $t = 0$ to $t = T - 1$ solve

$$\max_{K, K_t} \mathbb{E}_{\mathbb{P}(\mathbb{T}, K \circ \hat{P})} [H(x^{(k)}, \xi)] \quad (5.38)$$

s.t.

$$K = K_0^{(new)} \circ \dots \circ K_{t-1}^{(new)} \circ K_t \circ K_{t+1}^{(old)} \circ \dots \circ K_{T-1}^{(old)} \quad (5.39)$$

$$dl_r(\mathbb{P}(\mathbb{T}, K \circ \hat{P}), \mathbb{P}(\mathbb{T}, \hat{P})) \leq \varepsilon \quad (5.40)$$

then set $K_t^{(new)}$ as the solution K_t^* .

Update the worst-case probability distribution $P^{(k)} = K \circ \hat{P}_0^{(new)} \circ \dots \circ K_{T-1}^{(new)}$

Linearity in Algorithm 5.2 is ensured by only setting subkernels at stage t as variables within an iteration. The inner problem (5.27) is then approximated by iterating through all stages from the root down to the leaf nodes.

All other subkernels are set as constants, either as old constants $K_{t'}^{(old)}$ from $P^{(k-1)}$ for every t' such that $t' > t$, or as new constants $K_{t'}^{(new)}$ from $P^{(k)}$ if $t' < t$. An approach for achieving this is to initialize each iteration (k, t) from Algorithms 5.2 and 5.1 respectively, with a constant $C_{ji}^{(k,t)}$ such that

$$C_{ij}^{(k,t)} = K_0^{(old)}(j_1|i_1; i_0, j_0) \cdot \dots \cdot K_{t-1}^{(old)}(j_t|i_t; i_{t-1}, j_{t-1}) \\ \cdot K_{t+1}^{(new)}(j_{t+2}|i_{t+2}; i_{t+1}, j_{t+1}) \cdot \dots \cdot K_{T-1}^{(new)}(j|i; i_{T-1}, j_{T-1}) \quad (5.41)$$

Constraints (5.39) for iteration (k, t) is then linearized as

$$K(i, j) = C_{ij}^{(k,t)} K_t(j'_{t+1}|i'_{t+1}; k, l) \quad i, j \in \mathcal{N}_T \quad k, l \in \mathcal{N}_t$$

In addition we have to add (5.30) as a constraint to ensure that all mass that is transported from i' to all j' , conditional on k and l , adds up to one

$$\sum_{j \in l+} K_t(j|i'; k, l) = 1 \quad i' \in k+, (k, l) \in \mathcal{N}_t, t \in T \setminus \{|T|\}$$

Constraint (5.40), which ensures that the nested distance from the baseline model $\hat{\mathbb{P}}$ is no more than ε is in the general distributionally robust problem (5.37) written

$$\sum_{j \in \mathcal{N}_T} \sum_{i \in \mathcal{N}_T} d_{ij}^r K(i, j) \hat{P}(i) \leq \varepsilon$$

This constraint already linear and can therefore be applied directly.

In addition, we have to ensure that any alternative distribution $P(j)$ from the kernel-probability distribution relation (5.28) is a valid probability distribution, by constraining it to sum to 1

$$\sum_{j \in \mathcal{N}_T} \sum_{i \in \mathcal{N}_T} K(i, j) \cdot \hat{P}(i) = 1$$

Finally, the inner problem objective function (5.38) can be linearized by inserting (5.28), such that we get

$$\max_{K(i,j)} \sum_{j \in \mathcal{N}_T} \sum_{i \in \mathcal{N}_T} K(i, j) \cdot \hat{P}(i) \cdot x_j^{(k)} \quad (5.42)$$

The inner problem objective function (5.38) is in Analui and Pflug (2014) the expected objective value at the final stage of the stochastic process. Specifically, the objective is formulated as the product of scenario probabilities and accumulated cost function values at the leaf stage. We denote the accumulated cost function value at the leaf stage for scenario j as H_{jT} . The objective is then written as

$$\max_{K(i,j), K_t(j'|i';k,l)} \sum_{i,j \in \mathcal{N}_T} H_{jT} K(i, j) \hat{P}(i)$$

We have now seen how the multistage distributionally robust problem has been made solvable from the general formulation, and the assumptions that apply. We have also seen how this problem can be solved by iteratively solving two counteracting subproblems. To ensure a correct understanding of the problem formulation and solution method, we have implemented the inventory control example presented in Analui and Pflug (2014). The results correspond well with the theory, and are explained in detail in Appendix A.3.

A final problem that has to be overcome, is to ensure that the distributionally robust model can be formulated with the conditional value at risk (CVaR) risk measure, which is

the risk measure we apply in the hedging problem in Section 6.2.2. In the following section we show that in order to write a linear formulation of the inner problem when CVaR is included, we need to define a new optimization problem within the inner optimization problem.

5.3.3 The CVaR optimization problem

Recall that the convex distributionally robust optimization problem with fixed tree structure is written as

$$\min_{x \in \mathbb{X}} \max_{\mathbb{P}(\mathbb{T}, P) \in \bar{\mathcal{P}}_\varepsilon} \{ \mathbb{E}_{\mathbb{P}(\mathbb{T}, P)} [H(x, \xi)] : x \triangleleft \mathfrak{F} \} \quad (5.43)$$

where $H(x, \xi)$ is a cost function which we seek to minimize. In financial applications it is common to consider the objective as a utility function, which trades off expected costs and its related risk. We therefore expand the objective function in (5.43) such that the disutility from expected costs and the related risk is minimized. CVaR is used as risk measure.

Recall from Section 3.2 that CVaR is defined as the average costs exceeding the Value at Risk (VaR), where VaR is defined as the threshold that costs shall not exceed with a given confidence level α . Since we in a distributionally robust problem consider the probability model as a variable, CVaR has to be evaluated for varying probability distributions in addition to varying decisions. CVaR is therefore a function of both the probability model and the cost function. A formulation of CVaR, based on the work of Rockafellar and Uryasev (2000), can be written as the following optimization problem

$$\text{CVaR}_\alpha(H(x, \xi), \mathbb{P}(\mathbb{T}, P)) = \min_{\tau} \tau + \frac{1}{1 - \alpha} \mathbb{E}_{\mathbb{P}(\mathbb{T}, P)} [H(x, \xi) - \tau]_+ \quad (5.44)$$

The costs exceeding VaR, i.e. the shortfall costs, are here represented as the positive difference between $H(x, \xi)$ and the auxiliary variable τ . When solved to optimality τ is

the Value at Risk. By formulating the utility function as the disutility caused by expected costs and the related risk, the objective in problem (5.43) is reformulated as

$$\mathbb{E}_{\mathbb{P}(\mathbb{T}, P)}[U(x, \xi)] = \mathbb{E}_{\mathbb{P}(\mathbb{T}, P)}[H(x, \xi)] + \text{CVaR}_\alpha(H(x, \xi), \mathbb{P}(\mathbb{T}, P))$$

which transforms problem (5.43) to

$$\min_{x \in \mathbb{X}} \max_{\mathbb{P}(\mathbb{T}, P) \in \bar{\mathcal{P}}_\epsilon} \{\mathbb{E}_{\mathbb{P}(\mathbb{T}, P)}[U(x, \xi)] : x \triangleleft \mathfrak{F}\} \quad (5.45)$$

Formulating CVaR linearly in standard maximization or minimization problems, where the probability distribution is known, can easily be done. The procedure is for instance well documented in Zenios and Markowitz (2008). Including CVaR in the min-max problem (5.43) is on the other hand a complicated procedure. This is because the inner problem turns the probability model into a variable, which when solved simultaneously as the CVaR problem (5.44) causes the expected shortfall costs $\mathbb{E}_{\mathbb{P}(\mathbb{T}, P)}[H(x, \xi) - \tau]_+$ to be a product of two variables. This makes the optimization problem non-linear and hence difficult to solve.

A solution is to exploit that CVaR is an optimization problem, and that this makes problem (5.45) a min-max-min problem. By solving for one subproblem at the time and keeping all other variables fixed, this problem can be approximated in the same successive manner as in Algorithm 5.1¹. In the following section we explain that this can be done in practice by introducing a CVaR subproblem. By applying an iterative procedure between this and the inner problem, we ensure that the worst-case probability model is calculated for a correct value of CVaR.

¹We are grateful to Alois Pichler who pointed this out under a meeting in Trondheim in May 2018

Adapting the solution algorithm to the CVaR optimization problem

Integrating CVaR or any other risk measures to the solution algorithm of Analui and Pflug (2014) has to our knowledge not been covered in the literature. In this section we therefore present our proposed solution method.

In brief, this approach attempts to approximate the optimal solution, where the CVaR variables and worst-case probability model are decided simultaneously. This is done by iterating between solving the inner- and the CVaR optimization problem. While one of these subproblems is solved, the decision variables of the other are fixed. The optimal decision variables are then input as parameters when solving the other problem.

Recall that Algorithm 5.1 finds a solution to the general distributionally robust problem (5.16) by solving the opposing outer and inner problems in sequence, where each problem uses the opponent's newest optimal solution as input. As we saw in the preceding sections this eventually results in an equilibrium strategy. Even though this does not guarantee an optimal solution, the iterative procedure gives an approximation of what the equilibrium would be if the outer and inner problem decisions were decided simultaneously.

Our approach for approximating problem (5.45) is based on the same idea. The purpose is to iterate between the inner- and the CVaR problem until there are no further improvements in the worst-case probability model. This is an approximation of the optimal worst-case probability distribution when CVaR variables are decided simultaneously. We emphasize that we in this thesis do not aim to prove the convergence of this procedure. However, in Section 7.2.3 we compare the model performance when this approach is included, against the instance when the CVaR variables are confined to only vary in the outer problem.

The iterative procedure does not apply to the outer problem. This is because Algorithm 5.1 already treats the probability model from the inner problem as fixed when the outer problem is solved. This means that the outer problem can be solved simultaneously as the CVaR optimization problem. In practice, the CVaR problem (5.44) is integrated in the outer problem with a fixed $\mathbb{P}(\mathbb{T}, P)$.

We now describe the details of the iterations between the inner problem and the CVaR problem. Recall from Algorithm 5.2 that we in the inner problem iterate stagewise down the scenario tree. For our proposed CVaR extension, we for a given stage t add an inner

loop that iterates over the nodes $l \in \mathcal{N}_t$. For every node we solve problem (5.44) and store the CVaR variables. Once the final node at that stage has been reached, all CVaR variables are input to the inner problem. The inner problem then finds a new worst-case probability distribution for that stage, and then uses the new stagewise probabilities as input for the next CVaR problem iterations. A proposed stopping criterion is to require the absolute deviation of the probabilities between two iterations to be below a predetermined threshold. The inner problem then repeats the sequence for the succeeding stage. The iteration sequence between the inner problem and the CVaR problem are illustrated in Figure 5.7.

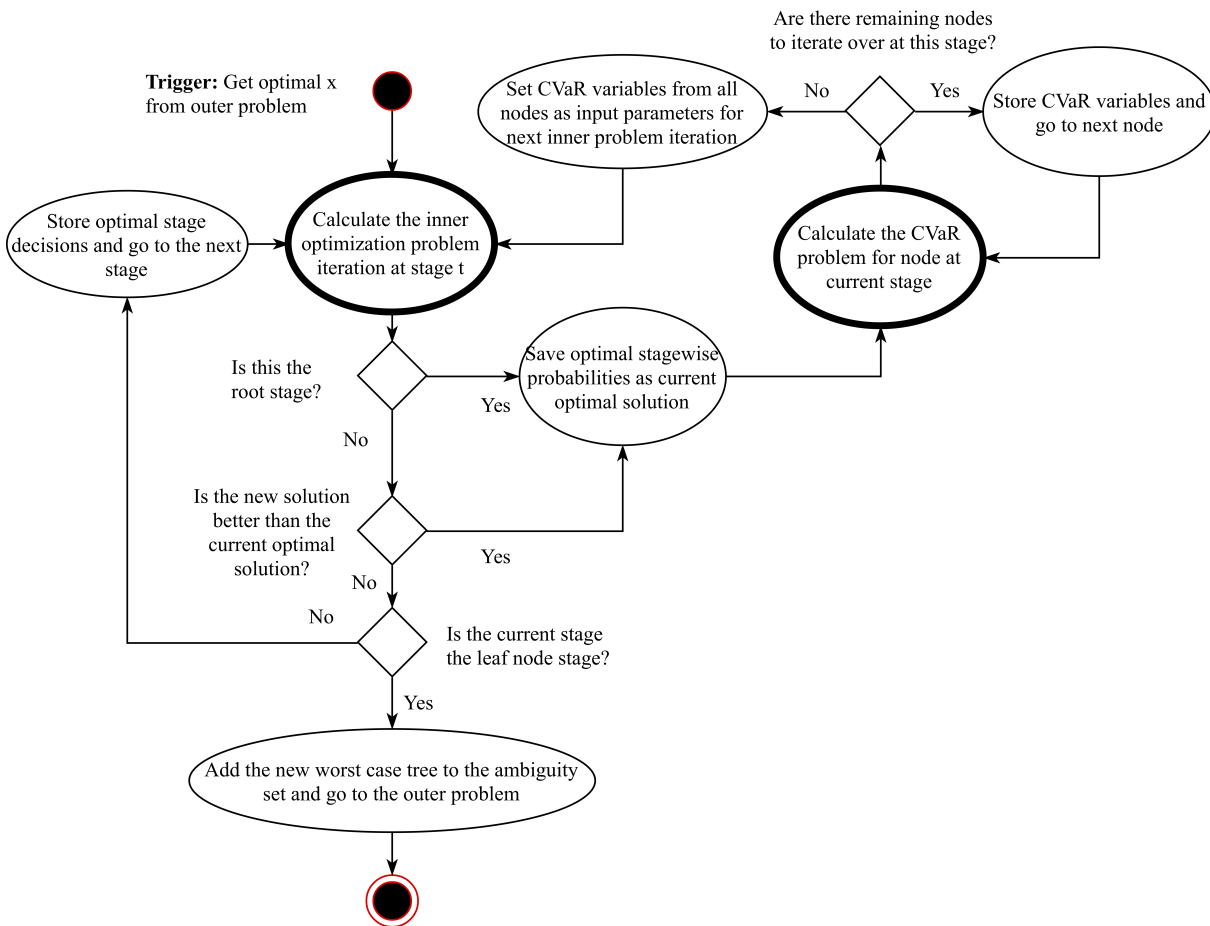


Figure 5.7: The inner problem solution algorithm for solving the distributionally robust optimization problem with CVaR.

With an understanding of how a distributionally robust optimization model with CVaR as risk measure can be formulated and solved, one remaining step remains before we can proceed towards adapting it to a hedging instance. Up to this point we have assumed to be in possession of a baseline probability model, or more specifically a scenario tree, which is estimated from empirically observed data. In the remainder of this chapter we present the approach for creating the scenario tree from a series of forecast paths.

5.4 Scenario tree generation

For multistage stochastic optimization problems, scenario trees are commonly used to represent the uncertainty in the underlying stochastic variable. A Scenario tree is a discrete representation of the uncertainty in a distribution, consisting of a finite number of stages and nodes per stage. In the tree, each node represents a potential realization of the stochastic variable at a certain stage, where new information is revealed (Kaut and Wallace, 2003). Further, a path of nodes through the tree represents a potential series of realizations of the stochastic variable, which is realized with a certain probability. This is known as a scenario.

As illustrated in Figure 5.8, a scenario tree consists of a branching structure where each node may have several successors. Hence, new information of the future distribution is obtained at each stage in the tree. An important property of a scenario tree is further that it accounts for the conditional probabilities between nodes at different stages (Shapiro, 2003).

When observing a set of simulations or historical time series, the information about a stochastic process is often observed as a *scenario fan*. A scenario fan is a set of paths evolving from the root node to the leaf nodes, where each intermediate node, except for the root node and the leaf nodes, has only one direct successor. Hence, as opposed to a scenario tree, full information is obtained after the first stage, with no uncertainty related to the succeeding stages (Pflug and Pichler, 2015). This property makes scenario fans unsuitable for multistage stochastic optimization. A methodology for transforming a scenario fan into a scenario tree is required in order to apply multistage stochastic optimization on this type of data.

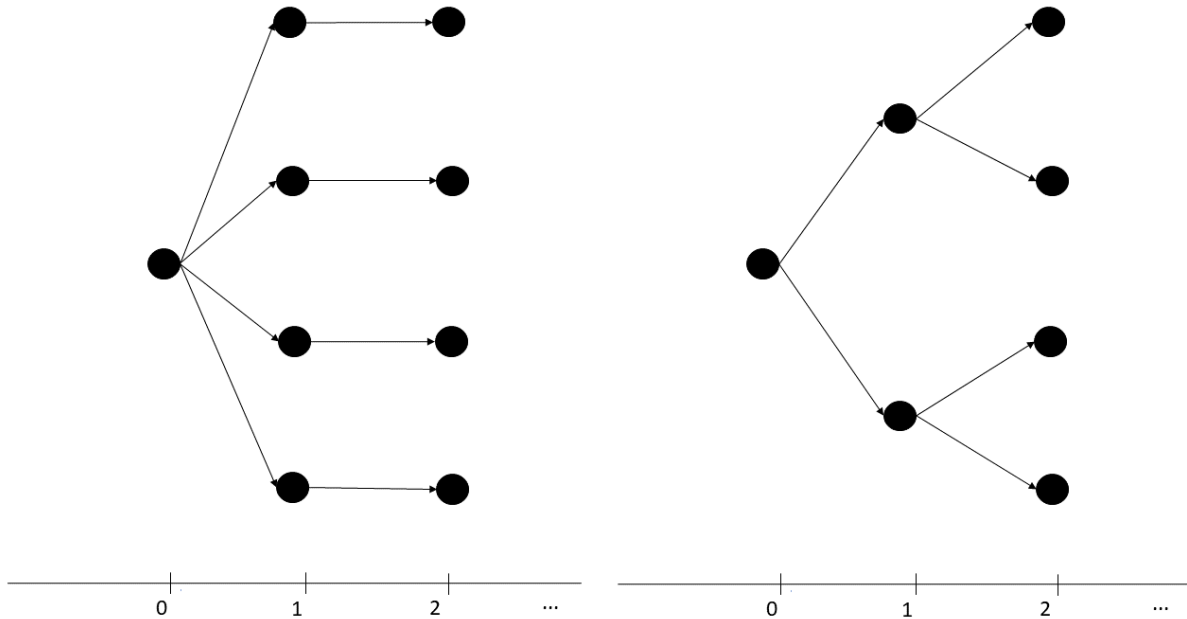


Figure 5.8: The scenario tree generation method transforms a scenario fan(left) to a scenario tree(right).

The literature on methods to generate scenario trees is rich, and an overview of the evolution of the most common approaches is found in Kaut and Wallace (2003). A more recent state of the art overview is also covered in Perez-Valdes et al. (2016). Briefly, one may divide the methods into two: Forecast-based methods and scenario clustering methods. The forecast-based approach typically involves fitting a forecast model to the input data before generating the scenario tree from its residuals. The scenario clustering methods generate scenario trees from scenario fans by clustering paths from the fan to achieve the desired tree structure.

A method developed by Nowak and Tomasgard (2007) is an example of a scenario tree generation method that applies a forecasting model in scenario generation. Based on time series of historical data, an autoregressive model is adapted to capture the underlying process of the stochastic variable, as for instance is explained Alexander (2008). Further, the 4 statistical moments of the error terms from the model are calculated, before scenarios that satisfy these statistical moments are generated by the moment matching method from Høyland et al. (2003). The scenario tree is then generated by combining the autoregressive

model with scenarios from the moment matching at each non-leaf node in the tree, starting from the root.

The method of Nowak and Tomasgard (2007) is well suited for transferring the serial correlation in the underlying process and the statistical moments of the underlying data to the scenario tree. Another advantage is that the sampled input data does not limit size of the scenario tree. On the other hand, this method does not prevent unrealistic outcomes of the random variable, such as price scenarios below zero, from being generated. This scenario tree generation method is for instance used in Schütz et al. (2009), where they also apply a principal component analysis (PCA) on the error term distribution to reduce the number of stochastic variables to the most important ones. This is done before conducting the moment matching.

There are also alternatives to the moment matching when generating scenarios based on the method of Nowak and Tomasgard (2007). Examples are to sample scenarios directly from the discrete distribution of the error terms, as done in Ladurantaye et al. (2009), and the copula-based method, as described in Kaut (2014). We do not cover these methods in further detail.

Instead of the forecast-based method, an alternative is to apply a clustering method when creating a scenario tree. In these methods a scenario tree is generated from a scenario fan, which is beneficial as stochastic information commonly is obtained on this form. Examples of such formats are historical time series and outputs from forecasting and simulation models. In general, clustering methods seek to generate a scenario tree with a predetermined structure that describes the statistical moments of the scenario fan as well as possible. A general drawback of clustering methods is however that they require assumptions to be made about the dependency between scenarios at different stages (Shapiro, 2018). A further drawback is that the number of scenarios at the last stage in the tree is limited to the total number of scenarios from the scenario fan it is made from. This limits the branching factor at each node. Generating scenario trees from scenario fans is therefore a tradeoff between the depth of the tree and the number of branches per stage.

An example of a clustering method is the approach of Heitsch and Römisch (2009). The method is for instance applied in Gabriel et al. (2009) and Fleten et al. (2002). Here, a fan is transformed to a tree by combining nearby nodes at a stage. The probabilities

of the clustered nodes are further aggregated, and the predecessor relations are retained. The number of nodes that are combined per stage depends on a predetermined number of scenarios per stage in the tree. A drawback with this clustering method is that only the node probabilities are adjusted during the clustering, while the new node value is simply assigned the value of the most likely old node. Hence, the expected value of the tree changes when this clustering method is conducted. This drawback is overcome by the clustering method applied in this thesis.

In the following sections give a detailed presentation of the clustering method applied for scenario generation in this thesis. This method uses method bases on stochastic approximation when generating the scenario tree from the scenario fan.

5.4.1 Scenario tree generation by stochastic approximation

The scenario tree generation method applied in this thesis is a clustering method based on the work of Pflug and Pichler (2015), Pflug and Pichler (2016) and Séguin et al. (2017). The method seeks to create a scenario tree from a scenario fan of forecasts, which from here are called the input paths, by minimizing the difference between the generated tree and the set of input paths. This is achieved by applying techniques from stochastic approximation, which is an iterative algorithm that is suitable for solving complex optimization problems (Pflug and Pichler, 2015).

Recall from Section 5.1.2 that the nested distance has been proven as a suitable metric for measuring the difference between two scenario trees. This metric is used to measure the extent to which the input paths and the generated tree differ, and hence to quantify the quality of the approximation. Nested distance for this reason also used as a convergence criterion. In addition to being beneficial relative to other scenario generation methods, this yields particularly interesting in the context of distributionally robust optimization. Recall from Section 5.2 that the nested distance defines the ambiguity radius of the ambiguity set in the distributionally robust optimization model. The decision maker can therefore use the nested distance convergence criterion to determine the lowest amount of ambiguity to account for, to address possible approximation errors from the scenario tree generation.

Another important advantage is that the new nodes values are randomly from the condi-

tional distribution of the input paths. According to Séguin et al. (2017) this ensures that all available data is applied in each iteration to improve the values of the tree, which to a large extent retains the statistical properties and serial correlation. This is a solid benefit relative to the method of Heitsch and Römisch (2009).

The scenario generation method considered in this thesis can be divided into two steps, initialization and improvements of the scenario tree. The first step is the initialization step, where the structure of the scenario tree is specified and a clustering algorithm is applied on the input data to initialize the node values in the tree. The second step aims to improve the values and assign probabilities to the nodes in the initiated tree. The improvement step is iterative. It first generates a random path, based on the distribution of the input paths and conditional on the preceding stages. Secondly, the randomly generated path is used to update the values and probabilities of the scenario tree with the application of stochastic approximation. The improvement step is repeated until the convergence criterion of a sufficiently small nested distance between the scenario tree and the input data is reached. The details of this method follow in the following subsections, and are summarized in Algorithm 5.3.

Step 1: Initialization of the scenario tree

First, the structure of the scenario tree is specified and an empty tree with this structure is created. Fixing the structure of the tree involves specifying the following properties: The number of stages, the number of nodes per stage, the predecessor relations between the nodes and the duration between the stages. Secondly, a clustering algorithm is applied on the input paths to cluster the paths at each stage according to the predetermined number of nodes per stage in the tree. The values obtained from the clustering are assigned as initial node values in the scenario tree. All scenario probabilities are initially set to zero.

K-means is applied as clustering algorithm, based on the work of Lloyd. (1982). The algorithm seeks to allocate data observations into a number of clusters, where the distance from the observations to the mean of the cluster they belong to, is minimized. This is done by denoting node $i \in \hat{\mathcal{N}}_t$ from the input paths at stage $t \in T$ as X_{it} , where X_{it} is a vector in \mathbb{R}^m and m is the dimension of each observation. Since $\hat{\mathcal{N}}_t$ is equal for all stages except of the first, we set $\hat{\mathcal{N}}_t = \hat{\mathcal{N}}$ for $t \in T \setminus \{1\}$ and let the node index i also denote the input path

index. The algorithm seeks to cluster the observations into $|\mathcal{N}_t| \leq |\hat{\mathcal{N}}|$ scenarios per stage, where \mathcal{N}_t is the set of nodes at stage t in the predetermined tree structure. The clustering is carried out so that the within-cluster sum of squares from the nodes to their cluster mean is minimized. Normally, an iterative algorithm is applied to achieve this. Initially, $|\mathcal{N}_t|$ clusters are arbitrarily chosen, for instance by randomly choosing observations from the input paths. Subsequently, each observation is assigned to the cluster to which its squared Euclidian distance is the least. The cluster means are then updated according to the centroid of the new cluster of observations. The algorithm iterates until there are no reassignments of observations to new clusters.

Step 2: Improvement of the scenario tree

Once the scenario tree is initialized, the values of the nodes are improved and their respective probabilities are calculated. The improvement step is iterative and consists of generating a random path from the distribution of the input paths, and applying this in a stochastic approximation to improve the values in the tree. The algorithm applies a nested distance criterion to determine convergence, and to quantify the quality of the approximation.

Generating a random path A random path is generated from the root stage to the leaf stage, where the value at each stage is based on the distribution of the input paths, conditional on the path's values at preceding stages. Accounting for the conditional probabilities ensures that the serial correlation in the data set is incorporated when generating the random path.

The process of generating a random path can be divided in two. First, a smoothed approximation of the conditional distribution of the input paths is generated with use of *kernel density estimation*. Secondly, random values are obtained based on the smoothed approximation of the conditional distribution until a path from the root stage to the leaf stage is generated. Specifically, we utilize the so-called *weight distribution* of the input paths, which is obtained from the expression of the smoothed conditional distribution, when generating the random values of the path. The detailed description of the method follows.

In order to generate multiple random paths based on the conditional distribution of input paths, we require a smoothed approximation of the conditional distribution. A smoothing convolution adapted from *kernel density estimation* is therefore employed to approximate the conditional density distribution of the input paths. From statistics, kernel density estimation is a well known non-parametric approach to approximate a smooth density function from a set of discrete observations. To achieve this, a smooth kernel function is applied, which allows interpolation between the observed samples to estimate the density at points where no samples have been observed (Chen, 2017).

We now proceed to the details on how the random path is generated. For the random variable $\xi_t \in \mathbb{R}^m$, the random path (ξ_0, \dots, ξ_T) is drawn from the conditional density distribution of the input paths (X_{i0}, \dots, X_{iT}) . In order to find the conditional density distribution over the input data, we first define the pure density distribution, which is unconditional on realizations of the random variable at previous stages. As discussed above, kernel density estimation is applied to make a smooth approximation of the distribution. The density distribution of the input paths is therefore referred to as the kernel density estimator. Based on the definition in Chen (2017), the unconditional kernel density estimator of the observed paths at stage t is then (5.46).

$$\hat{f}(\xi_t) = \frac{1}{|\hat{\mathcal{N}}| \cdot h_t} \sum_{i \in \hat{\mathcal{N}}} \kappa \left(\frac{\xi_t - X_{it}}{h_t} \right) \quad (5.46)$$

In equation (5.46), κ is the smooth kernel function which determines the shape of the distribution in the kernel density estimation (Chen, 2017). There exists a wide range of kernel functions that are commonly used. Jones (1990) proposes that the chosen function has no significant importance for the quality of the estimation. In this thesis the logistic kernel from (5.47) is applied. For further discussions on the choice of kernel functions, see Pflug and Pichler (2015).

$$\kappa(\xi) = \frac{1}{e^\xi + 2 + e^{-\xi}} \quad (5.47)$$

In equation (5.46), h_t is the bandwidth applied in the kernel estimation at stage t . The

bandwidth is a parameter that depends on standard deviation of the input paths at a given stage. In (5.46) it is used to control the degree of smoothing in the kernel density estimation. The choice of bandwidth has a stronger influence on the performance of the estimation than the kernel function, κ , and there exists a high variety of approaches to select the best bandwidth (Chen, 2017). We apply Silverman's rule of thumb, which is an approach that is particularly suitable when the underlying distribution is close to normal (Silverman, 1986). Equation (5.48) shows how h_t is found based on Silverman's approach, where $\sigma(X_{it})$ is the standard deviation of the data paths at stage t .

$$h_t = \sigma(X_{it}) \cdot |\hat{\mathcal{N}}|^{-\frac{1}{m+4}} \quad (5.48)$$

Samples from (5.46) can be applied to draw a random path that is unconditional on its previous stages. However, to make the random path conditional on its preceding stages, such that serial correlation is accounted for, (5.46) is combined with the definition of conditional density functions. From Pflug and Pichler (2015), a conditional density function is defined as follows: $f(x | y) = \frac{f(x,y)}{f(y)}$, where $f(x, y)$ is a multivariate density function. Further, by inserting ξ_t for x , and ξ_0, \dots, ξ_{t-1} for y , and then applying that the multivariate density function $f(x, y) = f(x) \cdot f(y)$, the kernel density estimator, conditional on the preceding stages, ξ_0, \dots, ξ_{t-1} , becomes (5.49).

$$\hat{f}(\xi_t | \xi_0, \dots, \xi_{t-1}) = \sum_{i \in \hat{\mathcal{N}}} \prod_{t'=0}^{t-1} \frac{\kappa\left(\frac{\xi_{t'} - X_{it'}}{h_{t'}}\right)}{\sum_{i' \in \hat{\mathcal{N}}} \kappa\left(\frac{\xi_{t'} - X_{i't'}}{h_{t'}}\right)} \cdot \kappa\left(\frac{\xi_t - X_{it}}{h_t}\right) \cdot \frac{1}{h_t} \quad (5.49)$$

By drawing samples from expression (5.49), a random path based on the distribution of the input paths and conditional on the values of previous stages, can be generated. According to Pflug and Pichler (2016), these samples from (5.49) can be found without calculating the analytical representation (5.49). In stead, one can utilize that $\prod_{t'=0}^{t-1} \frac{\kappa\left(\frac{\xi_{t'} - X_{it'}}{h_{t'}}\right)}{\sum_{i' \in \hat{\mathcal{N}}} \kappa\left(\frac{\xi_{t'} - X_{i't'}}{h_{t'}}\right)}$ in (5.49) is the weight $w_{it}(\xi_0, \dots, \xi_{t-1})$ of the input path i at stage t . The weight, w_{it} , indicates how close the randomly generated path has been to input path i at preceding

stages $0, \dots, t-1$ relative to the other input paths. The closer an input path i has been to the values of the randomly generated path at preceding stages, the higher is the weight of input path i at stage t . The higher the w_{it} , the higher is the probability that the randomly sampled path ξ_t is placed closer to this input data path X_{it} at stage t .

At each stage we therefore get a *weight distribution* of the different input path $w_{0t}, \dots, w_{|\mathcal{N}|t}$, which further is used to obtain a random sample from (5.49) at stage t . To obtain the weight of input path i at stage t , its weights from all preceding stages $0, \dots, t-1$ are multiplied. All the weights are positive and sum to 1 at each stage. At the first stage, the weights for all input paths are equal, meaning the weight distribution is uniform where each path's weight is $1/n$. By employing this, (5.49) becomes (5.46).

When using the weight distribution, a random sample from (5.49) is obtained in the following way: First, the input path with index i^* is randomly drawn from the weight distribution. Input paths with high weights, which are those that historically have been close to the randomly generated path, have higher probabilities of being drawn as i^* . The probability that the random path is close to the input path with index i^* at subsequent stages, is then increased. In that way, the randomly generated path depends on preceding stages, and conditional probabilities are therefore accounted for. In Séguin et al. (2017), i^* is identified as the input path index satisfying equation (5.50), where $rand_u$ is randomly drawn from the uniform distribution in the interval $[0,1]$.

$$\sum_{i=1}^{i^*-1} w_{it}(\xi_0, \dots, \xi_{t-1}) \leq rand_u \leq \sum_{i=1}^{i^*} w_{it}(\xi_0, \dots, \xi_{t-1}) \quad (5.50)$$

In (5.50), i^* at stage t is identified as the index of the first input path which makes the accumulated weights of input paths, $w_{1t} + \dots + w_{i^*t}$, exceed $rand_u$.

According to Pflug and Pichler (2015), the value of the randomly generated path ξ_t at stage t is found by adjusting the value of the identified input path i^* at stage t , with a random deviation drawn from the kernel density function. The size of the random deviation depends on the bandwidth of the input paths at stage t . (5.51) below gives the mathematical expression used to obtain the value of the random path ξ_t at stage t . In the expression, X_{i^*t} is the observed value of the input path with index i^* at stage t , Z is

a random deviation from the kernel density function and h_t is the bandwidth. Sampling the values of Z is done according to the composition method (for further details see Pflug and Pichler (2014a)).

$$\xi_t = X_{i^*_t} + (Z \cdot h_t) \quad (5.51)$$

Ultimately, to obtain a random path from root to leaf, ξ_0, \dots, ξ_T , the approach presented above is repeated for all stages from $t = 0$ to $t = T$.

Stochastic approximation After generating the random path ξ_0, \dots, ξ_T according to the input paths, it is used to improve the values and probabilities in the scenario tree through stochastic approximation. First, the closest path in the scenario tree to the randomly generated path is identified. Secondly, the values and probabilities of the nodes in the identified path are updated.

The path in the scenario tree that is closest to the randomly generated path is identified as the one minimizing some distance measure, for instance the r -order distance (5.13), to the random path. This path that is the closest to the random path is according to Pflug and Pichler (2015) found by solving (5.52). To understand (5.52) we denote a path of nodes in the scenario tree, from the root to a leaf, $\omega_{j_T} = j_0 \dots, j_T$, where $j_t \in \mathcal{N}_t$ is a node at stage t with predecessors j_0, \dots, j_{t-1} . For each stage t , $\mathcal{N}_t(j_0, \dots, j_{t-1})$ is the set of all the nodes with predecessors j_0, \dots, j_{t-1} . Further, the values of a scenario tree node j_t is Y_{j_t} on \mathbb{R}^m , corresponding to X_{i_t} from the input paths. Then, the path of nodes, $j_0 \dots, j_T$, with the minimum distance measure to the random path, ξ_0, \dots, ξ_T , is identified as follows:

$$j_t \in \underset{j'_t \in \mathcal{N}_t(j_0, \dots, j_{t-1})}{\operatorname{argmin}} d_r(Y_{j'_t}, \xi_t) \quad (5.52)$$

In practice, the path is identified by sequentially navigating through the stages in the tree, from the root node to the leaf nodes. For each stage, the node minimizing the distance measure to the random path at the respective stage is identified. Then, for the subsequent

stage, the children nodes of the identified node are considered in the similar way. This continues until a leaf node as been reached. The identified path, j_0, \dots, j_T , is then the path in the scenario tree that minimizes the distance measure to the randomly generated path, ξ_0, \dots, ξ_T .

The values of the nodes in the identified path are then improved to better approximate the randomly generated path, and hence the forecast fan. The improvement is done according to the stochastic gradient descent method, a frequently used method to update parameters of a function when an optimum is sought through stochastic approximation. According to Pflug and Pichler (2015), the values of the identified nodes j_0, \dots, j_T can then be updated according to (5.53) after each iteration $k = 1, 2, 3, \dots$. The states of the nodes at $k = 0$ are the initial values from the K-means clustering, which we explained in section 5.4.1. To simplify the calculations, the r -order distance $d_r(\cdot, \cdot)^r$ is applied as the distance measure between the nodes in the random path and the identified path. We denote $\alpha_{(p)}$ as the step-size, which can be interpreted as the learning factor of the approximation. In Pflug and Pichler (2016) it is proven that the approximation converges if $\alpha_{(p)} > 0$, $\sum_p \alpha_{(p)} = \infty$ and $\sum_p (\alpha_{(p)})^2 < \infty$. Séguin et al. (2017) set $\alpha_{(p)} = \frac{1}{p+30^{\frac{3}{4}}}$ in the following equation (5.53).

$$Y_{j_t}^{(k)} = Y_{j_t}^{(k-1)} - \alpha^{(k)} \cdot r d_r(\xi_t^{(k)}, Y_{j_t}^{(k-1)})^{r-1} \cdot \nabla d_r(\xi_t^{(k)}, Y_{j_t}^{(k-1)}) \quad (5.53)$$

Lastly, the probabilities of the nodes in the scenario tree are updated. Initially, all nodes are assigned a counter which is set to zero. For each iteration when a path that is chosen as the closest path to the randomly generated path, the counter for all nodes in this path is increased by one. When the approximation has converged, the unconditional probability of any node is obtained by dividing its counter by the total number of iterations. The conditional probability of a node is obtained by dividing its unconditional probability by the unconditional probability of its preceding node. Finally, the scenario probabilities in the tree is the product of the conditional probabilities of the nodes the scenario consists of.

Convergence The stochastic approximation of the scenario tree terminates when the approximation of the nested distance $\mathbb{E}(d_r)$ between the generated tree and the input paths converges. For instance, Séguin et al. (2017) require that the change in the approximated nested distance $\Delta\mathbb{E}(d_r)$ is within a certain ϵ for the ten latest iterations in order to determine convergence. The algorithm iterates between the random generation of a path and stochastic approximation, where the tree is improved for each iteration until this criterion is met.

Based on the works of Pflug and Pichler (2015) and Séguin et al. (2017), we explain how the nested distance between the tree and the input paths is calculated after each iteration in the algorithm. For clarity we denote the closest to the randomly generated path ξ_0, \dots, ξ_T as j_0^*, \dots, j_T^* hereafter. Recall that we write $\Omega_T = \omega_{0T}, \dots, \omega_{\mathcal{N}_T T}$ as the set of $|\mathcal{N}_T|$ paths in the scenario tree. Then, $\omega_{j_T^*}^{(k)}$ consists of the nodes $(j_1^*, \dots, j_T^*)^{(k)}$ and is the path that at iteration k is the one closest to the randomly generated path $\xi^{(k)} = \xi_1^{(k)}, \dots, \xi_T^{(k)}$. Following this, the distance measure at iteration k between the randomly generated path $\xi^{(k)}$ and the values of path $\omega_{j_T^*}^{(k)}$, which we denote $Y_{\omega_{j_T^*}^{(k)}} = (Y_{j_1^*}, \dots, Y_{j_T^*})^{(k)}$, is found from (5.54).

$$d_{\omega_{j_T^*}^{(k)}}^{(k)} = d_r(\xi^{(k)}, Y_{\omega_{j_T^*}^{(k)}})^r \quad (5.54)$$

The calculated distance $d_{\omega_{j_T^*}^{(k)}}^{(k)}$ is further added to the accumulated distance from previous iterations, for this specific path. We denote the accumulated distance after iteration k , for the path with index j_T^* at iteration k , as $D_{\omega_{j_T^*}^{(k)}}$. Be aware that $D_{\omega_{j_T^*}^{(k-1)}}$ refers to the accumulated distance after iteration $k-1$, but still for the path with index j_T^* at iteration k . Then, $D_{\omega_{j_T^*}^{(k)}}$ is found by (5.55).

$$D_{\omega_{j_T^*}^{(k)}} = d_{\omega_{j_T^*}^{(k)}}^{(k)} + D_{\omega_{j_T^*}^{(k-1)}} \quad (5.55)$$

For all paths except for $\omega_{j_T^*}$, the accumulated distance is unchanged from the previous iteration. That is, for all paths $\omega_{j_T} \in \Omega_T \setminus \{\omega_{j_T^*}\}$, $D_{\omega_{j_T}^{(k)}} = D_{\omega_{j_T}^{(k-1)}}$.

Finally, the nested distance between the generated scenario tree and the input paths is

approximated according to (5.56), where $P(\omega_{j_T})$ is the probability of path ω_{j_T} . It is written as

$$\mathbb{E}(\mathbf{dl}_r) = \sum_{j=1}^{N_T} \left(\frac{D_{\omega_{j_T}}^{(k)}}{k} \cdot P(\omega_{j_T})^{(k)} \right) \quad (5.56)$$

The approximation of the nested distance from (5.56) can be interpreted as the expected distance between a randomly drawn path from the distribution of the input paths to its closest path in the generated scenario tree. Hence, this is an indication of how well the generated tree approximates the scenario fan.

(5.56) is an approximation of the definition of nested distance, which is formulated exactly as problem (5.8) to (5.12) in Section 5.1.2. Since the approximation contains a distance measure and probabilities, which is the same components as the objective of the nested distance definition, there is a clear resemblance. A proof of why (5.56) can be applied as an approximation of the nested distance is found in Pflug and Pichler (2015).

In conclusion, we have generated a scenario tree from a set of forecast paths, based on an approach that applies stochastic approximation. The scenario tree is generated so that it minimizes the nested distance to the forecast paths. Algorithm 5.3 summarizes the approach in a structured way.

In Chapter 6 we formulate the specific distributionally robust hedging model for TrønderEnergi, based on the theoretical foundation established in this chapter.

Algorithm 5.3 Scenario tree generation by stochastic approximation

Input: The method takes in a scenario fan, i.e a set of input paths.**1) Initialization:**

Set $k = 1$ and the node counter $c(j_t)^{(k-1)} = 0$ for each node $j_t \in \mathcal{N}_t$ and each $t \in T$. Let the accumulated distance for each path in the tree, $\omega_{j_T} = j_0, \dots, j_T$, be $D_{\omega_{j_T}}^{(k-1)} = 0$.

1.1) Tree-structure: Decide the number of stages and the number of nodes per stage. For all nodes, set their values, Y_{j_t} , and probabilities, $Pr(j_t)$, to 0.

1.2) Clustering: Apply a clustering algorithm on the input paths to allocate initial values to the nodes in the tree.

2) Iteration: Improve the states and probabilities of the tree.

2.1) Random path: Generate a random path, ξ_0, \dots, ξ_T that is based on the distribution of the input paths and conditional on its preceding stages.

2.2) Stochastic approximation:

2.2.1) Identify the path of nodes in the tree, ω_{j_T} that minimizes the distance measure, d_r , to the random path.

2.2.2) Update the values of the identified path of nodes according to the gradient descent method:

$$Y_{j_t}^{(k)} = Y_{j_t}^{(k-1)} - \alpha^{(k)} \cdot r d_r(\xi_t^{(k)}, Y_{j_t}^{(k-1)})^{r-1} \cdot \nabla d_r(\xi_t^{(k)}, Y_{j_t}^{(k-1)})$$

2.2.3) For each node in the identified path, set: $c(j_t)^{(k)} = c(j_t)^{(k-1)} + 1$

For the identified path, set: $D_{\omega_{j_T}}^{(k)} = d_r(\xi^{(k)}, Y_{\omega_{j_T}}^{(k)})^r + D_{\omega_{j_T}}^{(k-1)}$.

2.2.4) The nested distance between the fan and the tree is approximated, $\mathbb{E}(dl_r^{(k)}) = \sum_{j=1}^{\mathcal{N}_T} \left(\frac{D_{\omega_{j_T}}^{(k)}}{k} \cdot P(\omega_{j_T})^{(k)} \right)$, where $P(\omega_{j_T}) = \frac{C(j_T)}{k}$ is the unconditional probability of a path ω_{j_T} .

3) Convergence: If $\Delta \mathbb{E}dl_r^{(k)} = (\mathbb{E}dl_r^{(k)} - \mathbb{E}dl_r^{(k-1)}) < \epsilon$ for the q last iterations: Terminate. Else: set $k = k + 1$ and go to "**2.1) Random path**".

Output: The method gives a scenario tree that minimizes the nested distance to the scenario fan.

Chapter 6

Electricity price hedging model

The purpose of this chapter is to develop the stochastic and distributionally robust hedging model, according to the problem description presented in Chapter 4, and based on the solution method presented in Chapter 5. Recall that the stochastic model approaches uncertainty in the input parameters by assigning probabilities to the potential scenarios, which are assumed to be known. The multistage distributionally robust model on the other hand considers model uncertainty, by also letting the scenario probabilities themselves be subject to uncertainty.

We first present any further assumptions and specifications to the hedging problem. Secondly, the multistage stochastic hedging model is developed. Lastly, this model is extended to the distributionally robust hedging model.

6.1 Assumptions

We assume that TrønderEnergi only considers hedging, and not speculation, as the underlying motivation for entering into derivatives. Recall that this is in line with how Stulz (1996) and Tanlapco and Liu (2002) define hedging, and how Fleten et al. (2002) approach a similar decision problem with stochastic optimization. To make sure that decisions are made with the purpose of reducing risk, and not to increase profits, the prices of the forward contracts are derived to be consistent with the conditional expected spot prices in the scenario tree. This is further elaborated in Section 7.1.1. In reality, entering

into forwards often combines hedging and speculative considerations, as we recall from Sanda et al. (2013) who investigate the Norwegian electricity market. The advantage of distinguishing between these decisions in an optimization model is however that the hedging decisions can be isolated. The alternative is to incorporate speculative decisions by including the decision maker's market view in the model. .

As hedging and not speculation is considered, the decision maker can only short, i.e sell, forward contracts written on the electricity price. Hence, speculating in terms of entering into long forward positions is not allowed. This assumption is common in other papers approaching optimal hedging decisions, such as in Shütz and Westgaard (2018).

For the hedging model, a planning horizon on 4 months with monthly stages is considered. At the first stage in the model, the decision maker decides how much of the expected production for succeeding stages that should be sold to an agreed forward price. At each succeeding stage, rebalancing decisions can be made based on new information. The remaining production at a stage that not is delivered through forward contracts, is sold to an uncertain spot price. In reality, hedging and rebalancing decisions can be made daily. This however requires an undesirably large scenario tree, which effects the computational performance of the model.

For simplicity, only forward contracts with monthly maturity periods are considered. For the regarded planning horizon, these contracts are considered more relevant as contracts with longer maturity periods normally are entered earlier relative to their maturity dates (Sanda et al., 2013). Further, we assume the contracts to be infinitely liquid so that there is no lack of potential buyers and sellers causing problems. As we consider forwards, we also assume that the required contract size is available. In addition, we assume there is no bid-ask spread on the contracts.

In reality, forward contracts written on the Nordic system price are settled in cash. For speculators who are not interested in ownership of the physical commodity, cash settlement is therefore suitable. As speculation is not considered in this model, it is however more convenient in a hedging setting to regard physical settlement of contracts. This has no direct implications for the outcome of the model, but makes it easier to relate to hedging as a physical amount of electricity that actually is being sold on contract. The same assumption is for example made in Fleten et al. (2002).

We assume that any lack of production, following situations where the quantity of elec-

tricity sold through contracts exceeds the actual production for the same period, has to be immediately covered from the spot market by the decision maker. This equals a short position in the spot market and we hereafter refer to shortage production in these situations. Shortage production is however assumed to be covered with a marginal additional cost to the spotprice, representing an additional effort of this process.

The transaction costs, corporate tax and natural resource tax are assumed to be the only decision relevant costs. Transactions costs are set to be proportional to the quantity of contracts shorted, according to the standards formulated in Zenios and Markowitz (2008), and are payed at the time the contract is entered. For further details on the tax system and why this is relevant for the hedging decision, see Chapter 2.

Ultimately, no other sources to risk than price risk and production risk are assumed relevant for the decision problem.

In the next section we formulate the hedging specific stochastic optimization model.

6.2 The multistage stochastic hedging model

In both the stochastic and distributionally robust hedging model, the purpose is to find the optimal hedging strategy, defined as the one maximizing the decision maker's utility. In achieving this, the model determines the optimal quantity of forward contracts with different maturities to be entered, depending on the decision maker's risk aversion. An important consideration is also when to enter the respective contracts, following the trade-off between entering into a position for an agreed price today and waiting for new information. In both the stochastic and distributionally robust hedging models, the electricity prices, production quantities and the evolution of forward prices with different maturity dates are considered as the uncertain parameters.

When formulating the multistage stochastic hedging model, we first explain the components of the objective with related constraints and secondly present the complete model.

6.2.1 Formulating the objective function

As indicated above, the objective is a utility function which represents the trade-off between expected profits and risk. The purpose is to find the allocation of forward contracts that maximizes the decision maker's utility. Since we consider multistage optimization models, a starting point is to ensure that the risk measured by CVaR is appropriately taken into account for each stage in the model. This is achieved by formulating the utility function recursively, which ensures a time consistent CVaR as presented in as presented in Section 3.2. Our implementation follows the recursive formulation of Rudloff et al. (2014). The authors emphasize that a major problem with utilizing the recursive formulation is the lack of a clear economic interpretation, as opposed to for formulations with time inconsistent risk measures. They however suggest that the recursive formulation should be interpreted as "the certainty equivalent of the portfolio value w.r.t the time consistent dynamic utility generated by one period preference functional". This can be interpreted as the deterministic profits that makes the decision maker indifferent to an uncertain alternative, whose expected utility is attained by following the optimal strategy.

The recursive formulation of the utility function consists of two components, the ensured profits in a current state and the expected utility over the directly succeeding states. The objective is then recursively formulated for each state in the scenario tree, from the root to the leave nodes. Since every direct successor also incorporates the expected utility from their direct successors, we get a recursive structure that ultimately nests up the expected utility from all succeeding nodes at the root node.

The first component of the objective function is the ensured profits at a state, which includes sales in the spot market and forward contracts entered into at the current state. These profits are deterministic, and hence ensured as they are independent of any scenario realizations at succeeding stages. The second component is the expected utility over the directly succeeding states. This component can be divided into the expected profits over directly succeeding states and the related risk. These are further weighted with a parameter $\lambda \in [0, 1]$ which represents the decision makers risk aversion. A higher λ represents the preferences of a more risk averse investor, who prefers lower risk over potential profits. $\lambda = 0$ models the preferences of a risk neutral investor, while $\lambda = 1$ models the preferences of a fully risk averse investor.

For simplicity we hereafter refer to the expected utility over the directly succeeding states

as *expected future utility*, and the expected profits over the directly succeeding states as the *expected future profits*.

The utility at a state in the objective function, constructed from ensured profits and the expected future utility, can therefore be written as follows:

$$\text{utility} = \text{ensured profits} + \text{expected future utility}$$

Where the two components are further denoted as follows

$$\begin{aligned} \text{ensured profits} &= \text{ensured revenues} - \text{ensured costs} \\ \text{expected future utility} &= (1 - \lambda) \cdot \text{expected future profits} + \lambda \cdot \text{risk} \end{aligned}$$

In order to provide a proper understanding of the objective function, the following paragraphs thoroughly explain the components in the objective function. The components composing ensured profits are first presented in detail, secondly the same is done for the expected future utility. We lastly present the mathematical formulation of the complete hedging model is presented.

Ensured profits

The profits that are *ensured* in a state include the revenues and decision relevant costs, whose uncertainty is eliminated. These profits are therefore independent of the succeeding scenarios. The revenues are determined by spot and forward sales, while taxes, transaction costs and shortage costs compose the costs.

The ensured revenues and costs occur as a consequence of receiving new information at a stage, and can further be distinguished into two groups - those that are *ensured* and *realized* at the stage, and those that are *ensured*, but *realized* at later stages. The revenues that are ensured and immediately realized include sales in the spot market. On the other hand, entered forward contracts represent revenues that are ensured but realized at later stages, when the contracts mature.

Costs that are ensured and immediately realized include corporate taxes on spot sales, in addition to the natural resource tax. The natural resource tax is included since it is fully

based on realized spot prices and production. The costs of buying electricity in the spot market in cases of shortage production are also realized immediately. While the corporate taxes on entered forward contracts are realized when the contracts mature, the transaction costs for entering into the contracts are both ensured and realized when the contracts are entered. As the latter gives incentives to wait before entering into contracts with a certain maturity date, the transaction costs are discounted according to the time when they are entered. All other cash flows than transaction costs are realized at a predetermined stage, independent on whether sales are through spot or forwards entered at other stages. Hence, discounting these revenues and costs is not decision relevant. In addition, since only a decision horizon on 4 months is considered, discounting is practically negligible.

In formulating the mathematical expression of the ensured profits, we denote each scenario as $i \in \mathcal{N}_T$, and write the monthly decision stages as $t \in T$. We then represent the ensured profit function in a given state, that is for a given scenario i at stage t , by h_{it} . Revenues and costs are denoted as r_{it} and costs c_{it} , respectively.

$$h_{it} = r_{it} - c_{it} \quad (6.1)$$

We proceed by presenting the detailed formulations of the ensured revenues and the ensured costs.

Ensured revenues The ensured revenues for scenario i at stage t are written r_{it} , and are a function of forwards entered and spot sales realized at stage t . For the set of remaining time periods M_t , we then define the maturity dates of forwards as m stages after the time t they are entered. The quantity bought of contracts with maturity in m months at stage t in scenario i is then denoted x_{imt} . Note that $m = 0$ refers to spot sales, making x_{i0t} the quantity sold in the spot market in scenario i at time t . Further, the corresponding prices for forwards entered at the state is F_{imt} , where F_{i0t} is the spot price for scenario i at time t . We then express the ensured revenues for scenario i at stage t , r_{it} , as (6.2).

$$r_{it} = \sum_{m \in M_t} F_{imt} x_{imt} \quad (6.2)$$

Ensured costs The second term in (6.1), c_{it} , represents the decision relevant costs that correspond to the ensured profits, and can further be divided into three components: Taxes c_{it}^{TAX} , transaction costs c_{it}^{TR} and the shortage costs, c_{it}^S , as shown in (6.3).

$$c_{it} = c_{it}^{TAX} + c_{it}^{TR} + c_{it}^S \quad (6.3)$$

The taxes, c_{it}^{TR} , can further be divided into natural resource taxes and corporate taxes. We denote T^R as the tax rate for the natural resource tax, and T^F as the tax rate for the corporate tax. Recall that the natural resource taxes for scenario i at stage t are based on the realized production Q_{it} and the spot price F_{i0t} . On the other hand, the revenues r_{it} form the basis for the corporate taxes.

The transaction costs on forwards for scenario i at stage t are written c_{it}^{TR} , and are assumed to be proportional with a factor C^{TR} to the amount of shorted contracts x_{imt} , where $m > 0$. In line with the discussion in the initial paragraphs of this section, these costs are discounted with a factor ρ according to the time t they are entered at.

The shortage costs c_{it}^S are obtained from the shortage quantity z_{it} , which represents the quantum delivered on contract that exceeds the production in scenario i at stage t . In cases where the produced quantity equals or exceeds the quantum delivered on contract, the shortage quantity z_{it} is zero and the excess production is sold in the spot market. Shortage production is covered in the spot market with a marginal additional cost, ϵ , representing the additional effort of this process.

Equation (6.3) can now be expressed in detail as

$$c_{it} = (T^F r_{it} + T^R F_{i0t} Q_{it}) + \left(\sum_{\{m \in M_t \mid m > 0\}} \frac{1}{(1 + \rho)^{t/12}} C^{TR} x_{imt} \right) + ((F_{i0t} + \epsilon) z_{it}) \quad (6.4)$$

Remarks concerning the initial stage Because realized spot sales at the first stage are not considered decision relevant, these are assumed to be zero. Consequently, neither natural resource taxes nor shortage costs are present at the initial stage. The same accounts for corporate taxes on spot sales.

Expected future utility

The second component of the recursive objective function is the expected future utility, which we recall is the expected utility over the directly succeeding states. This component is risk exposed and consists of two terms: the expected future profits and the related risk.

Expected future profits For node k at stage t the expected profits at stage $t+1$ is the sum of the profit functions for all nodes at stage $t+1$ that are direct successors of node k , denoted $i' \in k+$. The direct successors are weighted with their conditional probabilities $P_t(i' | k+)$. The expected future profits, conditional on the direct predecessor k , can then be expressed as in (6.5), where $v_{i'(t+1)}$ is the objective value for node i' at stage $t+1$.

$$\sum_{i' \in k+} P_t(i' | k) v_{i'(t+1)} \quad (6.5)$$

Observe that we use the objective function value when calculating what we refer to as the expected future profits. This follows as we apply the recursive formulation of Rudloff et al. (2014) that ensures a time consistent risk measure. The way we interpret and formulate (6.5) is also in line with Shütz and Westgaard (2018) and Pisciella et al. (2016). They apply a similar recursive formulation based on Rudloff et al. (2014), when approaching hedging of salmon prices and a capacity expansion problem of a power producer, respectively. By setting $\lambda = 0$ it becomes more intuitive that (6.5) represents the expected future profits.

Risk The risk related to potential profits in nodes that directly succeed node k , represents a disutility for a risk averse decision maker. In the following paragraphs we explain the time consistent CVaR formulation applied to quantify the risk.

Recall from Chapter 3.2 that CVaR is defined as the expected loss, given that the loss exceeds the value at risk (VaR). Further, VaR is defined as the threshold to which losses shall not exceed with a given confidence level. This general definition applies when considering a cost or a loss function. In our instance, where we consider positive profits, we

define CVaR as follows. CVaR is the expected profits, given that the profits are lower than the VaR. The VaR which is modelled by τ , is in this instance defined as the threshold that profits ought to be greater or equal to, with a confidence level α . We hereafter denote the positive difference between the VaR threshold and profits as shortfall profits, where the shortfall profits are zero if profits are greater than or equal to the VaR. When considering profits instead of costs, formulation (5.44) from Section 5.3.3 becomes (6.6).

$$\text{CVaR}_\alpha(H(x, \xi), \mathbb{P}(\mathbb{T}, P)) = \max_{\tau} \tau - \frac{1}{1 - \alpha} \mathbb{E}_{\mathbb{P}(\mathbb{T}, P)} [\tau - H(x, \xi)]_+ \quad (6.6)$$

Recall that the time consistent recursive CVaR formulation from Rudloff et al. (2014) is achieved by recursively formulating the objective function. This further ensures that the CVaR only is measured over the direct successors of a node at a given stage, and is often referred to as the *nested CVaR*. For node k at stage t , the nested CVaR over the direct successors of node k are expressed by (6.7). Here, $P_t(i' | k)$ is the conditional probability of reaching node i' at stage $t + 1$ from node k , while $y_{i'(t+1)}$ is an auxiliary variable representing the shortfall profit of node i' . The shortfall profit is expressed by constraints (6.8) and (6.9) which ensure that $y_{i'(t+1)}$ is the maximum of $\tau_{kt} - v_{i'(t+1)}$ and 0, where τ_{kt} is the VaR threshold at a given significance α for node k at stage t . Be however aware that τ_{kt} is evaluated over the uncertainty exposed direct successor nodes of k . Constraints (6.8) and (6.9) are necessary constraints in formulating CVaR linearly, see for instance Rockafellar and Uryasev (2000). Equations (6.7) to (6.9) are formulated for a specific $k \in \mathcal{N}_t$ at a given $t \in T$

$$\sum_{i' \in k^+} P_t(i' | k) y_{i'(t+1)} \quad (6.7)$$

$$\tau_{kt} - v_{i'(t+1)} \leq y_{i'(t+1)} \quad (6.8)$$

$$0 \leq y_{i'(t+1)} \quad (6.9)$$

Observe that the shortfall profits in (6.8) are calculated relative to the objective function

value. The discussion of this is in line with the discussion conducted above related to (6.5).

Based on the presented components of the objective function, the hedging specific optimization model with the time consistent CVaR is now formulated in its whole.

6.2.2 Formulating the optimization model with a time consistent risk measure

For the reader's reference, an overview of the nomenclature applied in the stochastic hedging model is presented.

Nomenclature

Sets

$k+$	Set of direct successors i' of k .	$k \in \mathcal{N}_t, t \in T \setminus \{ T \}$
M_t	Set of maturities at stage t .	$t \in T$
\mathcal{N}_t	Set of nodes at stage t .	$t \in T$
$\mathcal{N}_T \mid i \succ k$	Leaf node successors of k .	$k \in \mathcal{N}_t, t \in T \setminus \{ T \}$
T	Set of stages.	
T_t	Set of stages from 0 up to t .	$t \in T$

Indices

i	Scenario and leaf node.	$i \in \mathcal{N}_T$
k	Node at stage t .	$k \in \mathcal{N}_t, t \in T \setminus \{ T \}$
i'	Direct successor i' of k .	$i \in \mathcal{N}_{(t+1)}, t \in T \setminus \{ T \}$
m	Stages to maturity m , in months.	$m \in M_t, t \in T$
t	Decision stage.	$t \in T$

Parameters

$P_t(i' k)$	Probability of reaching node i' conditional on its direct predecessor.	$i' \in k+, k \in \mathcal{N}_t, t \in T \setminus \{ T \}$
X_t	Quantity from contracts bought in previous planning periods, to be delivered at stage t .	$t \in T$
α	CVaR confidence level at $\alpha * 100\%$.	$\alpha \in [0, 1]$
ϵ	Additional cost of buying electricity in case of shortage production.	
λ	Risk aversion parameter.	$\lambda \in [0, 1]$

Variables

h_{it}	Realized profits in scenario i , at stage t .	$i \in \mathcal{N}_T, t \in T$
h_{kt}^N	Auxiliary node indexed variable.	$k \in \mathcal{N}_t, t \in T \setminus \{ T \}$
v_{kt}	Objective function value in node k at stage t .	$k \in \mathcal{N}_t, t \in T$
x_{imt}	Quantity sold on contracts bought at stage t , to be sold after m stages, in scenario i . $m = 0$ are spot sales.	$i \in \mathcal{N}_T, m \in M_t, t \in T$
$y_{i'(t+1)}$	Shortfall profit in direct successor node i' of k at stage $t + 1$.	$k \in \mathcal{N}_t, t \in T \setminus \{0\}$
z_{it}	Quantity to buy in case of shortage production, in scenario i at stage t .	$i \in \mathcal{N}_T, t \in T$
τ_{kt}	Value at Risk for node k at stage t .	$k \in \mathcal{N}_t, t \in T \setminus \{ T \}$

The stochastic hedging model

In the stochastic model presented below, we have for ease of notation only shown the profits h_{it} , and not the decomposition, which we explained above. Nevertheless, a recapitulation of the profit term is presented as follows

h_{it}	Profits	$i \in \mathcal{N}_T, t \in T$
r_{it}	Revenue	$i \in \mathcal{N}_T, t \in T$
c_{it}	Costs	$i \in \mathcal{N}_T, t \in T$
c_{it}^{TAX}	Tax costs	$i \in \mathcal{N}_T, t \in T$
c_{it}^{TR}	Transaction costs	$i \in \mathcal{N}_T, t \in T$
c_{it}^S	Shortage costs	$i \in \mathcal{N}_T, t \in T$
F_{imt}	Price of a forward contract bought at stage t , with maturity m , in scenario i , $m = 0$ denotes spot sales.	$i \in \mathcal{N}_T, m \in M_t, t \in T$
Q_{it}	Produced quantum in scenario i , at stage t .	$i \in \mathcal{N}_T, t \in T$
T^F	Corporate tax	
T^R	Natural resource tax	
C^{TR}	Transaction cost per unit	
ρ	Discount rate	

$$\begin{aligned}
h_{it} &= r_{it} - c_{it} \\
&= \left(\sum_{m \in M_t} F_{imt} x_{imt} \right) - \left(c_{it}^{TAX} + c_{it}^{TR} + c_{it}^S \right) \\
&= \left(\sum_{m \in M_t} F_{imt} x_{imt} \right) - \left((T^F r_{it} + T^R F_{i0t} Q_{it}) + \right. \\
&\quad \left. \left(\sum_{\{m \in M_t \mid m > 0\}} \frac{1}{(1 + \rho)^{t/12}} C^{TR} x_{imt} \right) + ((F_{i0t} + \epsilon) z_{it}) \right)
\end{aligned} \tag{6.10}$$

The complete mathematical formulation of the hedging problem with nested CVaR to guarantee time consistency, follows. A similar recursive model formulation with time

consistent CVaR is also applied by Shütz and Westgaard (2018) to approach hedging of salmon with a multistage stochastic model. The same is also done in Pisciella et al. (2016) who approaches a capacity expansion problem of a price-taking power producer.

$$\max_x v_{00} \quad (6.11)$$

s.t.

$$\begin{aligned} & h_{kt}^N + (1 - \lambda) \sum_{i' \in k+} P_t(i' | k) v_{i'(t+1)} + \\ & \lambda \left(\tau_{kt} - \frac{1}{1 - \alpha} \sum_{i' \in k+} P_t(i' | k) y_{i'(t+1)} \right) = v_{kt} \quad k \in \mathcal{N}_t, t \in T \setminus \{|T|\} \end{aligned} \quad (6.12)$$

$$h_{iT} = v_{iT} \quad i \in \mathcal{N}_T \quad (6.13)$$

$$\tau_{kt} - v_{i'(t+1)} \leq y_{i'(t+1)} \quad i' \in k+, k \in \mathcal{N}_t, t \in T \setminus \{|T|\} \quad (6.14)$$

$$X_t + \sum_{t' \in T_t} \sum_{\{m \in M_{t'} \mid t' + m = t\}} x_{imt'} = Q_{it} + z_{it} \quad i \in \mathcal{N}_T, t \in T \quad (6.15)$$

$$x_{imt} = x_{(i-1)mt} \quad i \geq 1 \quad i \in \{\mathcal{N}_T \mid i \succ k\}, k \in \mathcal{N}_t, m \in M_t, t \in T \setminus \{|T|\} \quad (6.16)$$

$$z_{it} = z_{(i-1)t} \quad i > 0 \quad i \in \{\mathcal{N}_T \mid i \succ k\}, k \in \mathcal{N}_t, t \in T \setminus \{|T|\} \quad (6.17)$$

$$h_{it} = h_{kt}^N \quad i \in \{\mathcal{N}_T \mid i \succ k\}, k \in \mathcal{N}_t, t \in T \setminus \{|T|\} \quad (6.18)$$

$$x_{imt} \geq 0 \quad i \in \mathcal{N}_T, m \in M_t, t \in T \quad (6.19)$$

$$y_{kt} \geq 0 \quad k \in \mathcal{N}_t, t \in T \quad (6.20)$$

$$z_{it} \geq 0 \quad i \in \mathcal{N}_T, t \in T \quad (6.21)$$

Based on the discussion in the previous subsections, the objective with the nested CVaR implementation is formulated in constraints (6.11), (6.12) and (6.13). The essence of this formulation is given by the recursive expression (6.12), where we recognize the deterministic profit term, the expected future profits and the risk. The expression forms a recursive formulation as it incorporates expected utility for the directly succeeding states when calculating the objective for the current state. Since every direct successor also incorporates the expected utility from their direct successors, we get a structure that ultimately nests up the expected utility from the leaves up to the root. The recursive function (6.12) is solved stage-wise for every state in the scenario tree. Also recall that expected profits and risk are weighted according to the decision maker's risk aversion, represented by the parameter λ . The CVaR percentile α also strengthens these effects. A lower CVaR percentile increases the decision maker's propensity towards selling in the spot market.

The objective (6.11) maximizes the objective function of the initial decision stage, while constraints (6.13) represent the objective at the last stage. Since there are no states succeeding the last stage, (6.13) only consists of the deterministic profit term from (6.12).

Constraints (6.15) ensure consistency between produced quantity and quantity delivered through contracts and spot sales at a stage t . We therefore impose a constraint that requires the sum of spot sales and contract sales with maturity at stage t , to be as close to the produced quantity as possible. When the produced quantity exceeds the quantity sold on contract, the excess production is sold in the spot market. Oppositely, we recall from Section 6.1 that in case of shortage production, we assume that the remaining quantity is covered from the spot market.

For constraints (6.15), the quantity sold through forwards includes previously contracts entered within the current planning period ($x_{imt'} \mid m > 0$) and contracts entered in previous planning periods X_t . The reason why we include contracts from previous planning periods is to enable a rolling horizon approach to the decision making. This means that when running the optimization model at sequential stages, optimal decisions from previous planning periods are taken into account in the current planning period (Sethi and Sorger, 1991). For each stage in the planning horizon, the multistage hedging model is then solved again based on new information. Hence, only the first stage solution, which is the quantity of contracts shorted for different maturity periods at the time the model

is solved, is kept from each time the model is solved. The same approach is for instance applied in Séguin et al. (2017).

Constraints (6.16) and (6.17) are non-anticipativity constraints, ensuring that all decisions that are based on the same information ought to be equal. In practice this means that all scenarios that pass through a node have to make the same decision in that node.

Constraints (6.18) ensure that h_{kt}^N is equal to any h_{it} for $i \succ k$. This is a technical adjustment that enables us to use h_{kt}^N as a single term in the node indexed utility function (6.12).

The remaining constraints are non-negativity constraints.

In the next section we develop the distributionally robust hedging model by adapting the stochastic hedging formulation presented in this section to the general solution algorithm for distributionally robust problems, presented in Chapter 5.

6.3 The distributionally robust hedging model

Recall from Section 5.1 that we in distributionally robust optimization find the best decision for the worst-case probability model from a set of probability models in the proximity of the one we have observed. In this way we account for the model risk related to not knowing the underlying true probability distribution for certain.

In Section 5.3 we saw that the solution algorithm consists of three optimization problems. The outer problem that finds the optimal decision for the worst-case scenario tree. The inner problem that runs through every stage in the current worst-case tree to make it even worse. The last problem is the CVaR optimization problem that adapts CVaR to every new modification made to the probabilities in the inner problem. Also recall from Section 5.3.3 that these three problems formed a min-max-min problem, where the objective was a cost function. In the stochastic hedging model presented in Section 6.2 we however consider profits, and seek to maximize the utility. The distributionally robust hedging model we develop in this section therefore becomes a max-min-max problem.

In the following subsection we first modify the stochastic hedging model from Section 6.2, so that it can be applied as the outer problem (5.26) from Section 5.3. This involves

including the probability models within the ambiguity set, and to ensure that the outer problem maximizes the utility function for the probability model that provides the lowest expected utility.

We then consider the inner problem and the CVaR problem. The primary challenge is here to adapt the kernel solution method in Section 5.3.2 to the nested CVaR objective function (6.12).

6.3.1 Modifying the outer problem

In order to apply the stochastic hedging model (6.11) to (6.21) as the outer problem (5.26) in the distributionally robust problem, the following modifications need to be made. The first modification involves adapting the model with respect to the probability models \mathbb{P} in the ambiguity set \mathcal{P}_ε . We hereafter consider probability models as scenario trees. Recall the assumption that the tree structure is fixed, and from the outer problem (5.26) that any feasible decision is bounded by the worst-case scenario tree. We do for these reasons not need to extend all decision variables to include scenarios for every scenario tree in the ambiguity set. Rather, we only have to modify the constraints from the stochastic formulation that contain probabilities, which in this instance applies to the variables of the CVaR formulations (6.12) to (6.14).

We adapt the constraints with probabilities by rewriting the conditional probabilities to $P_{\mathbb{P}_t}(i' | k)$, such that they now are indexed according to their current tree. This modification also affects the auxiliary objective function variables v_{kt} , $y_{i'(t+1)}$ and τ_{kt} , which we now denote $v_{k\mathbb{P}_t}$, $y_{k\mathbb{P}_t}$ and $\tau_{k\mathbb{P}_t}$, respectively.

We then modify the nested CVaR objective function, given by constraints (6.12) and (6.13), so that they apply for every scenario tree in the ambiguity set

$$h_{kt}^N + (1 - \lambda) \sum_{i' \in k+} P_{\mathbb{P}_t}(i' | k) v_{i'\mathbb{P}(t+1)+} + \lambda \left(\tau_{k\mathbb{P}_t} - \frac{1}{1 - \alpha} \sum_{i' \in k+} P_{\mathbb{P}_t}(i' | k) y_{i'\mathbb{P}(t+1)} \right) = v_{k\mathbb{P}_t} \quad k \in \mathcal{N}_t, \mathbb{P} \in \mathcal{P}_\varepsilon, t \in T \setminus \{|T|\}$$

$$\begin{aligned}
h_{iT} &= v_{i\mathbb{P}T} & i \in \mathcal{N}_T, \mathbb{P} \in \mathcal{P}_\varepsilon \\
\tau_{k\mathbb{P}t} - v_{i'\mathbb{P}(t+1)} &\leq y_{i'\mathbb{P}(t+1)} & i' \in k+, k \in \mathcal{N}_t, \mathbb{P} \in \mathcal{P}_\varepsilon, t \in T \setminus \{|T|\}
\end{aligned}$$

A second modification is to reformulate the objective function so that it finds the maximum utility under the worst-case probability distribution. We first introduce the auxiliary variable u . We then apply a constraint assuring that u cannot be bigger than than the lowest root node objective variable $v_{0\mathbb{P}0}$

$$u \leq v_{0\mathbb{P}0} \quad \mathbb{P} \in \mathcal{P}_\varepsilon$$

By then setting the objective as

$$\max u$$

we know that the binding constraint is the one with the lowest of all $v_{0\mathbb{P}0}$, and that u therefore maximizes with regards to the worst-case distribution.

6.3.2 Formulating the inner problem

Recall from Section (5.3) that in order to solve the inner problem (5.27), we run a successive convex programming algorithm that iterates through every stage in the scenario tree. It then returns the scenario probabilities that make the expected objective function value as bad as possible. The inner problem objective is therefore similar to that of the outer problem, but with other decision variables. In the instance of the electricity hedging problem (6.11) to (6.21), we have to make the following modifications to the original formulation of Analui and Pflug (2014):

- The recursive formulation of the objective (6.11) - (6.13) requires us to consider the utility function for every stage, rather than for the terminal values at the leaf nodes. This requires us to formulate conditional probabilities as variables for every

stagewise iteration. The respective adaptations also have to be made to the kernel formulations.

- A further problem is to integrate the objective formulation with nested CVaR, which we recall from Section 5.3.3 turns the inner problem non-linear.
- The inner problem-CVaR iterations from Section 5.3.3 is already an extension of the risk neutral solution algorithm from Analui and Pflug (2014). This procedure has to be adapted to the nested CVaR formulation used in the hedging problem.

We first adapt the inner problem to the recursive objective formulation. We then make the adaptations to the inner problem to address the non-linearity of integrating nested CVaR. This allows us to formulate the hedging specific inner problem. Lastly, we adapt the CVaR optimization problem from Section 5.3.3 to the recursive objective formulation.

Adapting the inner problem to the recursive objective function

As discussed in Section 5.3, the inner problem in the original framework of Analui and Pflug (2014) inputs only the optimal objective values at the leaf stage from the outer problem. Recall that in this context the optimal objective values are profits. These profits are the realized profits at the end of each scenario path in the scenario tree, when having input the outer problem optimal decision variables. The worst-case scenario tree is then calculated so that the expected scenario profits are minimized. An intuitive example is the inventory control model presented in Section A.3. Here, the scenario profits are the accumulated profits from sales at every stage in the planning period.

The original accumulated objective function formulation does not fit well with the recursive structure we use to make CVaR time consistent. We emphasize that in the objective of the stochastic hedging model (6.11) it is the root node utility that is maximized. Attempting to accumulate the utility from the root down to the leaf nodes would eventually nest up the constraints to one single constraint, where all objective function variables cancel each other out. In addition to make little sense, the problem then becomes infeasible. An illustrative example is presented in Appendix A.4.

In the following paragraphs we first present the modifications of the inner problem because of the recursive objective formulation. We also present a set of necessary transportation kernel adaptations.

Presenting the approach Formulating the inner problem of Analui and Pflug (2014) for a recursive formulation of the objective is to our knowledge not covered in the literature. We therefore propose the following approach: We know from the stochastic problem (6.11) to (6.21) that the objective consists of sub-objective functions at each node and every stage, that calculate the node-wise expected utility. Recall from (6.12) that this recursive structure is written as

$$h_{kt} + (1 - \lambda) \sum_{i' \in k+} P_t(i' | k) v_{i'(t+1)} + \lambda \left(\tau_{kt} - \frac{1}{1 - \alpha} \sum_{i' \in k+} P_t(i' | k) y_{i'(t+1)} \right) = v_{kt} \quad k \in \mathcal{N}_t, t \in T \setminus \{|T|\} \quad (6.22)$$

We seek to adapt this formulation to Algorithm 5.2, which solves the inner problem by updating scenario tree probabilities one stage at a time. Note that in (6.22) the expected utility is calculated from the subtrees spanning out from each node at every intermediate stage in the scenario tree. We therefore have to formulate the conditional probabilities for every stagewise iteration. As can be seen in (6.22) it is in particular the conditional probability of reaching a node from its direct predecessor that needs to be calculated.

Recall that the inner problem objective for a given iteration is to adjust the probabilities, such that the utility obtained from the outer problem decision is as low as possible. Following constraints (6.22), this means that the inner problem has to find the conditional probabilities such that the expected value over all current stage objective variables v_{kt} are minimized. Because of the recursive structure, the new values of these variables propagate up along the path of predecessors up to the root node. Whereas the outer problem maximizes the root node utility, it is minimized in the inner problem. Note that we define the decision variables over the successor nodes of those at the current stage. In other words, the inner problem adapts the conditional probabilities of reaching a successor node, conditional on its predecessor at the current stage, so that the root node utility is minimized.

Formulating the stagewise objective function We now show how the the minimal expected root node utility can be formulated for every inner problem iteration. The prin-

principle is to formulate the root node utility as the expected leaf node utility in the subtrees spanning from the root down to the considered stage in the current iteration.

First, we note that the probability of reaching a current stage node is given from the path of predecessors. More specifically, we show that all other terms in (6.22) become constants when calculating the root node utility based on the current stage utilities. Note that it therefore is fine to fix the successor node utilities $v_{j'(t+1)}$. They are adjusted in the following iterations anyway.

Remember that we in Chapter 5 distinguished the new scenario tree, with changed probabilities, from the baseline scenario tree. Recall that the new scenario tree node $l \in \mathcal{N}_t$ corresponds to nodes k at the same stage in the baseline scenario tree. We similarly write the direct successor as $j' \in l+$. Further, we hereafter distinguish parameter probabilities $\hat{P}_t(j' | l)$ from variable probabilities $p_t(j' | l)$. We now show that the root node utility, which is captured in the root node objective variables v_{00} , can be written based on the current stage objective variables v_{lt} . For ease of notation we for now disregard the CVaR term. Consider the root node utility

$$v_{00} = h_{00} + \sum_{j' \in \mathcal{N}_1} P_0(j' | l_0) v_{j'1}$$

For node l at stage 1 the same utility function is

$$v_{l1} = h_{l1} + \sum_{j' \in l+} P_1(j' | l_1) v_{j'2}$$

By inserting the latter function into the first we get

$$v_{00} = h_{00} + \sum_{j' \in \mathcal{N}_1} P_0(j' | l_0) \left(h_{l1} + \sum_{j' \in l+} P_1(j' | l_1) v_{j'2} \right) \quad (6.23)$$

If we assume that the current iteration is at stage 2, so that we for a given node l have

the utility

$$v_{l_2} = h_{l_2} + \sum_{j' \in l_+} p_2(j' | l_2) v_{j'_3} \quad (6.24)$$

then adjusting $p_2(j' | l_2)$ makes v_{l_2} vary accordingly. This does not affect any of the profit terms h_{kt} in (6.23), which therefore can be considered constant in the current iteration. Adding the CVaR term to (6.24) would change the value of v_{l_2} , but would not affect the constant terms in (6.23). If we denote the constant profit terms as C , we can write the root node utility as the sum of current stage utilities, weighted by their path probabilities. We here refer to l_1 and l_0 as predecessors of l .

$$v_{00} = C + \sum_{l \in \mathcal{N}_2} P_0(l_1 | l_0) \cdot P_1(l_2 | l_1) v_{l_2}$$

Generally for the current iteration t we can write

$$v_{00} = C + \sum_{l \in \mathcal{N}_t} (P_{(t-1)}(l | l_-) \cdot \dots \cdot P_0(l_1 | l_0)) v_{j_2}$$

Since C does not affect the optimal decision it can be disregarded. We further denote the path of conditional probabilities $P_{(t-1)}(l | l_-) \cdot \dots \cdot P_0(l_1 | l_0)$ down to current stage node l as $P_t^{(new)}(l | l_0)$. We can now write the objective as

$$\min_{p_t(j' | l)} \sum_{l \in \mathcal{N}_t} P_t^{(new)}(l | l_0) v_{lt}$$

Formulating conditional probabilities with subkernels We emphasize that conditional probabilities only need to be calculated for the direct successor nodes $P_t(j' | l)$. Recall from the conditional probability kernel example (5.35) in Section 5.3.1 that for

direct successor conditional probabilities, the path of subkernels from relation (5.32) contains only a single subkernel. According to relation (5.34), we formulate the relation between kernels and conditional probabilities as

$$P_t(j' | l) = \sum_{i' \in k+} K_t(j'_{(t+1)} | i'_{(t+1)}; k, l) \cdot \hat{P}_t(i | k) \quad l \in \mathcal{N}_t \quad (6.25)$$

We set this relation as constraints in the inner problem, and thus get a relation between kernels and probabilities. As we know from Section 5.3.2, the actual decision variables in the inner problem are the subkernels at the current stage. Constraints (6.25) therefore restrain the optimization problem from finding conditional probabilities that cannot be formulated with the adjustable subkernels at that stage. The constraints that ensure consistency between kernels are the same as those described in Section 5.3.2.

Addressing non-linearity of the CVaR term

We now consider the CVaR term of the utility function. Recall from Section 5.3.3 that CVaR causes the inner problem to become non-linear. As we can see for constraints (6.12), this is caused by the product of the probability variables and the CVaR auxiliary variables in the terms

$$(1 - \lambda) \sum_{i' \in k+} p_t(j' | l) v_{i'_{(t+1)}}$$

and

$$\lambda \left(\tau_{kt} - \frac{1}{1 - \alpha} \sum_{i' \in k+} p_t(j' | l) y_{i'_{(t+1)}} \right)$$

Recall that this non-linearity issue is the reason why we in Section 5.3.3 extended Algorithm 5.1 such that it includes a CVaR optimization problem within the inner problem. This allows us to approximate the inner problem probabilities as if they were decided

simultaneously as the CVaR variables.

Nevertheless, in order to linearize the inner problem, we need to determine which variables to fix. To our knowledge there is no literature addressing this issue. Our two options are

1. Fixing $v_{j'(t+1)}$ and $y_{j'(t+1)}$ and allowing τ_{lt} to adapt according to the decision variables. As a consequence τ_{lt} and $p_t(j' | l)$ can be determined simultaneously, which is ideal. A problem is on the other hand that shortfall profit $y_{j'(t+1)}$, which by definition is a difference, is fixed.
2. An alternative is fix τ_{lt} and then replace $p_t(j' | l) y_{j'(t+1)}$ with the auxiliary variable w_{lt} . In this way we can allow $p_t(j' | l)$ to vary, while at the same time ensuring that the shortfall profit is consistent with τ_{lt} . We denote the fixed values of τ_{lt} and $v_{j'(t+1)}$ as \mathcal{T}_{lt} and $V_{j'(t+1)}$ respectively. We still need to ensure that the outer problem CVaR constraints (6.14), which are written

$$\tau_{kt} - v_{i'(t+1)} \leq y_{i'(t+1)} \quad i' \in k+, k \in \mathcal{N}_t$$

are satisfied. We do this by multiplying both sides with $p_t(j' | l)$ and get

$$p_t(j' | l)\mathcal{T}_{lt} - p_t(j' | l)V_{j'(t+1)} \leq w_{j'(t+1)} \quad j \in l+, l \in \mathcal{N}_t \quad (6.26)$$

A drawback with this approach is that τ_{lt} is fixed, which is suboptimal as the shortfall profits in this case are not calculated relative to the VaR threshold, with the intended confidential level α .

Alternative 2 is in our opinion the most attractive as it allows the shortfall profit to vary. While a fixed τ_{lt} on one hand is suboptimal, fixing the shortfall profit $y_{i'(t+1)}$ does on the other make little sense when calculating CVaR. The equivalence between the new inequality (6.26) and (6.14) can easily be seen by solving a linear equality and compare it to the same inequality, but with each side multiplied with a scalar.

We now have a linear formulation where conditional probabilities can be adjusted to

minimize the stagewise utility, with shortfall profit that adapts to these decisions. We can now proceed towards formulating the inner optimization problem for a given stage in the stagewise iteration of Algorithm 5.2.

Inner problem formulation

We now know that the purpose of the inner problem is to minimize the root node utility. This is done by adjusting current stage conditional probabilities with the use of the subkernels at the current stage. We recall that the root node utility is formulated as a product of the path of direct successor probabilities $P_t^{(new)}(l | l_0)$ and the current stage objective variables v_{lt} for all $l \in \mathcal{N}_t$. For the root node this probability is set to 1. From the outer problem, the inner problem inputs the following optimal variable values as parameters: The optimal profits H_{kt} , the successor stage optimal objective values $V_{j'(t+1)}$ and the Value at Risk \mathcal{T}_{kt} . We first write the nomenclature and then present the inner problem.

Sets

\mathcal{N}_t	Set of nodes in the current stage iteration.
\mathcal{N}_T	Set of leaf nodes.
$k+$	Set of direct successors i' of k .
$l+$	Set of direct successors j' of l .

Indices

k	Node index in the old scenario tree	$k \in \mathcal{N}_t$
i'	Direct successor node i' of k	$i' \in k+$
l	Node index in the new scenario tree	$l \in \mathcal{N}_t$
j'	Direct successor node j' of l	$j' \in l+$

Parameters

d_{ij}^r	The r-order distance between paths down to leaf nodes i and j .	$i, j \in \mathcal{N}_T$
H_{lt}	Profit function with fixed variables.	$l \in \mathcal{N}_t$
$V_{j'(t+1)}$	Successor stage objective function.	$j \in l+, l \in \mathcal{N}_t$
\mathcal{T}_t	Current state Value at Risk.	$l \in \mathcal{N}_t$
\hat{P}_i	Baseline model scenario probability.	$i \in \mathcal{N}_T$
$\hat{P}_t(i' k)$	Baseline model direct successor conditional probability.	$k \in \mathcal{N}_t$
$P_t^{(new)}(l l_0)$	Path of direct successor conditional probabilities down to node l .	$l \in \mathcal{N}_t$
ε	Ambiguity radius.	
λ	Risk aversion parameter.	
α	CVaR confidence level.	

Variables

$K(i, j)$	Transportation kernel.	$i, j \in \mathcal{N}_T$
$K_t(j' i'; k, l)$	Subkernel at stage t .	$k, l \in \mathcal{N}_t$
$p_t(j' l)$	Conditional probability of reaching direct successor node j' of l .	$j \in l+, l \in \mathcal{N}_t$
v_{lt}	Current stage objective value.	$l \in \mathcal{N}_t$
$w_{j'(t+1)}$	Auxiliary variable describing weighted expected loss for node j' .	$j' \in l+, l \in \mathcal{N}_t$

$$\min_{p_t(j' | l)} \sum_{l \in \mathcal{N}_t} P_t^{(new)}(l | l-) v_{lt} \quad (6.27)$$

s.t.

$$H_{lt} + (1 - \lambda) \sum_{j' \in l+} p_t(j' | l) V_{j'(t+1)} + \lambda \left(\mathcal{T}_{lt} - \frac{1}{1 - \alpha} \sum_{j' \in l+} w_{j'(t+1)} \right) = v_{lt} \quad l \in \mathcal{N}_t \quad (6.28)$$

$$p_t(j' | l) \mathcal{T}_{lt} - p_t(j' | l) V_{j'(t+1)} \leq w_{j'(t+1)} \quad j \in l+, l \in \mathcal{N}_t \quad (6.29)$$

$$\sum_{i \in k+} K_t(j'_{t+1} | i'_{t+1}; k, l) \hat{P}_t(i' | k) = p_t(j' | l) \quad j \in l+, k, l \in \mathcal{N}_t \quad (6.30)$$

$$C_{ij}^{(k,l)} K_t(j'_{t+1} | i'_{t+1}; k, l) = K(i, j) \quad i \succ k, j \succ l, (k, l) \in \mathcal{N}_t \quad (6.31)$$

$$\sum_{j \in \mathcal{N}_T} \sum_{i \in \mathcal{N}_T} d_{ij}^r K(i, j) \hat{P}_i \leq \varepsilon \quad (6.32)$$

$$\sum_{j' \in l+} K_t(j'_{t+1} | i'_{t+1}; k, l) = 1 \quad i' \in k+, (k, l) \in \mathcal{N}_t \quad (6.33)$$

$$\sum_{j \in \mathcal{N}_T} \sum_{i \in \mathcal{N}_T} K(i, j) \hat{P}_i = 1 \quad (6.34)$$

$$K_t(j' | i'; k, l) \geq 0 \quad i \in k+, j \in l+, k, l \in \mathcal{N}_t \quad (6.35)$$

$$w_{j'(t+1)} \geq 0 \quad j' \in l+, l \in \mathcal{N}_t \quad (6.36)$$

The objective (6.27) minimizes the expected value of the current stage objective variables, conditional on the preceding nodes.

This objective variable inputs the current stage utility function (6.28), which is similar to the utility function in the outer problem. For every node in the current stage, we vary the conditional probabilities of reaching the direct successor leaf nodes. Once a probability is varied and the objective function is changed, the shortfall profit in the CVaR term is adjusted accordingly.

The shortfall profit is ensured consistent as the positive difference between the fixed Value at Risk and the current objective function value by constraints (6.29).

As elaborated in the paragraphs above, constraints (6.30) ensure consistency between subkernels and conditional probabilities. Similarly, constraints (6.31) ensure that relation (5.32) holds by defining the transportation kernel as the path of subkernels from the root node. Recall that future and past stage subkernel values are stored in the $C_{ij}^{(k,t)}$ parameter.

Constraint (6.32) limits the nested distance between the new probability model and baseline model to be no more than ε . Recall from Section 5.2 that this constraint constructs the boundaries of the ambiguity set, which can be regarded as the nested distance ball (5.18).

Constraints (6.33) ensure that the total mass transported to all j from a given i' , conditional on k and l , forms a probability distribution that adds up to 1.

Similarly, constraint (6.34) requires that the new probability distribution, formed as the product of kernels and the baseline probability distribution, adds to 1.

Since subkernels form probability distributions, constraints (6.33) ensure that they are non-negative and no larger than 1. Lastly, the non-negativity constraints (6.36) ensure that shortfall profits are positive.

As discussed in the sections above, fixing the auxiliary CVaR variables in this problem is not optimal. With the purpose of finding an approximation of the CVaR values as if they were decided simultaneously as the inner problem probabilities, we iteratively run the inner optimization against the CVaR optimization problem. In the following section we elaborate on how the CVaR optimization problem is formulated.

CVaR optimization formulation

Before the inner problem continues on a new stagewise iteration, the optimal inner problem decision is input to the CVaR optimization problem to see whether if an updated Value at Risk can make the probability distribution even worse. In a similar manner as the inner problem runs from stage to stage, the CVaR optimization problem runs from node to node within the current stage. Therefore, for the optimization problem at a given

node l and stage t , we only have a single CVaR optimization problem decision variable τ_{lt} . The optimization problem considers only the CVaR term within the inner optimization problem for a given node l , and the uncertainty determined by the direct successors j' . We input the optimal probabilities from the inner problem $P_t(j' | l)$, and the optimal direct successor objective variables from the outer problem $V_{j'(t+1)}$. Since these values are fixed, we can formulate the shortfall profit balance constraint as is done in constraints (6.14) in the outer problem. For a given node l and stage t we formulate the CVaR optimization as

Sets

$l+$ The direct successors of l

Indices

j' Direct successor node j' of l . $j' \in l+$

Parameters

$V_{j'(t+1)}$ Successor stage objective function values from the outer problem. $j \in l+$

$P_t(j' | l)$ Optimal stagewise probabilities from the inner problem. $j' \in l+$

α CVaR confidence level.

Variables

τ_{lt} CVaR decision variable for the current node l at stage t .

$y_{j'(t+1)}$ Shortfall profit. $j' \in l+$

$$\max_{\tau_{lt}} \tau_{lt} - \frac{1}{1 - \alpha} \sum_{j' \in l+} P_t(j' | l) y_{j'(t+1)} \quad (6.37)$$

s.t.

$$\tau_{lt} - V_{j'(t+1)} \leq y_{j'(t+1)} \quad j' \in l+ \quad (6.38)$$

$$y_{j'(t+1)} \geq 0 \quad j' \in l+ \quad (6.39)$$

The objective function (6.37) is the CVaR term of the utility function, and constraints (6.38) and (6.39) ensure consistency between the CVaR variables.

Even though including the CVaR optimization problem improves the quality of the inner problem decisions, this is still an approximation. In the following chapter we discuss the computational results found by running the hedging model on a real data instance. One of these tests is to study how including the CVaR optimization problem impacts the worst-case probability distribution and the corresponding optimal decisions.

Chapter 7

Computational Study

The purpose of this chapter is to present the results of the hedging models developed in Chapter 6. This chapter is organized into two sections. In Section 7.1 information about the problem instance in this thesis is described, while the numerical results of the multistage stochastic and multistage distributionally robust hedging model are presented and analyzed in Section 7.2.

7.1 Problem instance

In this section we describe the input data applied in the hedging problem covered in this thesis. First, details about the scenario tree applied in the hedging models, are presented. These details include a description of the input data that the scenario tree is created from. Secondly, a description of the test data and the rolling horizon approach is presented. In addition, the remaining parameters that need to be defined in order to run the hedging models are quantified.

7.1.1 Scenario tree and input data

In this subsection we discuss the specific scenario tree used in this thesis, in addition to a presentation of the input data the scenario tree is generated from. While both the electricity spot price and production quantity scenarios are generated based on the same

method, the forward price scenarios are generated differently. This subsection is therefore organized in two. In the first part we present the input data for prices and production, and describe how scenarios are generated based on the input data. In the second part we similarly present a description of the input data and scenario generation for the forward prices. In brief, for the spot prices and production quantities, the scenario tree is generated from a scenario fan based on the clustering approach described in Section 5.4.1. The scenarios for forward prices are further constructed such that they are consistent with the spot price scenarios, according to the description in Chapter 6.

Scenario generation for spot prices and production quantities

In creating the scenarios of spot prices and production quantities, the scenario tree generation method from Section 5.4.1 is applied. Further, a scenario fan of forecasts from the EFI's Multi-area Power-market Simulator (EMPS) is applied as a basis for the applied scenario tree generation method. EMPS is a market equilibrium model, developed by SINTEF for optimization and simulation of hydro thermal power systems based on supply and demand of electricity (SINTEF, 2017). The model is for instance applied in forecasting inflow scenarios, which are highly correlated with production quantities, and in forecasting electricity price scenarios. The EMPS model uses statistical time series of 50 to 75 years as a basis when generating the different forecasts.

We first describe the data extracted from the EMPS model, and show how this data is processed in order to create the scenario fan that the scenario tree generation method is based on. The scenario fan from the EMPS model consists of 55 scenarios for each week over a time horizon of approximately 4 years. Further, each scenario gives a pair of spot prices and production quantities, based on market equilibrium calculations from one historical time series of inflow. The correlation between spot prices and production quantities is hence accounted for. TrønderEnergi considers only the 20 hydropower plants of most significant size as relevant for the hedging problem regarded in this thesis. In the EMPS model the spot prices and production scenarios are based on the assumption that the market participants are price takers in a market of perfect competition. This is a reasonable assumption for TrønderEnergi, which is a small company.

As the hedging models consider a time horizon of 4 months, with a monthly duration of each stage, the hourly price forecasts from the EMPS model are averaged, while the

production forecasts are aggregated over each month before being applied in the scenario tree generation approach. We assume that the inflow scenarios for similar months are independent from year to year. This allows us to apply all scenarios from the EMPS model as forecasts for the year which we run the optimization models for. Hence, the scenario fan is yielding a total of 220 ($55 \cdot 4$) scenarios.

As described in Section 5.4, a scenario fan as the one output from the EMPS model is not appropriate for multistage distributionally robust optimization models, because all information is received at the second stage of the fan, with no remaining uncertainty at the succeeding stages. Hence, in this thesis a scenario tree is generated from forecasts from the EMPS model, based on the clustering approach described in Section 5.4.1.

For the scenario generation approach we specify prices and production as uncertain variables, i.e. set $m = 2$. Recall from Section 5.4.1 that the scenario generation approach ensures that the correlation from the EMPS model between these dimensions is retained. Ultimately, the r -order distance is set to 2 as it simplifies the calculations, hence making the Euclidian distance the chosen distance metric. An illustration of how a scenario fan is transformed into a scenario tree is shown in Figure 7.1. Note that neither the scenario fan, nor the tree, are displayed with real scenario values as this information is confidential.

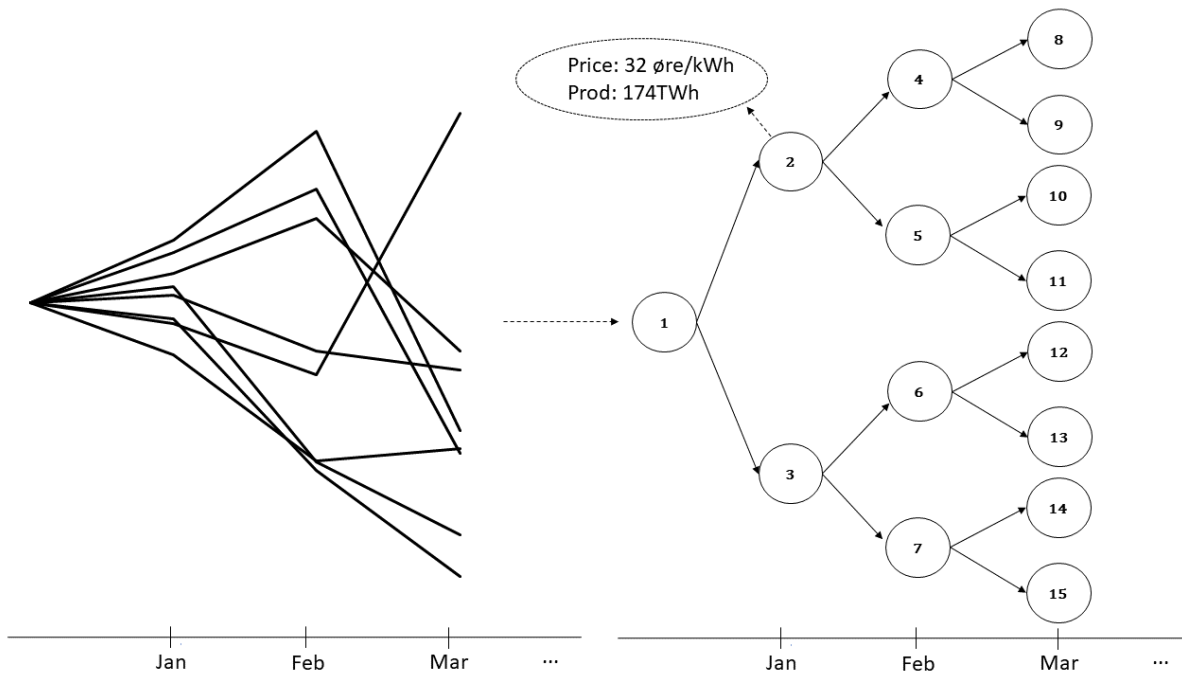


Figure 7.1: A scenario tree (right) is generated from a scenario fan (left) based on stochastic approximation. In the hedging models, the scenario fan consists of 220 forecast paths from the EMPS model. The generated scenario tree contains 5 stages and 192 scenarios at the last stage, where each scenario contains information about the spot price and production.

The structure of the generated scenario tree is chosen as the one that minimizes the nested distance to the scenario fan. Table 7.1 gives a summary of the alternative tree structures, the respective number of scenarios per stage and their average nested distance over five generated scenario trees. The structure represents the branching factor, i.e. the number of direct successors at each stage, for every node in the tree. The scenario generation method is implemented in MATLAB and the average computational time for the preferred structure, in **bold** in Table 7.1, is 26 seconds when $\epsilon = 0.0005$.

Structure	Scenarios pr stage	nested distance
6-4-3-3	6-24-72-216	2.5
4-4-4-3	4-16-64-192	1.9
3-3-4-6	3-9-36-216	3.1
3-4-4-4	3-12-48-192	2.4

Table 7.1: Alternative structures for the scenario tree, including the number of scenarios per stage and average nested distance to the scenario fan. The chosen structure is in **bold**.

Although the scenario generation approach enables us to generate a scenario tree from the EMPS model, a drawback is as we recall from Section 5.4, that the number of scenarios at the last stage in the tree is limited to the total number of scenarios from the EMPS model (220). Since we consider a five stage model, this limits the branching factor at each node and hence making it difficult to capture all potential outcomes. From table 7.1, we see that the tree with structure 4-4-4-3 and 192 scenarios gives the lowest nested distance. This structure is therefore applied in the optimization models.

Scenario generation for forward prices

Even though the prices on forward contracts with different maturities are known at the point of time they can be entered, the development of the forward prices is uncertain as time approaches the maturities. For instance, the price of a forward contract maturing three months from now is known today, but unknown tomorrow. Hence, as for spot prices and production quantities it is necessary to represent the uncertainty of the forward prices in the scenario tree.

As discussed under the assumptions in Chapter 6, this thesis only considers hedging, and not speculation, as the underlying motivation for entering into forward contracts. The forward prices are therefore required to be consistent with the conditional expectation of the spot price scenarios. By employing this, scenarios for the forward price can be derived as illustrated in Figure 7.2. Further, Figure 7.3 illustrates a forward curve, a graph showing current forward prices for contracts with different maturities. This graph clearly shows the seasonality in electricity prices, an observation that is common in the electricity market.

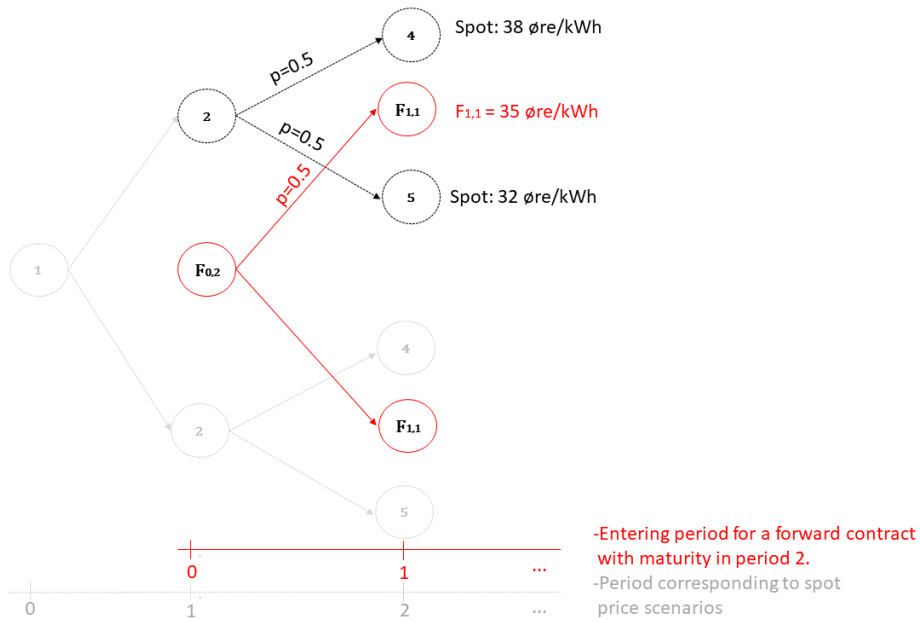


Figure 7.2: A scenario tree for the price of a forward contract with maturity in period 2. The forward price at a stage is consistent with the conditional expectation of the spot price at the maturity period

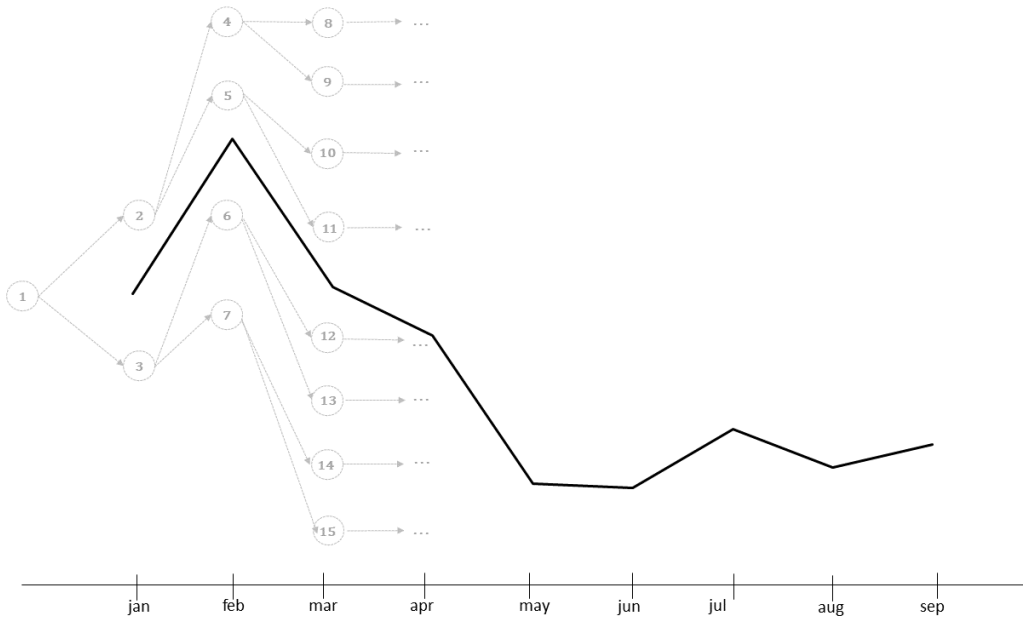


Figure 7.3: The forward curve for monthly contracts with maturities from January to September. The forward price is consistent with the expectation of the spot price at maturity.

7.1.2 Test data and the rolling-horizon procedure

The hedging models are run with monthly decision stages over a planning horizon of 5 months, for each year ranging between 2014 and 2017. Recall from Chapter 6 that the decision problem is to decide the proportion to be sold either on contract, or to an uncertain future spot price. For each year a rolling horizon approach is applied. This means that the multistage hedging models are solved five times each year, where the first stage decisions are stored each time a model is solved. These decisions are then input as parameters when the model is solved again for the next month. The initial decision is made at the end of December in the preceding year, hedging the prices from January to April. The model is then solved for January with a new scenario tree based on an updated fan from the EMPS model, hedging for February to May, and so forth. We assume that the only contracts with maturity within the planning period are relevant for the decision. TrønderEnergi is further assumed to hold no contracts at the initial decision stage, for a given year.

The transaction costs are set to 0.01 EUR/MWh (Nord-Pool, 2017b), and the buyback cost in cases where the quantity of electricity sold on contract exceeds the actual production for the same period, is set to 1EUR/MWh. Further, the discount rate ρ is set to 0.05.

Ultimately, the stochastic hedging model is solved for different degrees of risk aversion, λ , and CVaR confidence level α . The distributionally robust hedging model is also solved for varying ambiguity radius ε . According to Esfahani and Kuhn (2015), the ideal ambiguity radius should be large enough to capture the true probability distribution, but low enough so that the model does not give too conservative results. In this instance, we solve the distributionally robust hedging model for $\varepsilon = 2$ to $\varepsilon = 50$. Recall that the nested distance between the generated scenario tree and the input forecasts from the EMPS model was approximately 2, hence $\varepsilon = 2$ is chosen as the minimum ambiguity radius. As discussed in Section 5.1, we are able to capture the distributional differences between the applied scenario tree and input data. Further, higher ambiguity radii captures the decision makers increasing ambiguity aversion.

7.1.3 Computational performance

Before proceeding towards analyzing the numerical results, we make a brief comment on the runtime of the optimization models. Solving the distributionally robust model with a stop criterion of observing improvements less than 1% over the last three iterations, takes approximately 400 seconds. This corresponds to a 68 times higher computational effort than for the stochastic counterpart. Whether or not one should apply distributionally robust optimization with this framework should for this reason rely on the importance of computational efficiency. For instance, high need for frequently running the optimization problem would reduce the usefulness of the distributionally robust model. For running the optimization model on monthly basis, as is in this instance, distributionally robust optimization with the applied framework is adequate from a computational efficiency viewpoint.

7.2 Numerical results

In this section we present and interpret the results from having run the distributionally robust hedging model. Our approach is to examine the results from three perspectives. First, we present and evaluate the hedging decisions obtained from the respective models. We here aim to characterize how decisions made by the distributionally robust model differ from those of the stochastic model, and further discuss why we observe these differences. Secondly, we compare the differences in expected implications of applying decisions of the stochastic model against the distributionally robust model. Lastly, we conduct out-of-sample performance tests on the respective hedging strategies, which include simulations from potential probability distributions as well as backtesting on realized data. This is to investigate the extent to which an electricity producer can benefit from accounting for ambiguity when making hedging decisions.

We first present an in sample stability test in order to investigate the consistency of the scenario generation method.

7.2.1 In Sample Stability

The purpose of an in sample stability test is to verify if the applied scenario generation approach is consistent (Séguin et al., 2017). The test is based on King and Wallace (2012), and is conducted by investigating the variation in the optimal objective value for different scenario trees. In this hedging context, a relatively stable optimal objective is desirable as it indicates that the optimization model gives approximately equal expected utilities independent on the applied scenario tree. For a discussion on why it is more interesting to consider optimal objective values than optimal decisions when measuring in sample stability, see King and Wallace (2012).

We generate six scenario trees based on the scenario fan from January 2014, before running the hedging model independently on each of the trees. The test is conducted for both the stochastic model, as well as the distributionally robust model with two different degrees of ambiguity. The results are shown in Table 7.2. For the stochastic model, we observe that the optimal utilities are relatively stable for the different trees. Further, for the distributionally robust model we observe that the optimal objective values are slightly more sensitive to varying scenario trees. Intuitively, one could expect the in sample stability to increase for higher degrees of ambiguity since the distributionally robust model accounts for model risk. The explanation of why the in sample stability decreases is however that we observe the worst-case scenarios in the scenario tree to commonly be outliers, which tend to vary more than the rest of the scenario tree. Recall that the applied distributionally robust model handles model risk by accounting for more conservative probability distributions, which is achieved by weighting the worst-case scenarios more. It is therefore expected that the in-sample stability of the distributionally robust model is lower when the worst-case scenarios of the trees the test is conducted on tend to vary more than the rest of the scenario tree.

Data	SO			DRO, $\varepsilon = 5$		DRO, $\varepsilon = 10$	
	Tree	Utility, MEUR	diff, abs	Utility, MEUR	MEUR	Utility, MEUR	MEUR
Jan, 2014	1	10.025	1.2%	9.921	1.3 %	9.787	1.5 %
	2	10.025	1.2%	9.881	1.7 %	9.709	2.3 %
	3	10.152	0%	10.033	0.2 %	9.851	0.9 %
Jun, 2014	4	10.180	0.3%	10.063	0.1 %	10.065	1.3 %
	5	10.223	0.7%	10.142	0.9 %	9.955	0.2 %
	6	10.294	1.4%	10.259	2.1 %	10.259	3.2 %
Mean	10.150	0.8%	10.050	1.0%	9.938	1.6%	

Table 7.2: In-sample stability test for the stochastic hedging model (SO) and the distributionally robust model with $\varepsilon = 5$ and $\varepsilon = 10$. The difference is the represented as the percent-wise absolute deviation from the mean.

In the following section we present and evaluate the hedging decisions obtained from the stochastic- and distributionally robust hedging model.

7.2.2 Presenting and evaluating the hedging decisions

The purpose of this section is to understand how the hedging decisions from the distributionally robust model differs from the stochastic model. In order to uncover the principal differences between the models we first study the aggregated hedge-ratios over the four years of test data, for different levels of ambiguity. We then seek to explain these results by elaborating on how the scenario tree in the distributionally robust model changes with increasing ambiguity radius. We further study the detailed set of hedging decisions within a year in order to find differences in when contracts are entered, when they mature and the quantities hedged. Lastly, we conduct a sensitivity analysis on how varying the CVaR significance level and the degree of risk aversion affect optimal decisions.

Figure 7.4 shows the average hedge ratios in the period from 2014 to 2017 for the stochastic model and the distributionally robust model with different levels of ambiguity ε . For the four future months included in a planning horizon, the hedge ratios are calculated as the sum of all contracts that have maturity at a specific month, relative to the expected production in that month. The expected production for a month is calculated from the empirically based scenario tree \hat{P} . Further, the monthly expected hedge ratios are averaged in order to obtain the yearly expected hedge ratios. The CVaR significance level

and level of risk aversion are respectively set to $\alpha = 0.9$ and $\lambda = 0.5$.

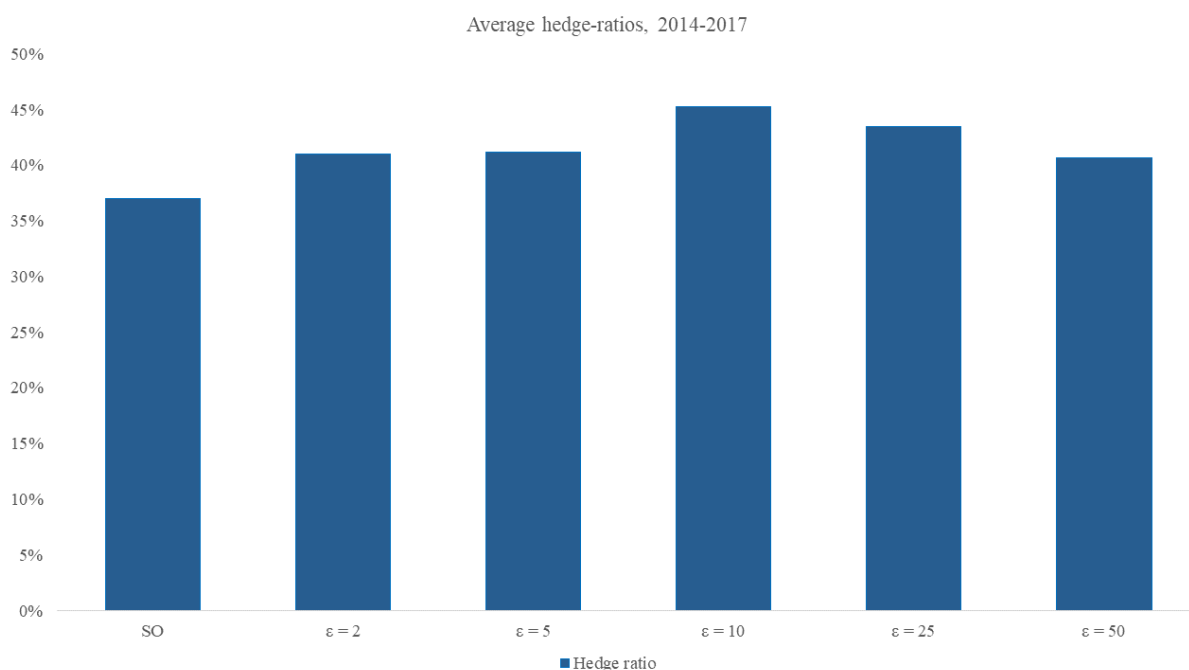


Figure 7.4: Average yearly hedge ratios over the period 2014-2017, achieved by the stochastic and distributionally robust hedging model with varying ambiguity radius.

On average we observe that the expected hedge ratios are approximately 40%. As we recall from Section 2.2, the tax system causes the position that neutralizes price risk under a deterministic production to be at 53%, which we hereafter refer to as a full hedge. This means that the models on average recommend hedging about 75% of the risk neutralizing hedge ratio. We observe that the expected hedge ratios tend to increase when the ambiguity radius ε increases, approaching a full hedge. This increased conservatism for higher ambiguity aversion is in accord with what one would expect from a more robust solution. As we however elaborate in the sections below, there are several factors influencing this relation. In general, the degree of conservatism is largely decided by the worst-case probability model in the ambiguity set. Our findings suggest that for small to moderate ε , an increased focus on low-price scenarios, and therefore an increased tail risk, incentivizes more hedging. A counteracting effect is however that low volume scenarios also are assigned more weight for increasing ambiguity. Hence, these hedging decisions are based upon estimates of lower production, causing lower hedged quantities. When ε

becomes very large, extremely low production levels tend to become the most significant contributor to low profits in the scenario tree. In a few of these scenarios the prices might even be higher than the forward price, despite that the correlation between spot prices and production quantities in general is positive. These effects explain the finding that the hedge ratios for $\varepsilon = 25$ and $\varepsilon = 50$ in average are lower than for the more moderate choices of ε .

Explaining the relationship between hedge ratios and increased ambiguity

In order to understand why hedging decisions change when ε increases, we study how the worst-case probability distribution evolves according to variations in ε . We therefore run the distributionally robust model on the same data for a broad specter of ε , and observe the evolution in the scenario probabilities in the scenario trees. Recall that the higher the ambiguity radius, the further we allow the worst-case distribution to deviate from the baseline probability model. Further, since we optimize for the worst-case probability model, we would expect that an increased ambiguity radius would narrow down the scenario probabilities to revolve around a few specific scenarios that cause the lowest profits and highest risk. In Figure 7.5, we present the findings for $\varepsilon = 0, 7.5, 10, 15, 20, 50$. For illustrative purposes, we have aggregated the total 192 scenarios into 16 discrete scenarios. Further, note that the depicted scenario tree is not intended to have the same structure as the one used in the model. It does rather serve the purpose of illustrating that a scenario probability represents a given path through the scenario tree.

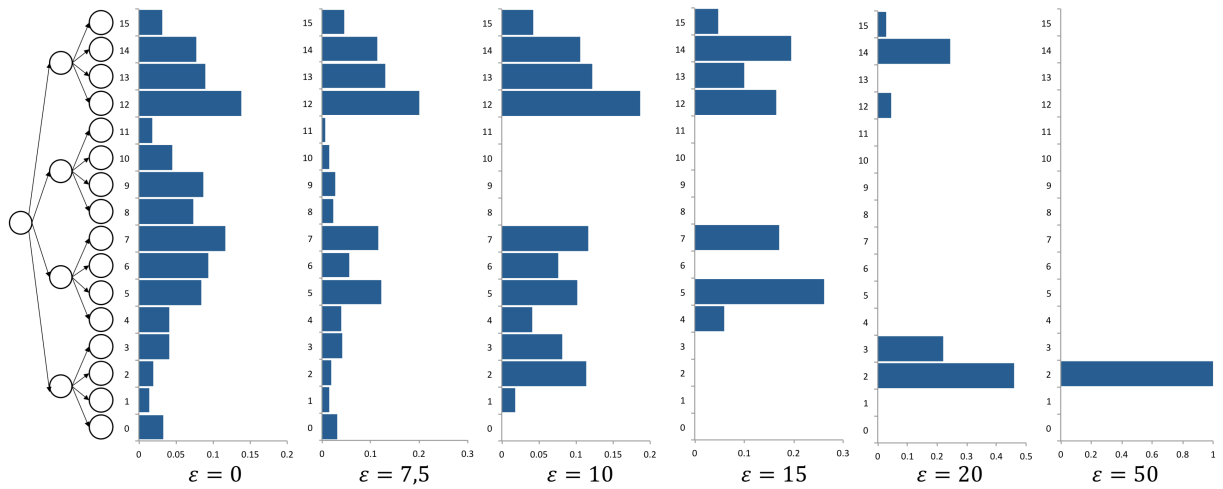


Figure 7.5: When the ambiguity radius increases, the worst-case scenario trees narrow down to single scenarios. Note that the scenarios are not presented in the order of increasing profits.

From Figure 7.5 we observe that when ε increases the tree complexity is reduced, including less and less scenarios. In the extreme case we observe that the worst-case probability distribution considers only the single scenario causing the most harm to the decision maker, making the model deterministic. This is in accord with how the scenario tree in the examples in Analui and Pflug (2014) evolves when the ambiguity radius increases. Recall that the structure of the scenario trees is the same, while only the scenario probabilities themselves are subject to variations.

In the following paragraphs we elaborate on how the changes in the probability model for increasing ε affects the hedge ratios. As mentioned in the previous subsection, a first finding is that the hedge ratios are affected by two counteracting effects. For higher values of ε we observe an increased weighting of the downside profit scenarios, which in general involves the scenarios with low prices. In turn, this increases the tail risk which makes higher hedge ratios more attractive. In addition, increased weighing of scenarios with low prices reduces expected future spot prices, which further reduces attractiveness of spot price exposure. A full hedge should occur if the deterministic forward price exceeds the expected spot price under deterministic production. On the other hand, a factor that moderates the effect of higher hedge ratios as ε increases, is that a higher ε also tends to increase the weights of scenarios with low production. Isolated, this incentivizes towards hedging lower quantities, which further gives lower hedge ratios relative to the baseline expected production. Too high hedge ratios under uncertain production involves a risk,

as it potentially leaves the decision maker incapable of supplying the agreed quantity in scenarios where the produced quantity is low.

From Figure 7.4 we recall that hedge ratios increase for higher ε , but only up to some point. We also observed in Figure 7.5 that increasing ε makes the model more and more deterministic, as fewer and fewer worst-case scenarios are increasingly weighted. A further explanation for why we observe the different hedging decisions can be derived by observing Figure 7.6 below, showing how expected price and production in every month are reduced when ambiguity increases.

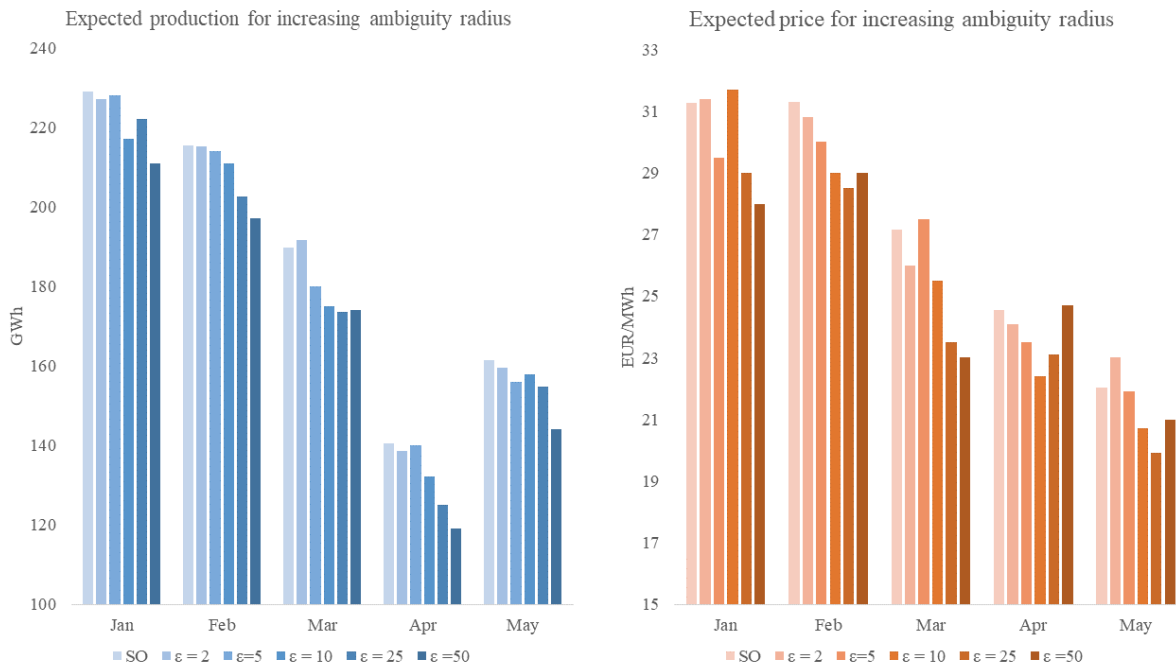


Figure 7.6: Expected production for increasing ambiguity radius (left), and expected spot price for increasing ambiguity radius (right). The figure is based on the optimizations conducted on an arbitrary year within the considered time horizon.

In general we observe the previously discussed trend of decreasing expected prices and production for increasing ε . However, for some values of ε we observe the opposite. In these cases a high ε has a higher expected price than for lower ε . For April it is even the case that the expected price under $\varepsilon = 50$ is higher than in the stochastic model. Seemingly paradoxically, even though expected profits are lower for high ε , the narrow focus on a few worst-case scenarios where the production is extremely low, in some

instances can cause reduced hedging.

A hypothesis is therefore that this effect is eliminated when production risk is removed. In this situation an increased ε only increases the weighting of the low price scenarios, which further gives strictly increasing hedge ratios. The following Figure 7.7 confirms this hypothesis. Note that to avoid immediate full hedges when increasing ε , a risk-premium is added.

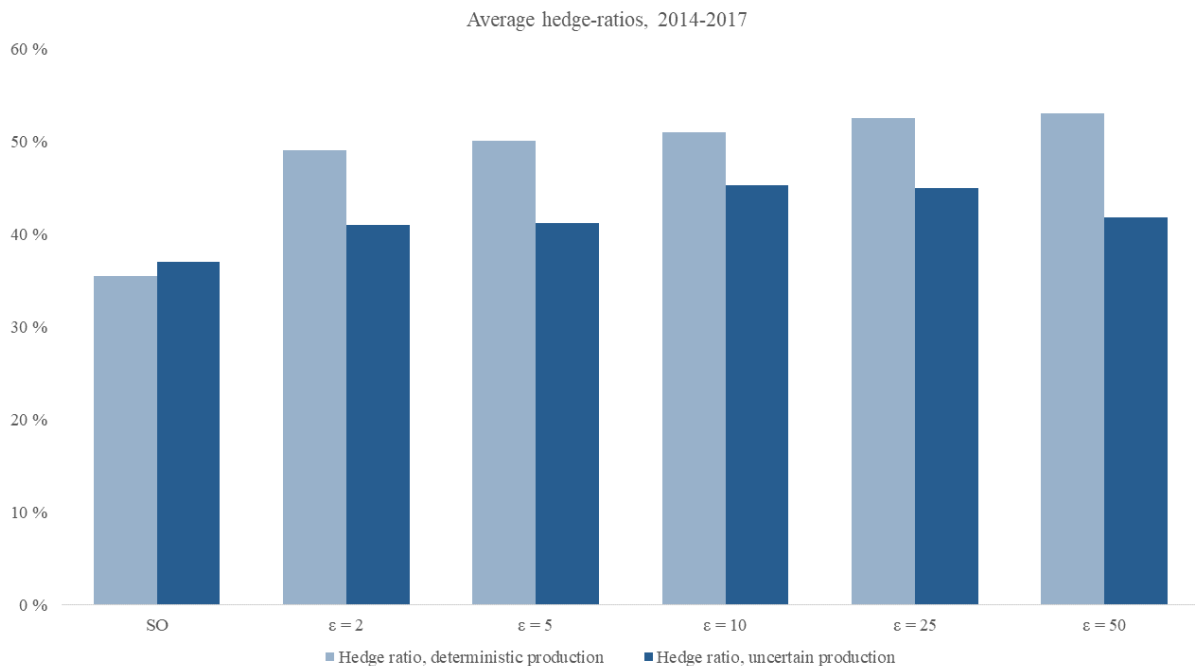


Figure 7.7: Expected hedge ratios for increasing ambiguity, illustrated with uncertain production (dark columns) and with deterministic production (light columns).

By eliminating the uncertainty in production, the previously discussed counteracting effect of increasing ambiguity radius is also removed.

Summarized we generally observe higher hedge ratios for increased ambiguity, following more weighting of the scenarios with low prices, which subsequently increases the risk. Lower expected production does on the other hand moderate the effect of increased hedge ratios. For very high ambiguity we observe decreased hedge ratios as a consequence of a more and more deterministic scenario tree, which narrows the focus on scenarios that at times have very low production. This makes the average hedge ratios lower for

the highest degrees of ambiguity. In the following paragraphs we regard the hedging decisions on a more detailed level, seeking to uncover differences between the stochastic and distributionally robust model concerning the time when contracts are entered, their maturities, and the quantities hedged. We also perform a sensitivity analysis to see the effects of modifying risk the risk aversion parameters.

Detailed hedging decisions and risk sensitivity analysis

Detailed hedging decisions The following results of the distributionally robust ($\varepsilon = 5$) and stochastic model are run on rolling horizon over five months of test data for 2017, with monthly decisions. For every decision the first-stage decisions from the previous planning period are kept. Figure 7.8 considers a single year and shows hedge ratios over the first five months, distinguishing between the month when contracts were entered. As for the figures presented in the preceding section, hedge ratios are relative to expected production for the baseline probability model. CVaR significance level and risk aversion are unchanged, respectively $\alpha = 0.9$ and $\lambda = 0.5$.

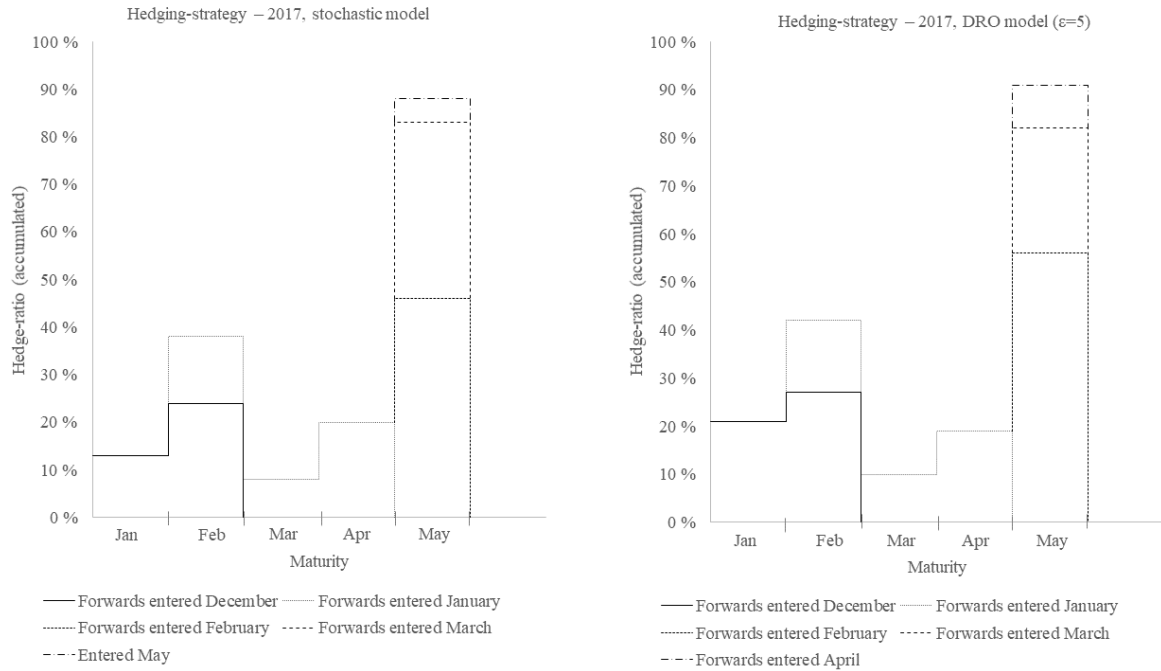


Figure 7.8: Hedge ratios for the considered months in 2017, split on entrance date of the contracts. The strategies from the stochastic model (left) and distributionally robust model with $\varepsilon = 5$ (right) are shown

For both models in Figure 7.8 we observe relatively low hedge ratios in March and high hedge ratios in May, which is because the scenario tree captures a higher degree of price risk in May than in March. In May the hedge ratios are at almost 90% of the expected production, meaning that the model overhedges relatively to a full hedge of 53% for a deterministic production. This is likely due to the uncertain production. A notable difference between the decisions from the stochastic and distributionally robust model for this year is that the distributionally robust model already in February hedges a more considerable proportion of the expected production in May than the stochastic model. The stochastic model does on the other hand adapt to the same hedge ratio in March. This is according to what one would expect from more risk averse actors, as there is a risk related to the development of forward prices with a specific maturity. However, as seen by studying the detailed hedging decisions for the other years of test data, which are attached in Appendix B.1, earlier contract entries are not significantly more common in the distributionally robust model.

Sensitivity analysis In the following paragraphs we study how the solutions differ when varying the CVaR significance level α and risk aversion parameter λ . Recall that for higher values of α the impact of losses exceeding VaR is increased, hence the risk increases. Similarly, an increased λ increases the decision maker's weighting of risk relative to expected profits. How variations in α affects the decisions as expected. We expect that a higher confidence level would lead to more conservative decisions, which as in Figure 7.9 is what we observe for all months when α increases.

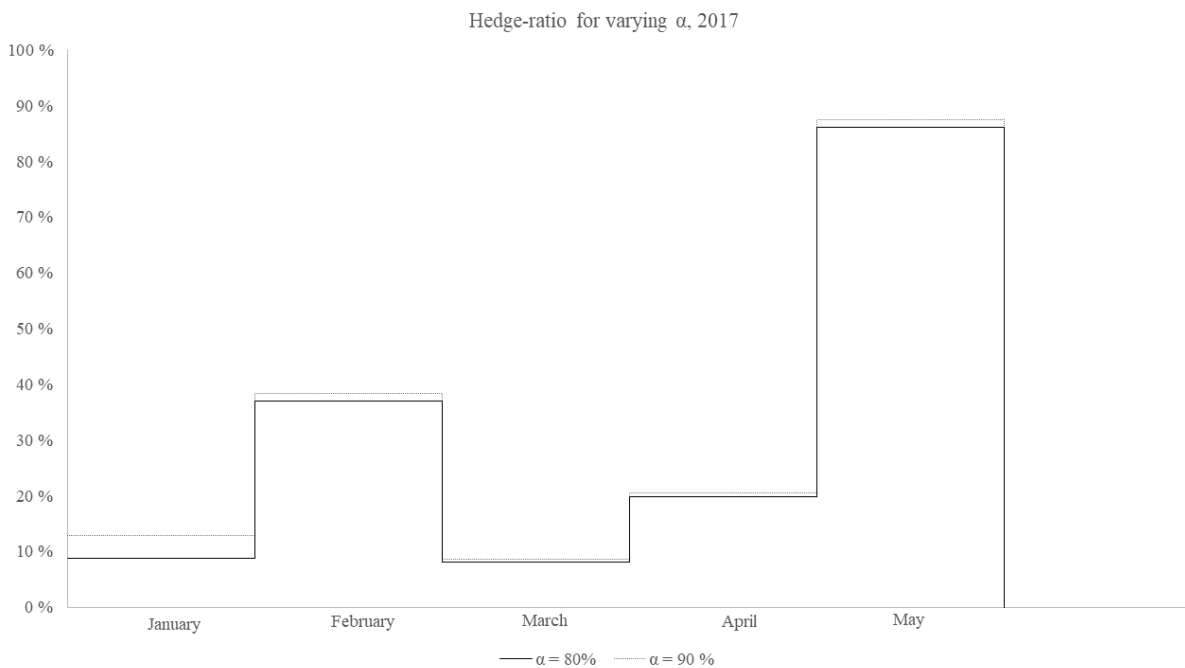


Figure 7.9: The monthly hedge ratios for 2017 for various levels of α .

We also study how varying λ affects the decisions in the stochastic and distributionally robust model. Figure 7.10 shows the hedge ratios for the considered months in 2017, for different values of the risk aversion, λ . The purpose is to study the effects of increasing risk aversion on the hedging strategies of the stochastic and distributionally robust optimization model.

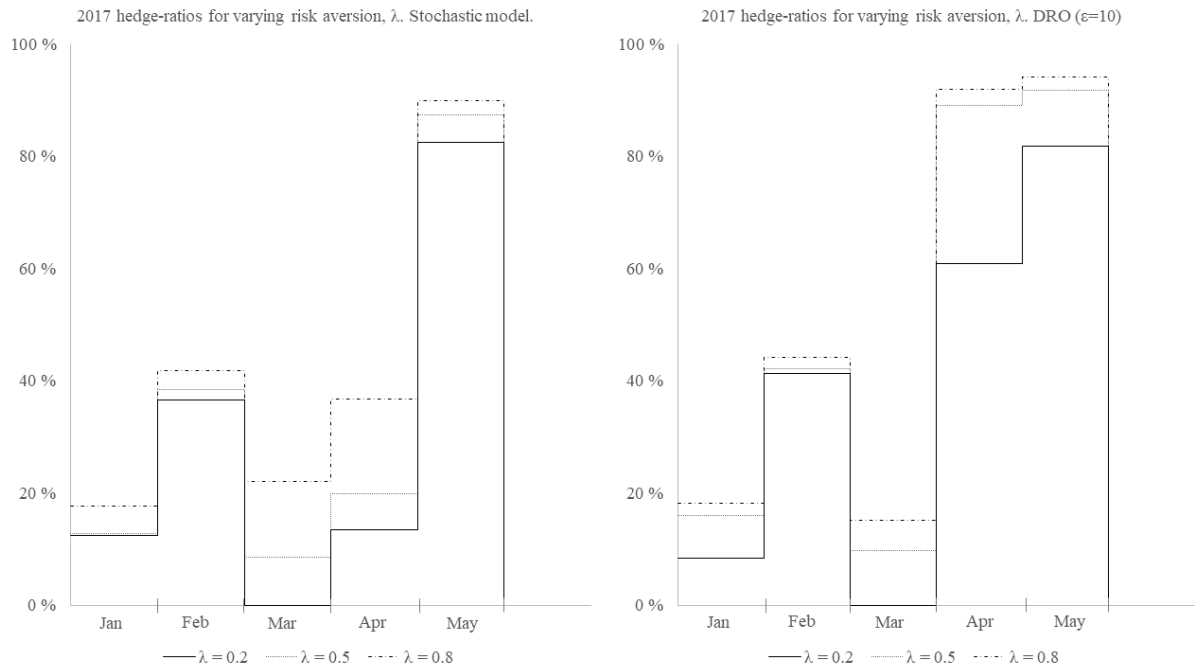


Figure 7.10: The monthly hedge ratios for 2017 for various levels of risk aversion, λ . The stochastic model (left) and distributionally robust model with $\epsilon = 10$ (right) are shown.

As expected, a higher degree of risk aversion gives higher hedge ratios. The most considerable difference between the two models is how they perceive the risk in April. While the stochastic model has relatively low hedge ratios in this month, the distributionally robust here weights the price risk as significantly higher, and therefore has a higher hedge ratio. This graph does however not show when contracts were entered. This information is exemplified for different contract types with maturity in February 2017 by the following Figure 7.11, which shows how hedge ratios vary with increasing λ for different types of contracts.

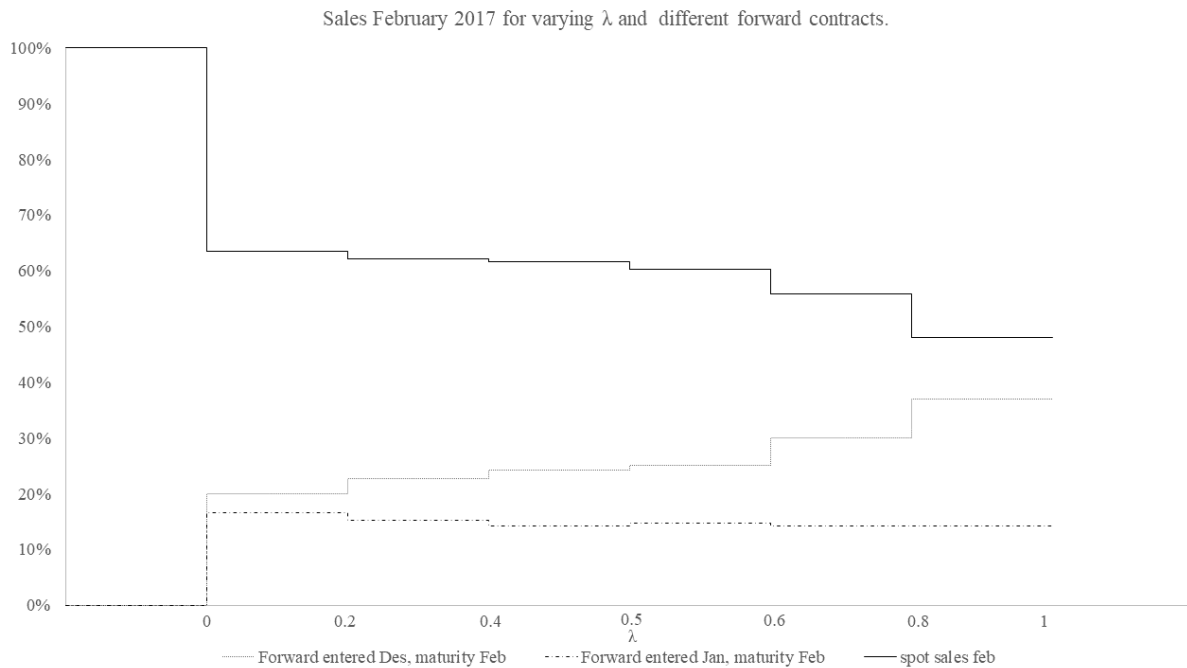


Figure 7.11: Hedge ratios in February 2017 for different types of contracts from the stochastic model.

Increasing risk aversion reduces spot sales and contracts entered in January, while contracts entered in December increase. This is in line with what one would expect from risk averse actors, who have increased willingness to enter into contracts earlier. This is a similar finding as those in Shütz and Westgaard (2018). We also observe that by increasing risk aversion, the optimal hedge ratios approach a full hedge of 53%. The results are taken from the stochastic model, but the findings are similar for the distributionally robust model.

Remarks on the overall findings

A question that arises is how the effects of increasing ambiguity on the optimal decisions differ from the effects of increasing the risk aversion. Recall that the principal difference between risk aversion and ambiguity aversion is that while risk aversion applies to the uncertainty within an observed probability model, ambiguity aversion subjects the observed probability model itself to uncertainty.

The common observation for both increased ambiguity and risk aversion is increased hedge ratios. A factor that on the other hand solely applies when increasing the ambiguity aversion under situations where production risk also is present, is that more conservative production scenarios also are weighted more, as we saw in Figure 7.6. This effect incentivizes towards hedging smaller quantities. Besides of this we have not uncovered any apparent differences between the decisions made by the distributionally robust model and those of the stochastic. Understanding further differences between risk aversion and ambiguity aversion is an interesting venue for future research.

We have up to this point evaluated the differences between decisions of the distributionally robust and the stochastic model. In the following section we compare the expected implications of applying decisions from the stochastic model compared to decisions from the distributionally robust model. Specifically, we focus on the expected performance under varying distributional assumptions.

7.2.3 The expected results of implementing the models

The purpose of this section is to evaluate the expected benefits of having applied a distributionally robust model instead of a stochastic model, in cases when there is distributional uncertainty. We therefore compare the performance of the decisions from each model under different assumptions about the realized probability model. Lastly, we evaluate the performance of the distributionally robust model with and without the CVaR-inner problem iterations explained in Section 5.3.3.

Comparing the performance under different realized probability models

We are interested in comparing how the optimization models perform if the baseline probability model is realized, and similarly how they perform if the worst-case probability model is realized. This allows us to evaluate the value of applying distributionally robust optimization. Since the objective function is an utility function based on the trade-off between risk and expected profits, the performance is measured as the expected utility under the given assumptions about the realized probability model. We could alternatively have used expected profits, but this would disregard the risk term in the objective.

We study the decisions made in January 2014. In order to increase the significance of our findings we generate 6 scenario trees on the January 2014 data, and find the optimal decisions for every scenario tree under different levels of ambiguity. Our approach follows that of Pflug and Pichler (2014c), where the model performance is evaluated from two perspectives, the instance when the realized probability model in fact is the baseline model $\hat{\mathbb{P}}$ and the one where the realized probability model is the worst-case model $\mathbb{P}^* \in \mathcal{P}_\varepsilon$. Naturally, one pays a price for making the solution distributionally robust. Comparing the two perspectives allows us to compare the loss of having robustified the solution against the loss suffered if \mathbb{P}^* is realized. We denote the distributionally robust solution $x^*(\mathcal{P}_\varepsilon)$ and the stochastic solution $x^*(\hat{\mathbb{P}})$. The value of having implemented the distributionally robust model is measured by two key figures.

The first key figure is a measure of the expected utility sacrificed for making the solution distributionally robust. It is therefore called the *price of ambiguity*. Specifically, the price of ambiguity is defined as the expected utility loss of having implemented $x^*(\mathcal{P}_\varepsilon)$ instead of $x^*(\hat{\mathbb{P}})$ in the instance when $\hat{\mathbb{P}}$ is the realized probability model. The price of ambiguity is defined as

$$\mathbb{E}_{\hat{\mathbb{P}}}[U(x^*(\hat{\mathbb{P}}), \xi)] - \mathbb{E}_{\hat{\mathbb{P}}}[U(x^*(\mathcal{P}_\varepsilon), \xi)] \quad (7.1)$$

The second key figure considers the instance when the worst-case probability model \mathbb{P}^* is realized. We are here interested in measuring how much the decision maker gains from having robustified the solution instead of having chosen the stochastic model. For this reason the key figure is called the *gain for distributional robustness*. Note that \mathbb{P}^* is the \mathbb{P} that minimizes the objective value. The key figure is then written

$$\min_{\mathbb{P} \in \mathcal{P}_\varepsilon} \mathbb{E}_{\mathbb{P}}[U(x^*(\mathbb{P}), \xi)] - \min_{\mathbb{P} \in \mathcal{P}_\varepsilon} \mathbb{E}_{\mathbb{P}}[U(x^*(\mathcal{P}_\varepsilon), \xi)] \quad (7.2)$$

An increased ambiguity radius includes worse and worse probability models. This does on one hand increase the conservatism of the distributionally robust solution, but does on the other hand reduce the potential downside of the stochastic solution. Therefore, both key figures increase when the ambiguity radius increases. We show that TrønderEnergi can gain from implementing the distributionally robust model when facing model risk. The reason is that gain for distributional robustness increases faster than the cost of

distributional robustness. Consider the following presentation of how these key figures evolve for an increasing ambiguity radius

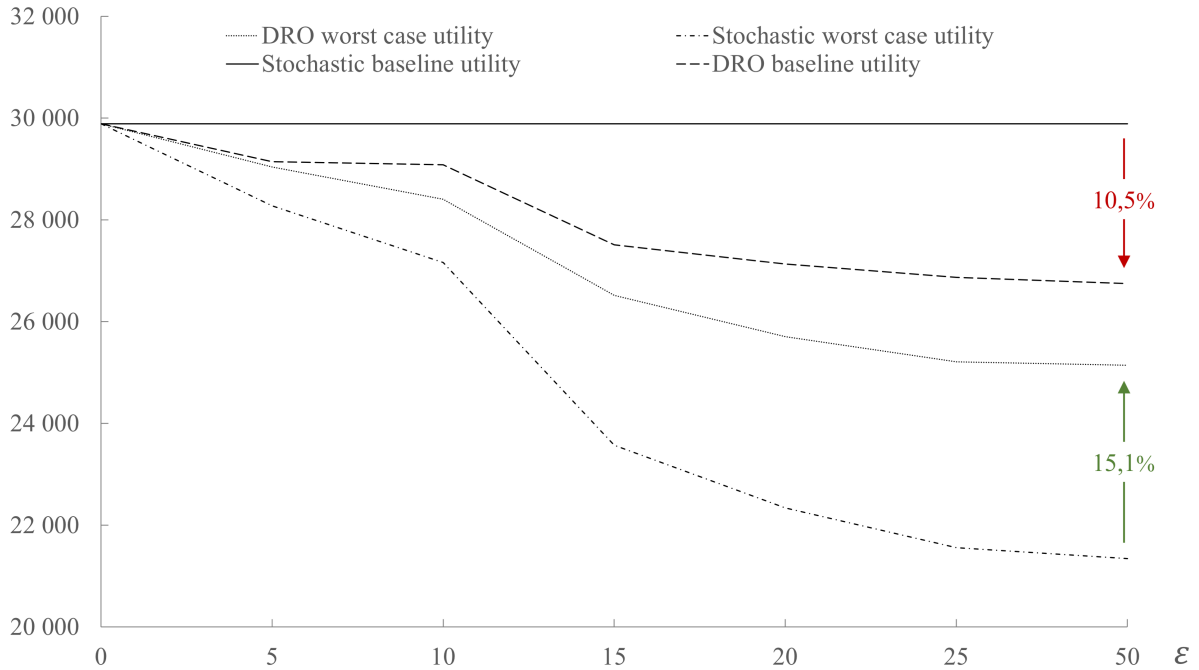


Figure 7.12: The price of ambiguity, marked as difference between the *solid* and the *dashed* line, and the gain for distributional robustness, marked as the difference between the *dotted* and the *dash-dotted* line. The gain for distributional robustness is higher than the price of ambiguity.

In the extreme setting of $\varepsilon = 50$ the cost of ambiguity at 10,5% shows that the decision maker is expected to sacrifice this proportion to robustify the solution. On the other hand, the gain for distributional robustness at 15,1% shows that the worst-case expected utility is increased more than what is traded off for robustness. Note that these findings are in line with the results in the simple inventory control example presented in Analui and Pflug (2014).

A practical interpretation of these results is that planning for profits of 25 000 and potentially attaining an additional 1 500, is more preferable situation than planning for profits of 30 000 but possibly ending up with only 22 000. Note that since the objective in this instance considers utility, to relate profits directly to these figures is not entirely precise. The principle holds all the same.

Further elaborations on the calculations and data are described in Appendix B.2. The

input data is here presented in tables from which one can confirm what can be glimpsed from Figure 7.12, that gain for distributional robustness in this instance exceeds the cost of ambiguity for any choice of ε .

In order to study whether these results apply out-of-sample, we conduct the same analysis under simulations under different probability models in Section 7.2.4.

The effects of including the inner problem-CVaR iterations

We here study how including the iterative CVaR optimization procedure from Section 5.3.3 and Section 6.3.2 affects the performance of the distributionally robust optimization problem. We study two versions of the distributionally robust optimization problem, one with the CVaR optimization iterations and one without. Recall that the CVaR optimization problem updates CVaR for every new worst-case distribution provided by the inner problem, while the one without the CVaR optimization problem only adjusts CVaR in the outer problem. We therefore expect the version that includes the CVaR optimization problem to find a worse probability distribution than the version without the CVaR optimization problem. We compare the two versions by running them on equivalent data from January 2014. In Figure 7.13 below we show how the objective value of the two versions decreases for every outer problem iteration

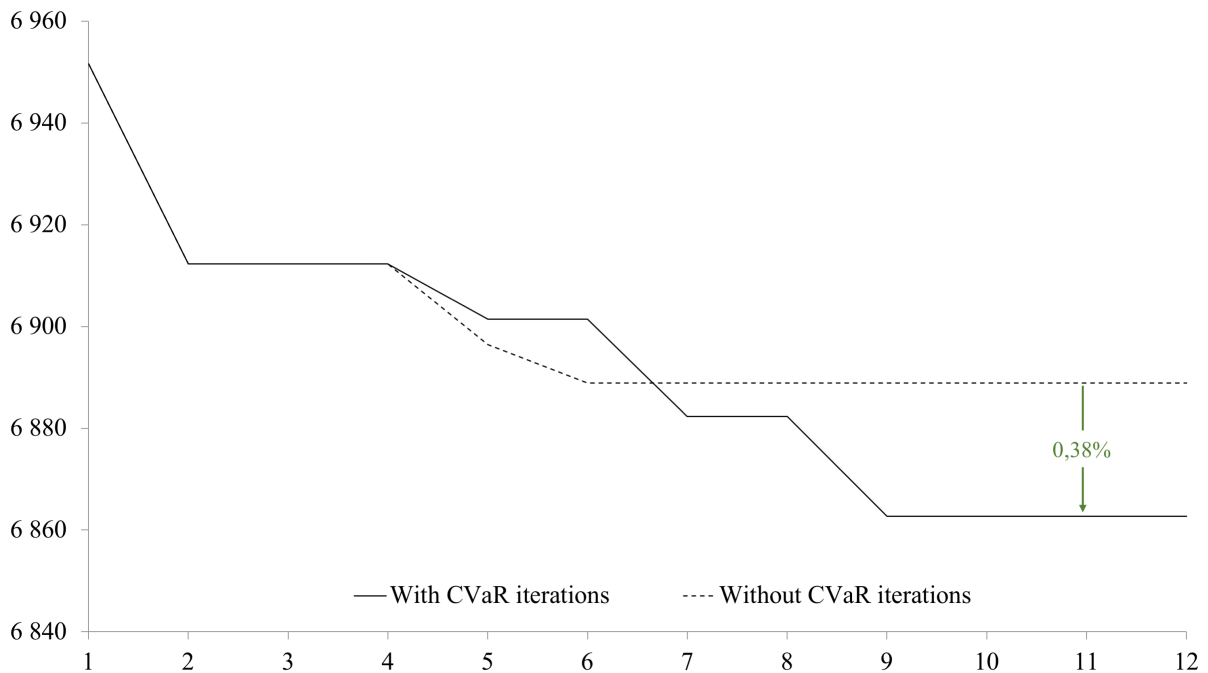


Figure 7.13: Reduction in objective function value per outer problem iteration for problems with and without the CVaR problem iterations.

As expected, we see an improvement in the objective value of the problem that includes the CVaR iterations. This is however a very slight improvement, which comes at the cost of considerably decreased computational efficiency. We also observe that the problem without CVaR iterations converges after fewer iterations than the alternative approach. As a consequence, fixing CVaR in the inner problem might be viable if computational efficiency is more preferable than finding the optimal solution.

We have seen that distributionally robust optimization is beneficial when there is distributional uncertainty, because the gain for distributional robustness exceeds the price of ambiguity. The benefit of using distributionally robust optimization is amplified by iteratively updating CVaR between the inner problem iterations. Having studied the expected performance of the two models, we can proceed towards studying the out-of-sample performance of the two optimization models on simulated and historical data.

7.2.4 out-of-sample performance

The purpose of this subsection is to investigate the out-of-sample performance of both the stochastic and distributionally robust hedging model. In doing this, we conduct simulations as well as backtesting on historical spot prices and production quantities. A naive hedging strategy with no hedging is also tested and applied as a benchmark for the hedging models. This represents a risk neutral strategy. When evaluating the performance of the respective models we focus on the risk and potential profits, under varying degrees of ambiguity aversion.

Simulations

The applied simulation approach is based on Banks et al. (2010) and involves randomly generating potential realizations of spot prices and production quantities. Here, one realization is simulated by randomly drawing a path of spot prices and associated production quantities from the EMPS model. Recall that the EMPS model is the set of forecast paths that the scenario tree is generated based on. We therefore assume that the EMPS model covers the potential realizations of spot prices and produced quantities. The performance of the respective hedging strategies is further evaluated based on multiple simulations.

The primary purpose of the simulations is to investigate the robustness of the stochastic and distributionally robust hedging model against distributional uncertainty. An optimization model has high distributional robustness if deviations from the empirically observed probability model has low impact on the performance of the optimization model. Testing for distributional robustness is achieved by studying the performance of the respective models under simulations from the baseline probability distribution, as well as from potential worst-case distributions. Specifically, we first simulate over the baseline probability model, where the scenarios in the EMPS model are equally weighted. We then simulate again, but this time with higher probabilities assigned to the low-profit scenarios in the EMPS model.

Although simulation does not give the actual realizations of the spot price and production quantity for the considered time horizon, it enables us to conduct multiple imitations of the outcome, making it possible to analyze expected profits and risk with significant

precision. Simulation also permits us to predetermine the probabilities of the potential spot price- and production scenarios, hence making it possible to test the performance under alternative probability distributions.

The simulations are conducted to evaluate the hedging strategies from 2015, where the subjective risk aversion λ is set to 0.5, and CVaR is calculated with a 90% confidence level. The risk aversion in the stochastic and distributionally robust hedging decisions are therefore equal, while the ambiguity varies. Note that all following hedging decisions can be found in tables in Appendix B.1.

We first present a brief presentation of the optimal decisions that the following simulations are based on. The average hedge ratios over the planning horizon for the stochastic and distributionally robust model, under varying degrees of ambiguity, are presented in Figure 7.14. This figure gives no information regarding the maturity or entrance of forward contracts. The purpose of showing the figure is, however, to illustrate that the general hedging activity is higher for the distributionally robust than for the stochastic model, and tends to increase further with higher degrees of ambiguity aversion. Ultimately, for the highest degrees of ambiguity aversion, the average hedge ratios are slightly lower. This is in line with the discussion in Section 7.2.2.

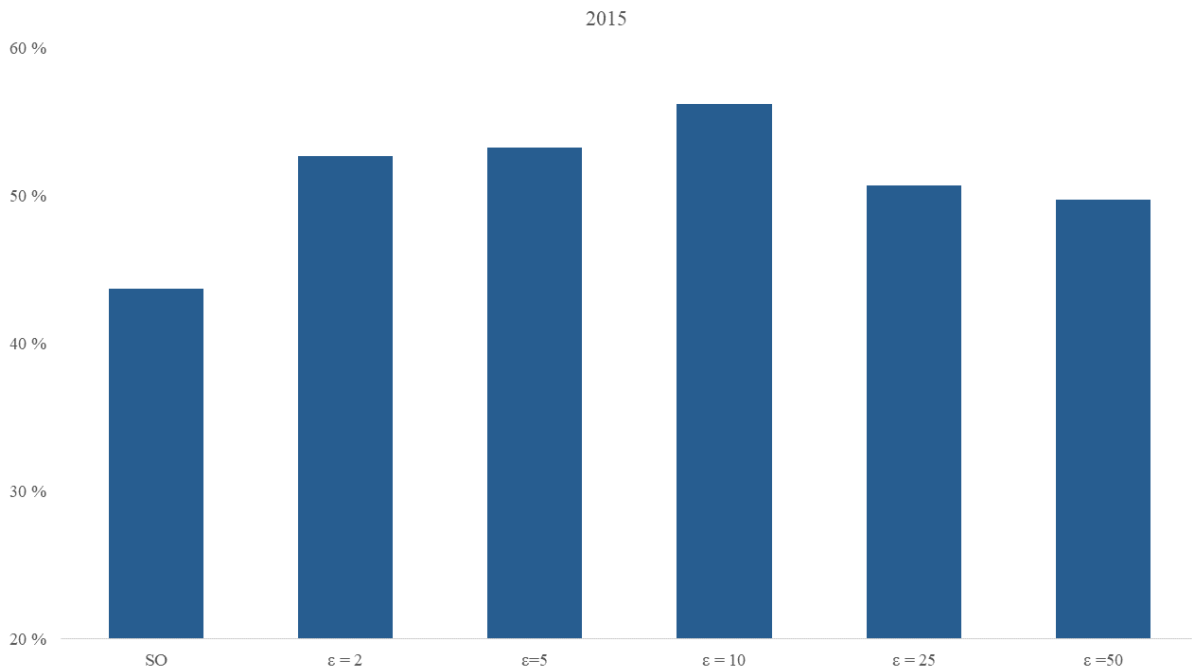


Figure 7.14: Average hedge ratios 2015 for the stochastic- and distributionally robust hedging model with varying ambiguity radius.

As we consider the hedging strategies for 2015, the random realizations of the spot price and production quantities are drawn from the EMPS model generated at the end of December 2014. Further, 25.000 simulations are conducted in order to achieve a significant indication of the performance.

Simulations from the baseline probability distribution We first conduct the simulations by drawing realizations from the baseline probability model, where all scenarios in the EMPS model are equally weighted, as illustrated in Figure 7.19. These simulations evaluate the performance of the models in the instances when the empirically observed baseline probability model actually is realized.

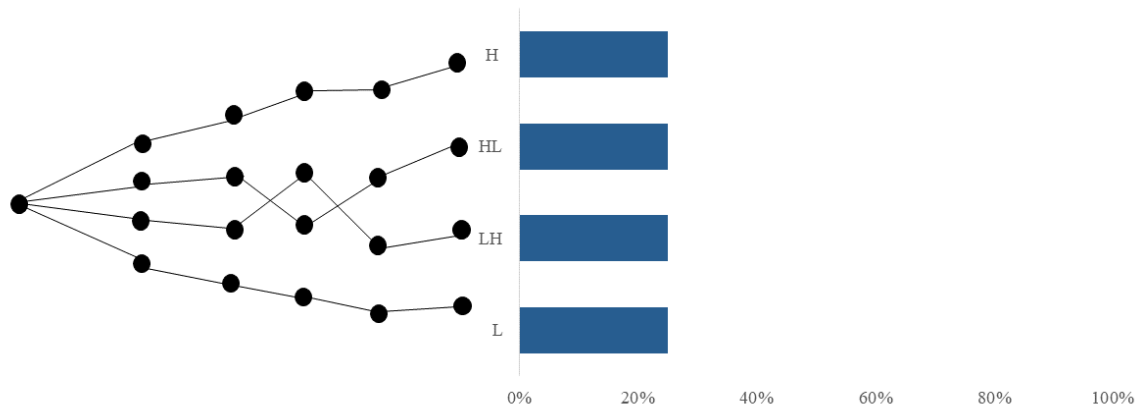


Figure 7.15: Illustration figure: The scenarios from the EMPS model and their corresponding probabilities under the assumed baseline probability distribution. All the scenarios have equal probabilities of being drawn in the simulation.

In analyzing the performance of the hedging models, we calculate the average profit and risk over the 25.000 simulations, where risk is measured by the standard deviation, 90%VaR and 90%CVaR. In Figure 7.16, the outcome of the simulations is illustrated. The expected profits are highest for the naive strategy, followed by the stochastic and distributionally robust model, respectively. Further, among the considered ambiguity radii in the distributionally robust model, profits tend to decrease for $\varepsilon = 2$ to $\varepsilon = 10$, before slightly increasing for higher degrees of ambiguity aversion. The latter is related to lower hedged quantities for the distributionally robust model with high ambiguity radius, according to the discussion in Section 7.2.2.

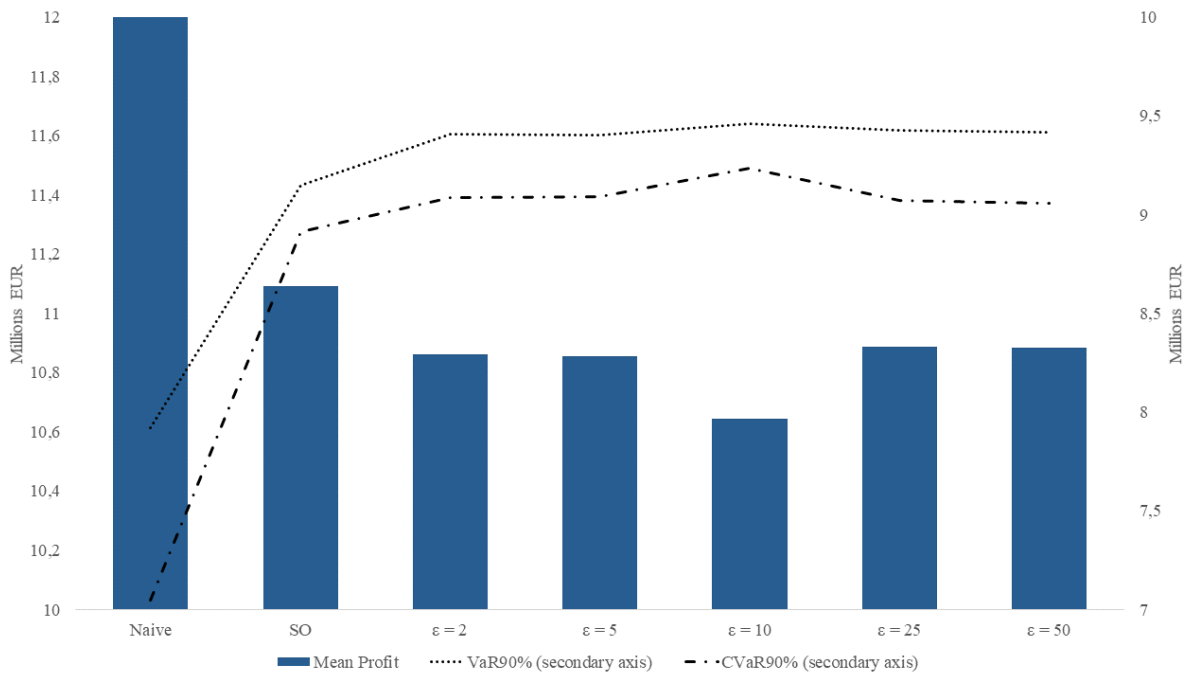


Figure 7.16: Mean profits and tail risk under the simulations from the baseline probability distribution. The performance is shown for the naive-, stochastic- and distributionally robust hedging model

Higher expected profits, however, come at the cost of higher risk, which is in line with traditional financial theory, see e.g. Zenios and Markowitz (2008). Both the expected profits and the risk of the different models are clearly correlated with the average hedge ratios for the different degrees of ambiguity, as presented in Section 7.2.2.

Specifically from Figure 7.16, the expected profits are approximately 12 million EUR for the naive strategy, 11 million EUR for the stochastic strategy and 10.6 million EUR for the distributionally robust strategy with $\varepsilon = 10$. On the other hand, the 90%CVaR is 7, 8.9 and 9.2 million EUR respectively, indicating a significant reduction in tail risk for strategies with higher hedge ratios. The risk-reward trade-off for the different models is further illustrated in Figure 7.17, where the deviation of expected profits and the risk measures are calculated relative to the naive strategy.

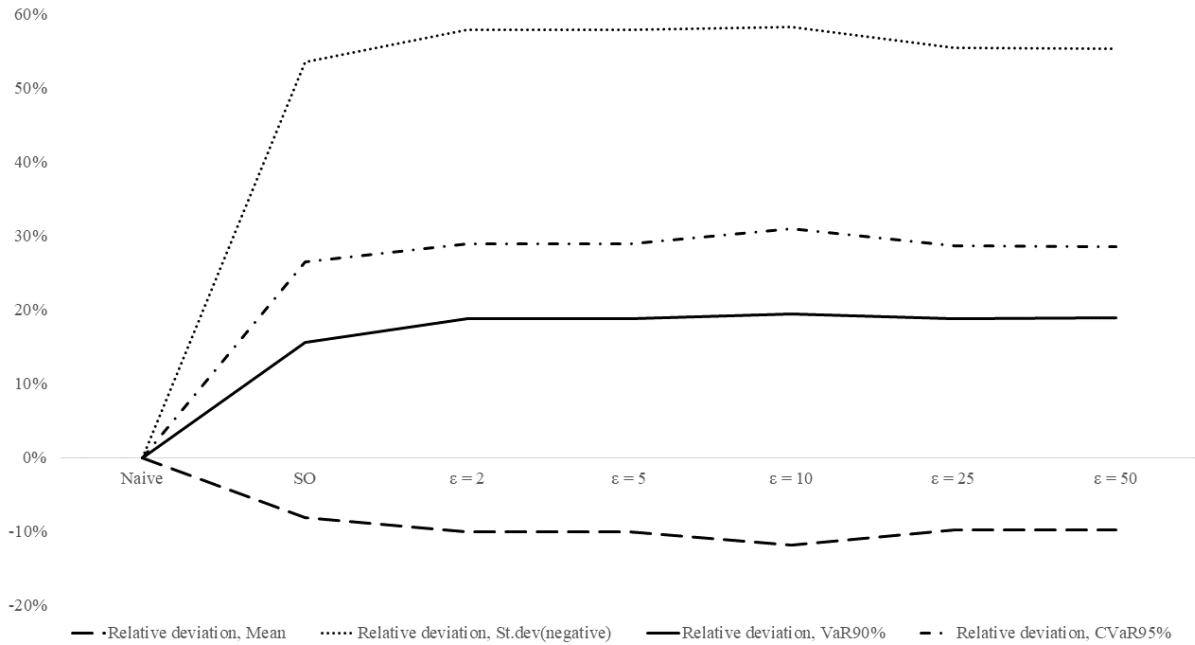


Figure 7.17: Deviations in expected profits and tail risk under the simulations from the baseline probability distribution. Shown for the stochastic- and distributionally robust hedging model, relative to the naive strategy

As can be seen from the figure, the decision maker is able to reduce the tail risk with 26% by trading away 8% of expected revenue in the stochastic model. Also notice that the volatility in the profits, measured by the standard deviation, is reduced by more than 50% when comparing the stochastic model against the naive strategy. Further, by comparing the stochastic model with the distributionally robust model with $\varepsilon = 10$, an additional tail risk reduction of 5 percentage points is achieved by reducing the expected profits with an additional 3 percentage points. Hence, this illustrates trade-off between risk and expected profits achieved by a distributionally robust hedging model, in the situation where the probability distribution is correctly assumed.

Under the assumption that there is no distributional uncertainty, the optimal strategy is the one maximizing the decision marker's utility under the simulations from the baseline distribution. The utility is calculated similarly as in the objective function (6.12) in Section 6.2.2, as

$$(1 - \lambda) \cdot \text{expected profits} + \lambda \cdot 90\%CVaR$$

where the risk aversion λ is set to 0.5. Based on the simulations, the achieved utility for the naive strategy, and the stochastic- and distributionally robust model, are shown in Figure 7.18. The utility is maximized for the stochastic model. This is as expected, since the stochastic model is optimized under the assumption that the baseline probability model is perfectly known. We also observe that the simulated utility for both the stochastic and the distributionally robust model are higher than for the naive strategy. We lastly note that the strategies from the distributionally robust model turns out to only be slightly worse performing than the stochastic model, in terms of utility under the estimated baseline distribution.

By subtracting the utility of the distributionally robust model from the utility of the stochastic model, we get an indication on how suboptimal the distributionally robust model are under the assumed baseline distribution. That is, the reduction in utility when accounting for distributional uncertainty, if the probability distribution is correctly estimated. Hence, through the simulations, we achieve a measure of the price of ambiguity, similar to to what was calculated in Section 7.2.3. The simulated price of ambiguity is illustrated in Figure 7.18 as the bars, measured on the right vertical axis. For ambiguity radii less than $\varepsilon = 25$, the price of ambiguity increases with higher values of ε , since the deviation between the alternative distribution and the estimated baseline distribution in the optimization increases. This is in line with what one would expect. For very high degrees of ambiguity aversion, however, the price is reduced as hedge ratios closer to the stochastic model is recommended, again because of the extraneously low production expectations that are accounted for. For $\varepsilon = 10$, the price of ambiguity is 0.6%, while for $\varepsilon = 2$ and $\varepsilon = 5$ it is just above 0.2%.

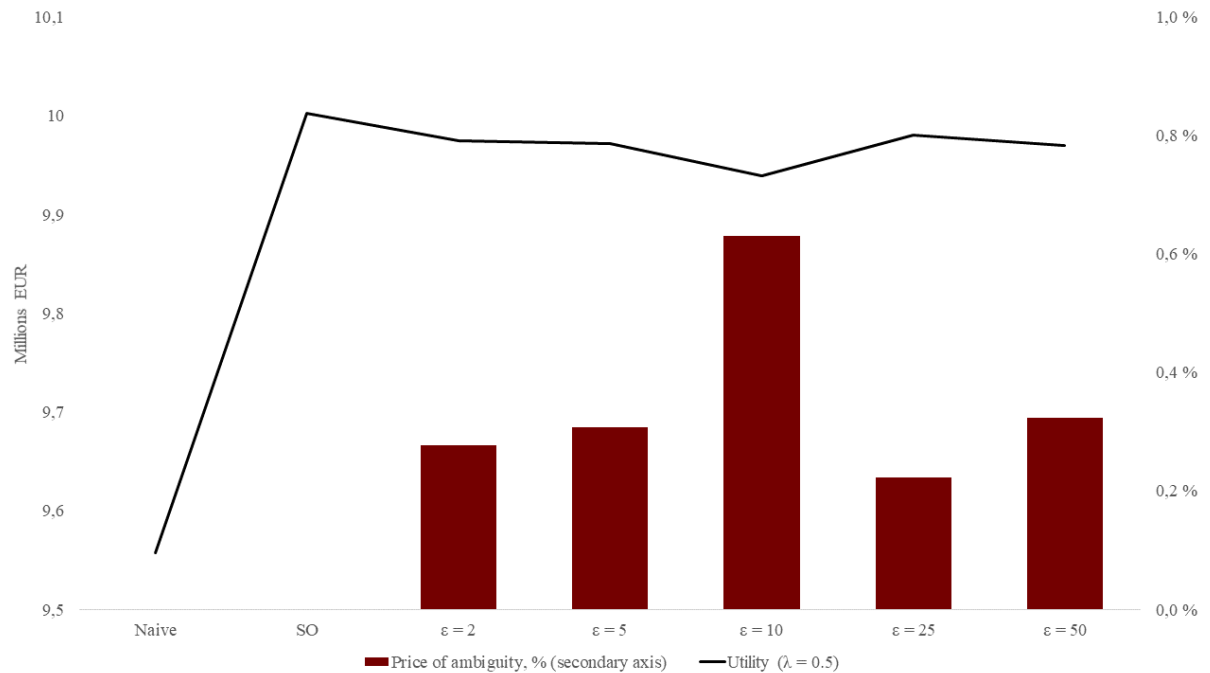


Figure 7.18: Utilities for the naive-, stochastic- and distributionally robust hedging strategies under the assumed baseline distribution, measured on the primary axis. The price of ambiguity for the distributionally robust strategies, measured on the secondary axis.

In cases where there is distributional uncertainty, but the distribution is correctly estimated, the performance of the stochastic hedging model gives, as expected, the highest utility. The performance of the distributionally robust model are, however, only slightly worse performing. It is further interesting to investigate the performance and reduction in utility of the respective hedging models in situations where the baseline distribution does not turn out to be correctly estimated.

Simulations under alternative probability distribution We now analyze the performance of the stochastic and distributionally robust hedging model for probability models that deviate from the one that is empirically observed. This is done by conducting simulations where the scenario probabilities from the EMPS model are changed. Specifically, as we are interested in potential distributions that can cause more harm for the decision maker, we first simulate from a distribution where the low-profit scenarios are given a 50% higher probability. The new distribution is illustrated in Figure 7.19. The scenario values are still the same as in the baseline probability model, which is in line

with how we model the distributional uncertainty in the distributionally robust hedging model. In Appendix B.3, we show the results from a simulation where solely the lower 25% percentile of the revenue scenarios are weighted with higher probabilities. We hereafter refer to this distribution as the *alternative distribution*.

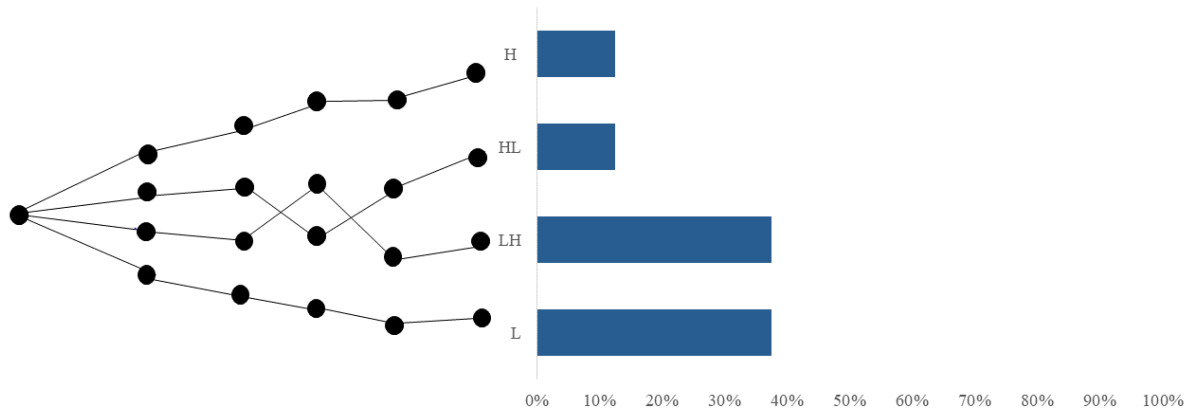


Figure 7.19: Illustration figure: The potential scenarios from the EMPS model and their corresponding probabilities under the alternative probability distribution. Compared to the assumed distributions, the downside revenue scenarios are weighted with higher probabilities.

In Figure 7.20, the expected profits and risk related to the respective hedging strategies is shown. Relative to the simulations from the baseline distribution, a significant drop in expected profits is observed for the naive strategy and the stochastic strategy. Specifically, for the stochastic model the expected profits drops approximately 14%, from about 12 millions EUR under the baseline distribution, to 10.3 millions EUR under the alternative distribution. On the other hand, for the distributionally robust model with $\varepsilon = 10$, which now gives the highest expected profits, the simulated expected profits have dropped less than 1%, from 10.6 millions EUR to 10.5 millions EUR. We further observe that this makes the expected profits about 3% higher for the distributionally robust model than for the stochastic model.

This emphasizes that a problem with stochastic optimization models is that the performance tends to be highly sensitive to badly estimated distributional parameters, as is widely recognized in the literature (Conigli et al., 2016). This proves that such models are significantly vulnerable under distributional uncertainty. Further, it gives rise to what we referred to as the Optimizer’s curse in Section 3.2, where stochastic models tend to

be overfitted on in-sample data, causing poor performance on out-of-sample tests. The distributionally robust hedging model, however, yields more stable and less risky profits under distributional uncertainty. These findings are in line with Esfahani and Kuhn (2015), who find the post decision disappointment to be higher for stochastic models than for distributionally robust model.

From Figure 7.20, we further observe that 90%VaR and 90%CVaR is increasing for ambiguity radii up to $\varepsilon = 10$, before being slightly reduced for higher ambiguity radii. Recall that higher values of VaR and CVaR is equivalent with lower risk. Therefore, as opposed to the simulations from the baseline probability model, we experience lower risk for models with higher expected profits. The distributionally robust optimization model with ambiguity radius $\varepsilon = 10$ is the most attractive under simulations from the alternative probability distribution, since it is the model providing the lower risk and highest expected profits. We further observe that the 90%CVaR is about 5% lower for the distributionally robust model with $\varepsilon = 10$, than for the stochastic optimization model.

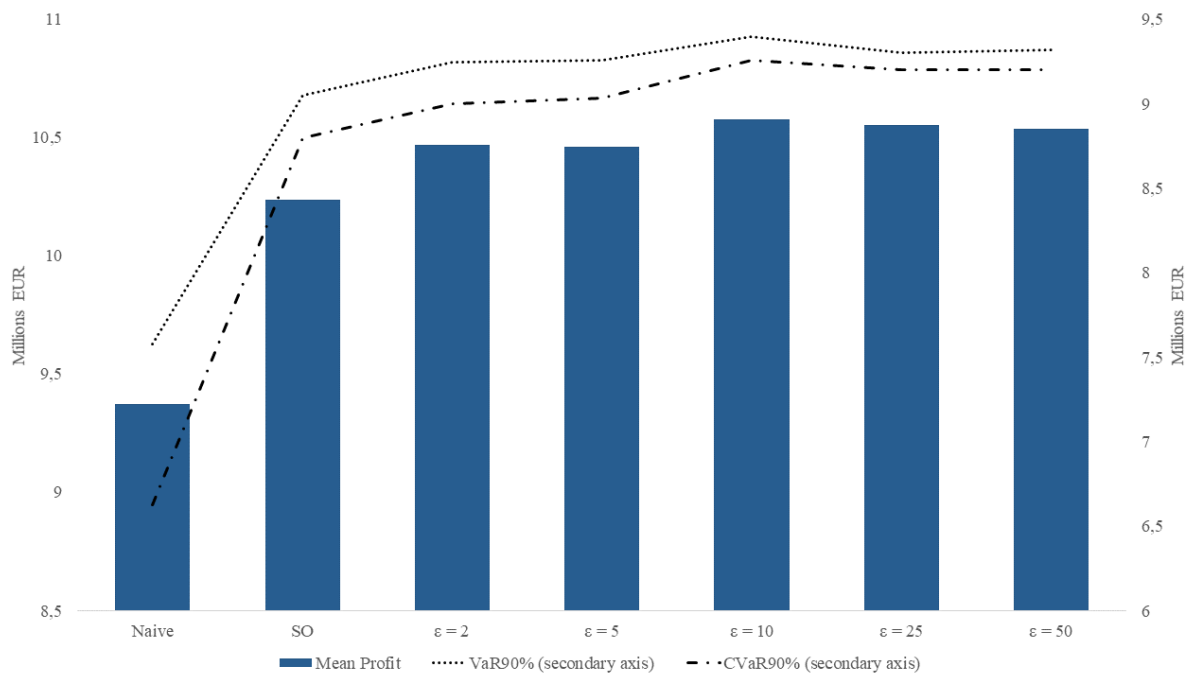


Figure 7.20: Mean profits and tail risk under the simulations from the alternative probability distribution. The performance is shown for the naive-, stochastic- and distributionally robust hedging model

Figure 7.21 shows the deviations in the risk measures and expected profits for the respective models, relative to the naive strategy, for the simulations conducted under the alternative distribution. This figure emphasizes the findings discussed related to Figure 7.20.

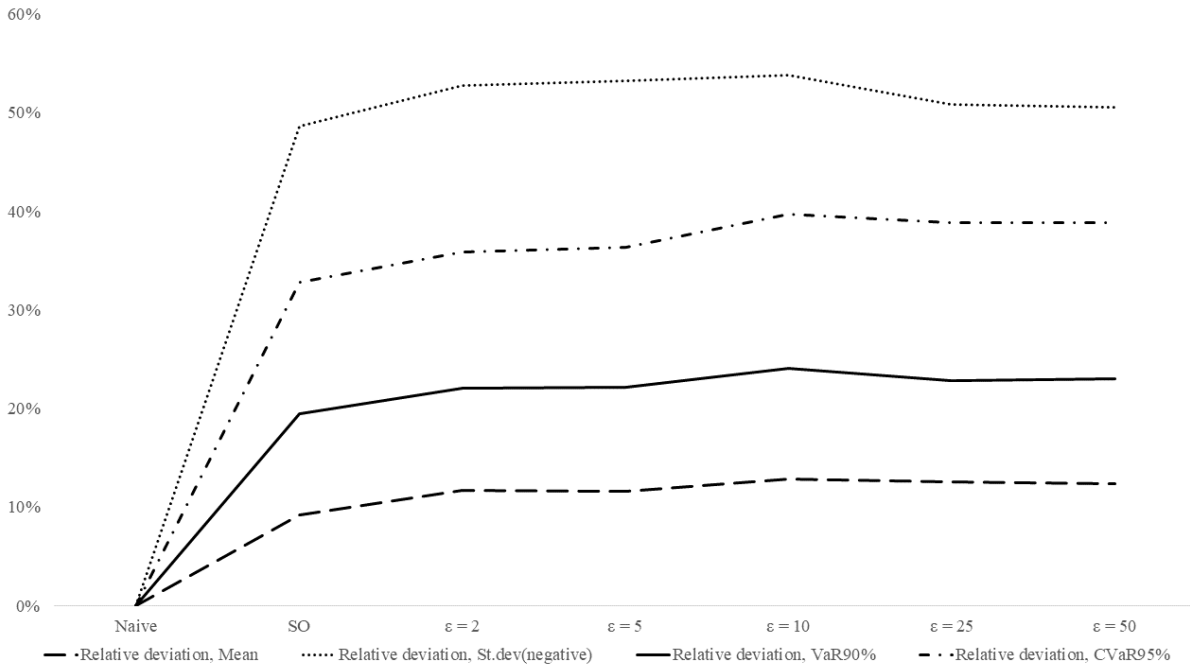


Figure 7.21: Deviations in expected profits and tail risk under the simulations from the alternative probability distribution. Shown for the stochastic- and distributionally robust hedging model, relative to the naive strategy

Further, Figure 7.22 gives the profit density distributions from the 25.000 simulations for three different strategies: The naive-, the stochastic- and the distributionally robust strategy with $\epsilon = 10$. The figure shows expected profits 90%VaR and 90%CVaR. By observing the leftmost figure against the center and right we get a clear indication of reduced risk and increased expected profits by hedging with a stochastic and distributionally robust model. As expected, we also observe that the potential profits are highest under the naive strategy.

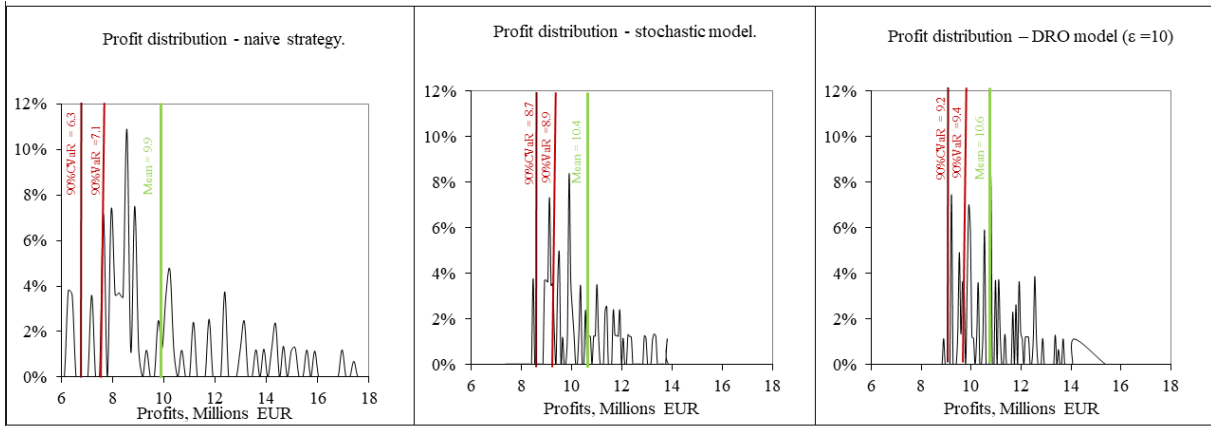


Figure 7.22: The profit density functions of different hedging strategies under the alternative probability distribution

Similarly as for the simulations from the estimated baseline probability model, we calculate the utility of the stochastic and distributionally robust optimization models with varying degrees of ambiguity, but now based on the simulations from the alternative probability model. The risk aversion λ is still 0.5 for both optimization models. The results are shown in Figure 7.23. While the stochastic hedging model gave the highest utility under the simulations from the baseline distribution, the distributionally robust model with $\varepsilon = 10$ is the best in terms of utility under the simulations from the alternative distribution. Further, we observe that all the distributionally robust model give higher utility than the stochastic model.

In the same manner as we calculated the price of ambiguity for simulations from the baseline probability model, simulating from the alternative probability model allows us to calculate the gain for distributional robustness. This is achieved by subtracting the utility of the stochastic hedging model from the respective distributionally robust model with varying ambiguity radii, and is illustrated by the bars in Figure 7.23. This gives an indication on the relative gain of applying a distributionally robust hedging model in a situation where the empirically observed probability distribution is over-optimistically estimated.

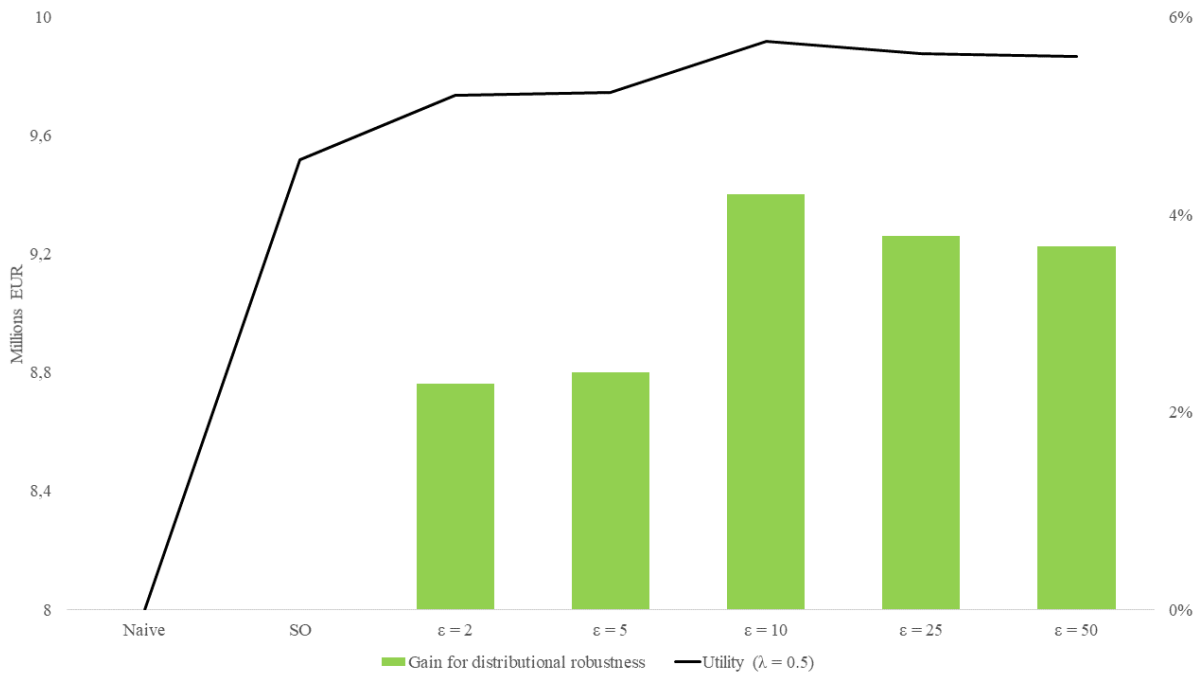


Figure 7.23: Utilities for the naive-, stochastic- and distributionally robust hedging strategies under the alternative distribution, measured on the primary axis. The gain for distributional robustness is measured on the secondary axis.

We find that the price of ambiguity is lower than the gain for distributional robustness. Specifically, the gain for distributional robustness when $\varepsilon = 10$ is above 4%, while the price of ambiguity for the same model is just above 0.6%. The results from the simulations also correspond well with the expected price of ambiguity and gain for distributionally robustness that we calculated in Section 7.2.3.

Final remarks on the simulations The results from the conducted baseline- and alternative probability model simulations show that distributionally robust hedging model is highly attractive for electricity producers when exposed to distributional uncertainty. Based on the simulations, the distributionally robust model tend to be slightly suboptimal in situations where the probability model is correctly estimated, while being highly advantageous when the probability model is wrongly estimated.

Compared to the stochastic model, the distributionally robust model therefore give more stable profits, independent on the underlying probability model. Hence, the model risk

in situations with ambiguity is reduced. In turn, this reduces the post decision disappointment of decisions from distributionally robust model, which is highly favorable for electricity producers that seek a hedging strategy that ensures more stable profits.

We lastly emphasize the benefits of carefully specifying the ambiguity radius of the distributionally robust model. From the simulations we observe that a too low ambiguity radius tend to give suboptimal solutions as the true distribution not is captured by the ambiguity set. A too high ambiguity radius tends to give supoptimal solutions, as overly conservative assumptions regarding the production quantities and prices are made.

The simulation from the distribution where the lower 25% percentile of the revenue scenarios are weighted with higher probabilities, as attached in Appendix B.3, give similar results as the simulation from the alternative distribution presented in this subsection.

With an understanding of the benefits of using distributionally robust optimization to support hedging decisions for an electricity producer, we proceed towards backtesting the hedging strategies from 2014-2017 on real historical data. This enables us to show how TrønderEnergi would have benefitted from accounting for distributional uncertainty over the considered time horizon.

Backtesting

Backtesting involves testing both the naive-, stochastic- and distributionally robust hedging strategies on historical spot prices and production quantities over the considered time horizon. The purpose of doing this is to show that TrønderEnergi would have benefitted from applying a distributionally robust hedging model.

Figure 7.24 shows the yearly average profits of the respective hedging strategies, for the period 2014-2017. The absolute profits are shown by the bars, while the line gives the deviation of the respective strategies, relative to the naive strategy. Recall that we only consider profits for the first five months of each year, as this is the planning horizon of the hedging strategies.

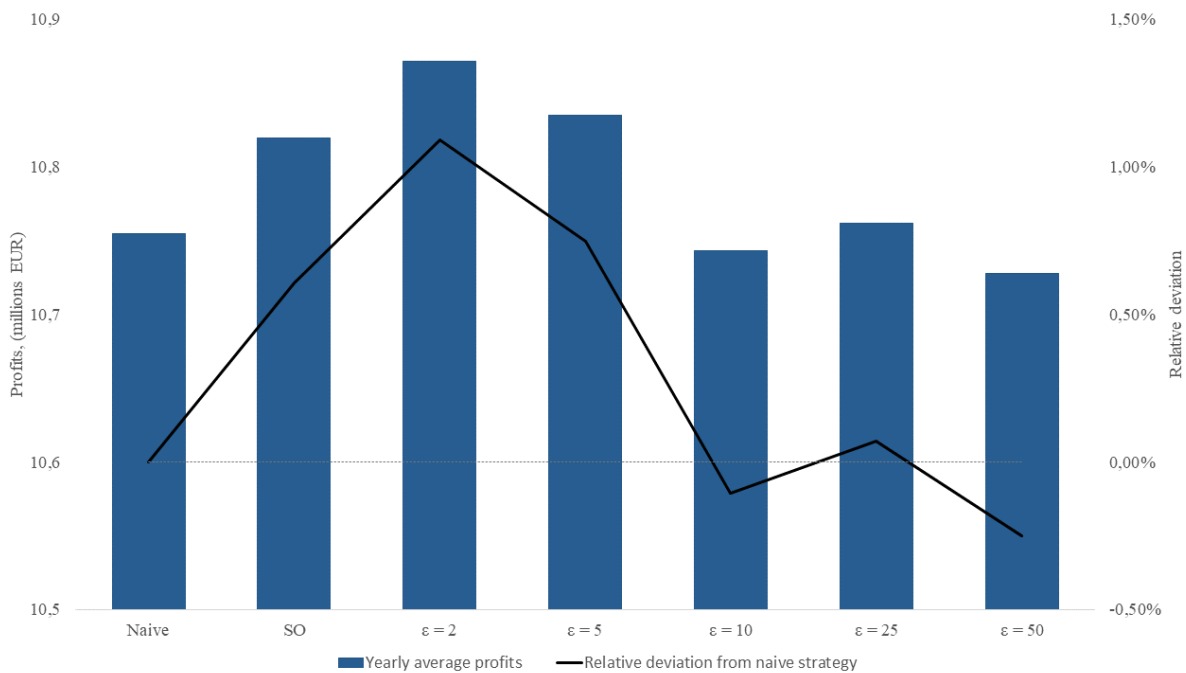


Figure 7.24: Average yearly profits from backtesting the naive-, stochastic- and distributionally robust hedging strategies for 2014-2017 on historical data. The relative deviation in profits from the naive strategy is shown on the secondary axis.

Over the considered time horizon, the distributionally robust hedging model with $\varepsilon = 2$ shows the highest profits, with an average close to 10.9 million EUR. Further, this is about 0.1 million EUR, or 1%, above the yearly average profits from the naive strategy. The distributionally robust model with $\varepsilon = 10$ and $\varepsilon = 50$ give marginally lower profits than the naive strategy. In addition to the distributionally robust model with $\varepsilon = 2$, the hedging model with $\varepsilon = 5$ also shows higher profits than the strategy from the stochastic optimization model.

These findings show that TrønderEnergi could have increased their profits with up to 1% in the period 2014-2017 by applying a stochastic or distributionally robust hedging strategy. In addition, accounting for low degrees of ambiguity in the hedging models turned out to be more profitable than applying a stochastic model. This indicates that the spot prices and production quantities were exposed to some degree of ambiguity.

We further analyze the volatility in the yearly profits from the respective hedging models. In Figure 7.25, the standard deviation of yearly profits from 2014-2017 is shown together with the yearly average profits.

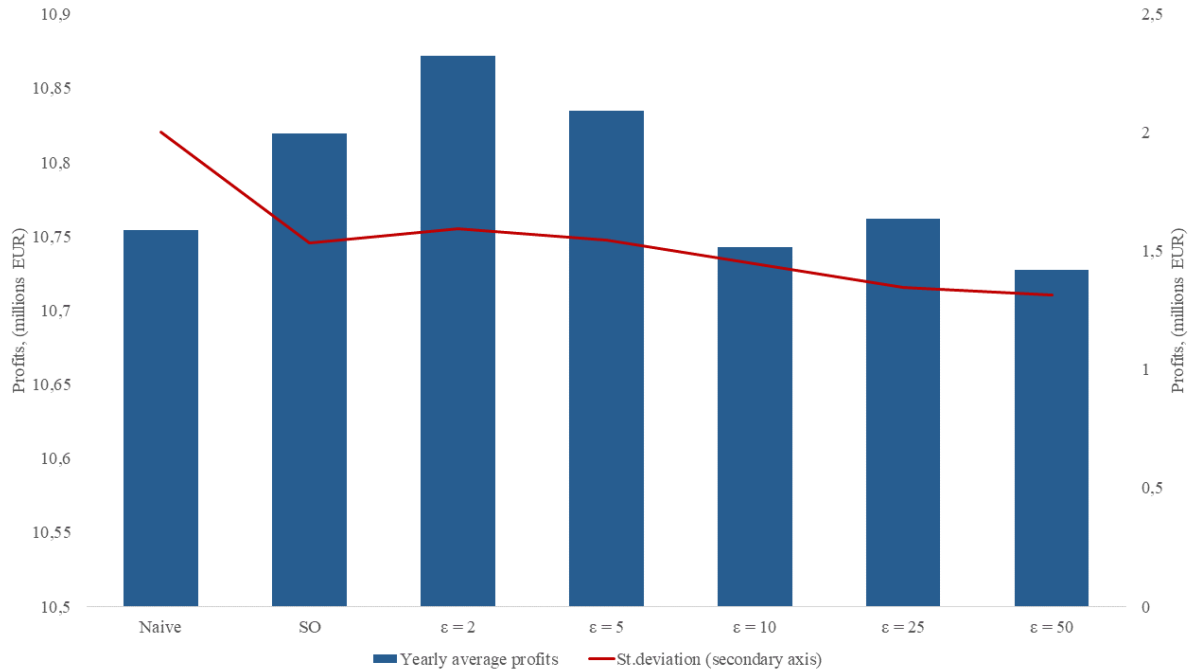


Figure 7.25: The standard deviation of the average yearly profits from backtesting the respective hedging models, measured on the secondary axis. Average yearly profits are shown for the hedging models are shown on the primary axis.

Compared to the naive strategy, there is a decreasing tendency in the volatility of profits for higher ambiguity radii. This is expected as we recall from Section 7.2.2 that the general hedging activity tends to increase with higher ambiguity radii. For the stochastic model and the distributionally robust model with $\varepsilon = 2$ and $\varepsilon = 5$ however, the volatility over the four years is rather stable. Since we observed higher profits for these ambiguity radii, this indicates an improved risk-reward trade-off for decisions of the distributionally robust model.

The findings above are emphasized in Figure 7.26, which shows the yearly profits for the naive-, stochastic- and distributionally robust strategy with $\varepsilon = 5$ and $\varepsilon = 50$ over the time horizon, 2014-2017. The figure shows that profits tend to be slightly more stable for the distributionally robust model than for the stochastic and naive strategies.

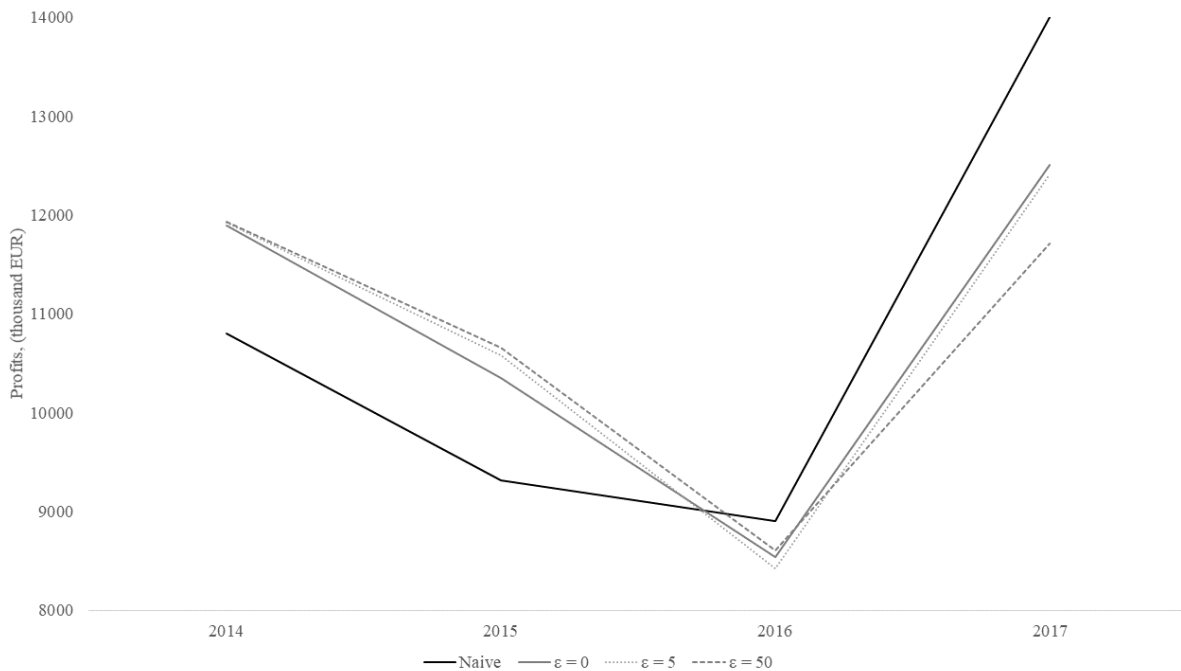


Figure 7.26: Yearly profits for the naive-, stochastic- and distributionally robust strategy with $\epsilon = 5$ and $\epsilon = 50$. Backtested on historical data from 2014-2017

We observe that the stochastic and distributionally robust model gave higher profits than the naive strategy in 2014 and 2015. This is because of the higher hedge ratios in these models, which has proven to be beneficial since spot prices in general were below the achieved forward prices in this period. For 2016 and 2017 the opposite is observed.

As only four years are considered, the 90%CVaR is simply the minimum yearly profit for the respective models over the considered period. As seen from Figure 7.26, these minimum profits occur in 2016 for all of the strategies, where the forward prices in general were significantly lower than through the remaining years. The average spot prices in 2016, however, turned out to exceed the forward price in all considered months, except for February. A perhaps contra-intuitive finding is therefore that the stochastic and distributionally robust hedging models show a lower 90%CVaR than the naive strategy over the studied period. This is likely to change when considering a greater time perspective.

In Appendix B.4, further details on the backtests of the different hedging strategies are found for each of the considered years. Specifically, the realized profits and hedge ratios for the respective models is illustrated on a yearly basis, as well as the difference between

the achieved forward price and the historical spot price for the different months. These figures explain the performance of the hedging models for each year. In addition, the relative deviations between the expected and realized production is shown for each of the considered months. The principal findings are that both the deviation in production quantities and spot prices tend to be highly volatile relative to their expectations, hence emphasizing the risk related to uncertain production quantities and spot prices. This emphasizes the need for an active risk management strategy for electricity producers. However, it also accentuates the risk related to hedging too much of the expected production, as the producer can end up with a production shortage.

We should ideally have backtested for more than four years in order to get a significant indication. However, a larger amount of data for both developing and backtesting the hedging strategies was not available. The clear results from these tests are nevertheless good indications of the effects of accounting for ambiguity when developing hedging strategies.

By backtesting the hedging models on historical data from 2014-2017, higher profits and lower volatility is generally observed for the respective optimization models, relative to the naive strategy. Further, the highest profits are observed for the distributionally robust hedging model with low ambiguity radius. This indicates that accounting for some degree of ambiguity was favourable over the considered time horizon, in terms of profits.

Although the backtesting is performed on too few years to conclude that the findings apply on a significant basis, the findings clearly show the benefits of applying distributionally robust optimization relative to a naive hedging strategy and a stochastic optimization model. With an additional support from the conducted simulations, we conclude that distributionally robust optimization outperforms both a naive strategy and stochastic optimization for risk- and ambiguity-averse electricity producers seeking to reduce the downside risk under distributional uncertainty.

Chapter 8

Concluding Remarks

Following the shortcoming of traditional stochastic optimization models in handling distributional uncertainty, known as ambiguity, we have developed a multistage distributionally robust optimization model for finding optimal hedging decisions for TrønderEnergi, a Norwegian electricity producer. In this thesis, scenario uncertainty as well as ambiguity are considered for both spot prices and production quantities. Forward contracts are applied as hedging instruments. The optimal strategy is found as the one maximizing the utility under a given degree of risk aversion and ambiguity.

The optimal hedging decisions are backtested on historical price and production realizations, and profits are compared to what was realized for TrønderEnergi. The practical perspective of this thesis is therefore a unique contribution to the otherwise highly theoretical literature on distributionally robust optimization.

Ambiguity is managed by optimizing with respect to the probability distribution that inflicts the most harm, chosen from an ambiguity set of possible distributions. The size of the ambiguity set is limited to the probability distributions that differ from the empirically observed distribution up to a certain extent, measured in nested distance. An increased tolerance in nested distance therefore ensures higher robustness towards distributional uncertainty. The method for solving the distributionally robust problem is based on an existing successive convex programming algorithm for approximating the optimal solution. We have extended this algorithm to account for risk aversion by incorporating a time consistent formulation of CVaR as risk measure.

The scenario tree that the optimization model is based on is generated from a clustering approach that seeks to minimize the nested distance to a set of forecast paths. The approach involves generating a random path of spot prices and production volumes based on kernel density estimation, and then identifying the path of nodes in the scenario tree that is closest to the generated path. Ultimately, the values of the identified path are improved with stochastic approximation. Iterations are conducted until convergence in nested distance between the generated tree and the set of forecast paths.

By applying a rolling horizon approach on data from 2014 to 2017, the following findings rise from analyzing the performance of the distributionally robust hedging model and comparing it to the stochastic model:

Accounting for distributional uncertainty in general gives higher hedge ratios, which tend to increase further for larger ambiguity sets. This comes as a consequence of a higher weighting of low-price scenarios as more ambiguity is accounted for in the model. The effect of increasing hedge ratios is, however, moderated as scenarios with lower expected production also tend to be increasingly weighted for higher degrees of ambiguity. In turn, this makes risk management relevant for lower expected production quantities, which for the highest degrees of ambiguity actually gives slightly decreasing hedge ratios.

Our findings show that there is a considerable benefit of applying a distributionally robust model under distributional uncertainty. By simulating from the empirically observed baseline distribution we observe only a slight reduction in the performance of the distributionally robust optimization model, compared to the stochastic model. Under alternative worst-case distributions, however, the distributionally robust model shows a significantly better performance than the stochastic model. This shows that the gain for having made the solution distributionally robust when the realized distribution differs from the one that was estimated, is higher than the price of accounting for ambiguity. Specifically, we find the gain for distributionally robustness and price of ambiguity to be 4% and 0.6% respectively, for the distributionally robust model with ambiguity radius $\varepsilon = 10$. These findings demonstrate that a relatively small proportion of expected returns can be traded away for a significant reduction in risk when applying a distributionally robust hedging model under ambiguity.

The results from backtesting on TrønderEnergi's historical data from 2014 to 2017 show that the hedging strategies of a distributionally robust optimization model that accounts

for small degrees of ambiguity give the highest average profits. Compared to the current situation where hedging is absent, the hedging strategies of the distributionally robust model increases average profits over the period by 1%, while also ensuring more stable profits. The distributionally robust model outperforms a stochastic model with an increase in average profits of 0.5%, while the volatility in profits remains stable. In conclusion, our findings suggest that accounting for ambiguity is favourable when developing hedging strategies for electricity producers.

Chapter 9

Future Research

In this chapter we discuss the potential for further research based on the findings of this thesis. We here cover two principal topics. The first considers venues for improving the distributionally robust optimization solution model. The second topic regards measures for increasing the relevance of the hedging problem for the decision maker.

Our approach towards solving the general distributionally robust optimization problem extends the original approximation approach of Analui and Pflug (2014). There is no literature evaluating the extent to which the original approach deviates from the optimal solution, nor is this covered in this thesis. We do on the other hand test how including the CVaR-inner problem iterations affects the optimality. Tests on a larger dataset would increase the significance of the findings. Studies of the approximation quality of both these approaches would increase their usefulness, and is therefore of interest for future studies.

A considerable impairment of our proposed distributionally robust optimization problem is the considerably low computational efficiency. A venue for further studies is the formulation of computationally tractable multistage distributionally robust problems that incorporate time consistent risk measures.

We consider a fixed scenario tree structure and accounted for distributional uncertainty by only varying scenario probabilities. A desirable field of study is development of computationally tractable algorithms that handle varying tree structures as well as uncertain scenario probabilities.

The principle improvements to the hedging model is to increase the number of scenarios and stages of the scenario tree, as well as to improve the time resolution. This would provide a more accurate representation of the possible uncertain outcomes. It is also relevant to study the performance of the hedging model with speculative considerations. This can be achieved by using actual forward prices, and by adapting the scenario tree so that it considers subjective opinions. Other interesting extensions of the model includes incorporating options as hedging instruments.

Our findings uncovered that studies of the differences between the effects of varying the ambiguity aversion and risk aversion would be valuable contributions to understanding the usefulness of distributionally robust optimization in a hedging context.

The literature on distributionally robust optimization lacks practical applications. Even though this thesis is a contribution, the literature would benefit from increasing the significance of our findings.

Bibliography

- Alexander, C. (2008). *Practical financial econometrics*. John Wiley & Sons, Ltd.
- Analui, B. and Pflug, G. (2014). On distributionally robust multiperiod stochastic optimization. *Computational Management Science*, 11(3):197–220.
- Banks, J., Carson, J., Nelson, B., and Nicol, D. (2010). *Discrete-Event System Simulation*. Prentice Hall, 5 edition.
- Bayraksan, G. and Love, D. (2015). Data-driven stochastic programming using phi-divergences. *Tutorials in Operations Research*, pages 1–19.
- Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). Robust optimization. *Princeton University Press*.
- Ben-Tal, A. and Nemirovski, A. (1997). Robust truss topology design via semidefinite programming. *SIAM J. Optim.*, 7:991–1016.
- Benth, F. E. and Koekebakker, S. (2008). Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157.
- Bertocchi, M., G., C., and Dempster, M. (2011). *Stochastic Optimization Methods in Finance and Energy*. Springer US, New York.
- Bertsimas, D., Gupta, V., and Kallus, N. (2014). Robust sample average approximation. *arXiv:1408.4445*.
- Best, M. and Grauer, R. (1991). On the sensitivity of meanvariance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, 4(2):315–342.

- Bisschop, J. (2006). *AIMMS optimization modeling*. Lulu. com.
- Broll, U., Welzel, P., and Wong, K. P. (2015). Futures hedging with basis risk and expectation dependence. *International Review of Economics*, 62(3):213–221.
- Bystrøm, H. (2003). The hedging performance of electricity futures on the nordic power exchange. *Applied Economics*, 35(1):1–11.
- Carter, D., Rogers, D., and Simkins, B. (2006). Hedging and value in the u.s. airline industry. *Journal of Applied Corporate Finance*, 18(4):21–33.
- Chen, Y.-C. (2017). A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1):161–187.
- Chopra, V. (1993). Mean-variance revisited: Near-optimal portfolios and sensitivity to input variations. *Journal of Investing*, 2(1):51–59.
- Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11.
- Conejo, A. J., Garcia-Bertrand, R., Carrion, M., Caballero, , and de Andrés, A. (2008). Optimal involvement in futures markets of a power producer. *IEEE Transactions on Power Systems*, 23(2):703–711.
- Consigli, G., Kuhn, D., and Brandimarte, P. (2016). Optimal financial decision making under uncertainty. *International Series in Operations Research & Management Science*, 245.
- Delage, E. (2017). Quantitative risk management using robust optimization.
- Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.
- Deng, S. and Oren, S. (2006). Electricity derivatives and risk management. *Energy*, 31(6):940 – 953. Electricity Market Reform and Deregulation.
- Ek, K. and Thorbjørnsen, I. (2014). Optimizing heging strategies for hydropwer producers using forwards.

- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.
- Esfahani, P. M. and Kuhn, D. (2015). Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*.
- Finansdepartementet (2018). *Direkte Skatter*.
URL:<https://www.regjeringen.no/no/tema/okonomi-og-budsjett/skatter-og-avgifter/direkte-skatter/id2353512/>.
- Fleten, S., Wallace, S., and Ziemba, W. (2002). *Hedging electricity portfolios using stochastic programming*.
- Fonseca, R., Zymler, S., Wiesmann, W., and Rustem, B. (2009). Robust optimization of currency portfolios. *COMISEF Working Papers series*, 12.
- Froot, K., Scharfstein, D., and Stein, J. (1993). Risk management: Coordinating corporate investment and financing policies. *The journal of finance*, 48(5):1629–1658.
- Gabriel, S. A., Zhuang, J., and Egging, R. (2009). Solving stochastic complementarity problems in energy market modeling using scenario reduction. *European Journal of Operational Research*, 197(3):1028 – 1040.
- Geman, H. (2008). *Risk Management in Commodity Markets: From Shipping to Agricultural and Energy*. Wiley Finance, WestSussex, England.
- Goh, J. and Sim, M. (2010). Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4):902–917.
- Graham, J. and Rogers, D. (2002). Do firms hedge in response to tax incentives? *The Journal of finance*, 57(2):815–839.
- Gómez-Villalva, E. and Ramos, A. (2004). Risk management and stochastic optimization for industrial consumers. 1.
- Gökgöz, F. and Atmaca, M. E. (2012). Financial optimization in the turkish electricity market: Markowitz’s mean-variance approach. *Renewable and Sustainable Energy Reviews*, 16(1):357 – 368.

- Hanasusanto, G. and Kuhn, D. (2016). Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Arxiv*.
- Heitsch, H. and Römisch, W. (2009). Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118(2):371–406.
- Higle, J. L. (2005). *Stochastic Programming: Optimization When Uncertainty Matters*, chapter Chapter 2, pages 30–53.
- Høyland, K., Kaut, M., and Wallace, S. W. (2003). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2):169–185.
- Jin, Y. and Jorion, P. (2006). Firm value and hedging: Evidence from u.s. oil and gas producers. *The Journal of finance*, 61(2):893–911.
- Jones, M. (1990). Variable kernel density estimates and variable kernel density estimates. *Australian Journal of Statistics*, 32(3):361–371.
- Kaut, M. (2014). A copula-based heuristic for scenario generation. *Computational Management Science*, 11(4):503–516.
- Kaut, M. and Wallace, S. (2003). Evaluation of scenario-generation methods for stochastic programming. 3.
- Keppo, J. (2002). Optimality with hydropower system. *IEEE Power Engineering Review*, 22(6):57–57.
- King, A. and Wallace, S. (2012). *Modeling With Stochastic Programming*. Operations Research and Financial Engineering. Springer.
- Kusy, M. I. and Ziemba, W. T. (1986). A bank asset and liability management model. *Operations Research*, 34(3):356–376.
- Ladurantaye, D. D., Gendreau, M., and Potvin, J.-Y. (2009). Optimizing profits from hydroelectricity production. *Computers & Operations Research*, 36(2):499 – 529. Scheduling for Modern Manufacturing, Logistics, and Supply Chains.
- Lloyd., S. P. (1982). Least square quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137.

- Luenberger, D. (1980). On estimating the expected return on the market: an exploratory investigation. *Oxford University Press*.
- Markowitz, H. (1952). Portfolio selection. *The Journal of finance*, 7(1):77–91.
- McDonald, R. (2014). *Derivatives Markets*. PEARSON.
- Merton, R. (1961). Risk, ambiguity, and the savage axioms. *J. Financ. Econ.*, 8(4):323–361.
- Michaud, R. (2001). Efficient asset management: A practical guide to stock portfolio management and asset allocation. *Oxford University Press*.
- Mo, B., Gjelsvik, A., and Grundt, A. (2001). Integrated risk management of hydro power scheduling and contract management. *IEEE Transactions on Power Systems*, 16(2):216–221.
- Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48(3):261–297.
- NASDAQ (2017a). *Electricity Price Are Differentials*.
URL:<http://www.nasdaqomx.com/transactions/markets/commodities/markets/power/epads>.
- NASDAQ (2017b). *Nordic Power Products*.
URL:<http://www.nasdaqomx.com/commodities/markets/power/nordic-power>.
- Nord-Pool (2017a). *Bidding areas*. URL:<https://www.nordpoolgroup.com/the-power-market/Bidding-areas/>.
- Nord-Pool (2017b). *Nordic and Baltics*.
URL:"<https://www.nordpoolgroup.com/trading/Fees/Nordic-Baltic/>".
- Nord-Pool (2017c). *The power market*. URL:"<https://www.nordpoolgroup.com/the-power-market/>".
- NVE (2016). *Norway and the European power market*. URL:<https://www.nve.no/energy-market-and-regulation/wholesale-market/norway-and-the-european-power-market/>.
- Oum, Y. Pren, S. and Deng, S. (2005). Volumetric hedging in electricity procurement. *Paper presented at the Power Tech*.

- Oum, Y. and Oren, S. S. (2010). Optimal static hedging of volumetric risk in a competitive wholesale electricity market. *Decision Analysis*, 7(1):107–122.
- Papoulis, A. (1965). Probability, random variables, and stochastic processes.
- Perez-Valdes, G., Kaut, M., Fleten, S.-E., and Midthun, K. (2016). Methods for the electricity market - a state of the art. *SINTEF Report*.
- Pflug, G. and Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442.
- Pflug, G. C. and Pichler, A. (2012). A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23.
- Pflug, G. C. and Pichler, A. (2014a). *Introduction*, pages 1–39. Springer International Publishing, Cham.
- Pflug, G. C. and Pichler, A. (2014b). *The Nested Distance*, pages 41–93. Springer International Publishing, Cham.
- Pflug, G. C. and Pichler, A. (2014c). *The Problem of Ambiguity in Stochastic Optimization*, pages 229–255. Springer International Publishing, Cham.
- Pflug, G. C. and Pichler, A. (2015). Dynamic generation of scenario trees. *Computational Optimization and Applications*, 62(3):641–668.
- Pflug, G. C. and Pichler, A. (2016). From empirical observations to tree models for stochastic optimization: Convergence properties. *Society for Industrial and Applied Mathematics*, 26(3):1716–1740.
- Pindyck, R. S. and Rubinfeld, D. L. (2014). Microeconomics.
- Pineda, S. and Conejo, A. (2012). Managing the financial risks of electricity producers using options. *Energy Economics*, 34(6):2216 – 2227.
- Pisciella, P., Vespucci, M., Bertocchi, M., and Zigrino, S. (2016). A time consistent risk averse three-stage stochastic mixed integer optimization model for power generation capacity expansion. *Energy Economics*, 53:203 – 211. Energy Markets.

- Postek, K., den Hertog, D., and Melenberg, B. (2014). Tractable counterparts of distributionally robust constraints on risk. *SIAM Review*.
- Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2(3):21–41.
- Rudloff, B., Street, A., and Valladão, D. M. (2014). Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences. *European Journal of Operational Research*, 234(3):743 – 750.
- Sanda, G., Olsen, E., and Fleten, S. (2013). Selective hedging in hydro-based electricity companies. *Energy Economics*, 40:326–338.
- Schütz, P., Tomasgard, A., and Ahmed, S. (2009). Supply chain design under uncertainty using sample average approximation and dual decomposition. *European Journal of Operational Research*, 199(2):409 – 419.
- Sethi, S. and Sorger, G. (1991). A theory of rolling horizon decision making. *Annals of Operations Research*, 29(1):387–415.
- Shapiro, A. (2003). Inference of statistical bounds for multistage stochastic programming problems. *Mathematical Methods of Operations Research*, 58(1):57–68.
- Shapiro, A. (2009). On a time consistency concept in risk averse multistage stochastic programming. 37:143–147.
- Shapiro, A. (2018). Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming.
- Shapiro, A. and Kleywegt, A. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542.
- Shen, s., Pelsser, A., and Schotman, P. (2013). Robust hedging in incomplete markets. *Working Paper, Maastricht University*.
- Shütz, P. and Westgaard, S. (2018). Optimal hedging strategies for salmon producers.
- Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Routledge., New York.

- SINTEF (2017). *Samkjøringsmodellen*.
URL:<https://www.sintef.no/prosjekter/samkjoeringsmodellen/>.
- Smith, C. and Stulz, R. (1985). The determinants of firms' hedging policies. *The Journal of Financial and Quantitative Analysis*, 20(4):391–405.
- Smith, J. E. and Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322.
- Smithson, C. and Simkins, B. (2005). Does risk management add value? a survey of the evidence. *Journal of Applied Corporate Finance*, 17(3):8–17.
- Stulz, R. (1996). Rethinking risk management. *Journal of Applied Corporate Finance*, 9(3):8–25.
- Séguin, S., Fleten, S.-E., Côté, P., Pichler, A., and Audet, C. (2017). Stochastic short-term hydropower planning with inflow scenario trees. *European Journal of Operational Research*, 259(3):1156 – 1168.
- Tanlapco, E., Lawarree, J., and Liu, C.-C. (2002). Hedging with futures contracts in a deregulated electricity industry. 22:54–54.
- Tanlapco, E., L. J. and Liu, . C. (2002). Hedging with futures contracts in a deregulated electricity industry. *Power Systems, IEEE Transactions*, 17(3):577–582.
- von Neumann, J. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320.
- Žáčková, J. (1966). On minimax solution of stochastic linear programming problems. *Casopis pro Pestovani Matematiky*, 91:423–430.
- Wallace, S. W. and Fleten, S.-E. (2003). Stochastic programming models in energy. In *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 637 – 677. Elsevier.
- Woo, C.-K., Horowitz, I., Horii, B., and Karimov, R. I. (2004). The efficient frontier for spot and forward purchases: an application to electricity. *Journal of the Operational Research Society*, 55(11):1130–1136.

- Woodard, J. D. and Garcia, P. (2008). Basis risk and weather hedging effectiveness. *Agricultural Finance Review*, 68(1):99–117.
- Wu, J. and Sen, S. (2000). A stochastic programming model for currency option hedging. *Annals of Operations Research*, 100(1):227–249.
- Xin, L., Goldberg, D. A., and Shapiro, A. (2013). Distributionally robust multistage inventory models with moment constraints. *arXiv preprint arXiv:1304.3074*.
- Yumi, O., Shmuel, O., and Shijie, D. (2006). Hedging quantity risks with standard power options in a competitive wholesale electricity market. *Naval Research Logistics (NRL)*, 53(7):697–712.
- Zenios, S. and Markowitz, H. (2008). *Practical Financial Optimization: Decision Making for Financial Engineers*. Wiley.

Appendices

Appendix A

Distributionally robust optimization

This appendix contains the supplementary work on distributionally robust optimization referred to in Chapters 3, 5 and 6. Section A.1 gives a broader description than the one found in Chapter 3 on the different approaches to generate the ambiguity set. Section A.2 proves that it is not satisfactory to formulate a convex combination between two probability models within the ambiguity set. Section A.3 presents the results from our implementation of the inventory control problem example in Analui and Pflug (2014). Ultimately, Section A.4 shows why the nested CVaR does not allow an accumulated formulation of the stage-wise utilities.

A.1 Approaches to generate the ambiguity set

Considering the approaches towards handling ambiguity, a shared assumption is that there exists an unknown underlying probability distribution that governs the uncertainty. This distribution is assumed to be somewhere in the proximity of the empirically observed data, and is more specifically assumed to reside within a so-called ambiguity set of possible distributions that might govern the uncertainty. There are no assumptions made about the probability of choosing one probability distribution over another. Rather, the chosen distribution is the one that causes the worst possible expected objective value. There is

no superior way of forming an ambiguity set. The different approaches towards handling ambiguity therefore concern how the ambiguity set is formed. The primary approaches are:

- *Statistical moments.* This approach regards the ambiguity set as intervals around the statistical moments of a distribution. Hence, distributional information has to satisfy a set of statistical moment constraints, typically being not to deviate more than a certain amount from the empirically observed moments. Two primary benefits of this approach is that it has proven to have high computational tractability, in addition to giving the modeller high flexibility in adjusting ambiguity sets, since the moments of a distribution can be customized. Delage and Ye (2010) provide a thorough description of this approach.
- *Goodness-of-fit confidence regions.* The ambiguity set in this approach contains distributions that have passed a statistical hypothesis test relative to the empirical distribution with some level of confidence. This is for instance done by Bertsimas et al. (2014). A benefit of this method is that a statistical guarantee that the underlying probability distribution is within the ambiguity set is obtained. In addition, the method is computational tractable.
- *Probability metrics.* This type of ambiguity set can be considered as a ball around the empirically observed data with radius according to some statistical distance measure, or probability metric. The ambiguity set contains all probability distributions within the ball. Benefits of these ambiguity sets are that the degree of conservatism can be specified by adjusting the radius of the ball, as well as often being able to give statistical guarantees in the same manner as for goodness-of-fit ambiguity sets. Typical statistical distance measures are ϕ -divergences (Bayraksan and Love, 2015), or the nested distance, which is the basis for the distributionally robust models used this paper.

A.2 An incorrect convex combination

Following the example of Pflug and Pichler (2014a), we show why formulating the convex combination between two probability models in the ambiguity set is an approach that does not satisfy the following criterion. The nested distance between a baseline model and the

convex combination has to be no greater than the maximal nested distance between the baseline model and one of the two probability models.

We first consider a baseline model $\mathbb{P}^{(0)}$ and the alternative models $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$. According to Pflug and Pichler (2014a), in order to satisfy the above mentioned criterion, we have to evaluate the nested distance from $\mathbb{P}^{(0)}$ to the probability model formed by the convex combination of leaf node probability distributions $\lambda P^{(1)} + (1 - \lambda)P^{(2)}$, which we denote $\mathbb{P}(\mathbb{T}, P^{(\lambda)})$. This nested distance has to be less than the convex combination $\text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})$ and $\text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(2)})$. More specifically, we have to find an instance where

$$\text{dl}_r(\mathbb{P}(\mathbb{T}, P^{(0)}), \mathbb{P}(\mathbb{T}, P^{(\lambda)}))^r \leq \lambda \text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})^r + (1 - \lambda) \text{dl}_r(\mathbb{P}^{(0)}, \mathbb{P}^{(2)})^r \quad (\text{A.1})$$

does not hold. We now show that such an instance can be found for convex combinations of leaf node probabilities.

Example 1. It suffices to only use the convex combination between two trees in this instance, hence we write $\mathbb{P}^{(2)}$ in (A.1) as $\mathbb{P}^{(1)}$, and write $\mathbb{P}^{(1)}$ in (A.1) as $\mathbb{P}^{(0)}$. We then compare the convex combinations of $\text{dl}_1(\mathbb{P}^{(0)}, \mathbb{P}^{(0)})$ and $\text{dl}_1(\mathbb{P}^{(0)}, \mathbb{P}^{(1)})$ against the nested distance from $\mathbb{P}^{(0)}$ to the probability distribution formed by the convex combination $\mathbb{P}^{(0,5)} = 0,5P^{(0)} + 0,5P^{(1)}$, which we denote $P^{(0,5)} = \mathbb{P}(\mathbb{T}, P^{(0,5)})$. Now consider probability models as scenario trees, as shown in the following Figure A.1.

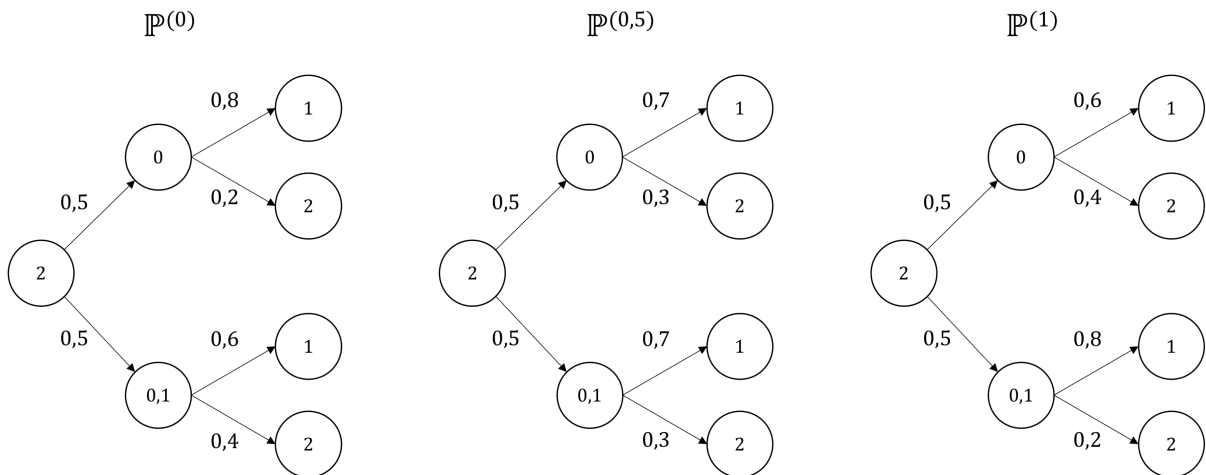


Figure A.1: The arc values are conditional probabilities, the nodes values are the values in the state space.

We can easily see that $\mathbf{dl}_1(\mathbb{P}^{(0)}, \mathbb{P}^{(0)}) = 0$. We find $\mathbf{dl}_1(\mathbb{P}^{(0)}, \mathbb{P}^{(1)}) = 0,1$ thus requiring the left hand side of (A.1) to be less than or equal to $0,5 \cdot 0 + 0,5 \cdot 0,1 = 0,05$. However $\mathbf{dl}_1(\mathbb{P}(\mathbb{T}, P^{(0)}), \mathbb{P}(\mathbb{T}, P^{(0,5)})) = 0.1$, hence (A.1) does not hold.

A.3 Results from the Inventory Control Problem Example

We here present the results from our implementation of the inventory control problem example in Analui and Pflug (2014). We first present the problem and the instance information. We then confirm the diminishing complexity of the scenario probabilities and reduced expected objective function value when the ambiguity radius increases.

The authors present an inventory control problem with the objective of maximizing total expected profits. The nomenclature is defined as follows

Parameters	
pr^b	Selling price for each product $b = 1, 2$
pc^b	Production cost of each product $b = 1, 2$
ic^b	Inventory cost of each product $b = 1, 2$
ec^b	External supply cost of each product $b = 1, 2$
c	Maximum overall production capacity
\bar{x}_i	Maximum inventory capacity
$init_b$	Initial stock level of product $b = 1, 2$
d^b	Demand for product $b = 1, 2$
Decision Variables	
x_f^b	Production volume of product b for $b = 1, 2$
Decision Dependent Variables	
x_i^b	Inventory level of each product $b = 1, 2$
x_e^b	External supply of each product $b = 1, 2$
v	Profit

The model is presented below. For elaboration on the constraints, we refer to Analui and Pflug (2014).

$$\begin{aligned}
& \max \sum_n P(n)v(n) && \forall n \in \mathcal{N} && (5.1) \\
\text{subject to} & \sum_b x_f^b(n_-) \leq c && \forall n \in \mathcal{N} \setminus \mathcal{N}_0 && (a) \\
& x_i^b(n_-) + x_f^b(n_-) + x_e^b(n) - d^b(n) = x_i^b(n) && \forall n \in \mathcal{N} \setminus \mathcal{N}_0 && (b) \\
& \sum_b x_i^b(n) \leq \bar{x}_i && \forall n \in \mathcal{N} && (c) \\
& x_i^b(n_-) + x_e^b(n) \geq d^b(n) && \forall n \in \mathcal{N} \setminus \mathcal{N}_0 && (d) \\
& \sum_b pr^b d^b(n) - \sum_b [pc^b x_f^b(n_-) + ic^b x_i^b(n) + ec^b x_e^b(n)] = v(n) && \forall n \in \mathcal{N} \setminus \mathcal{N}_0 && (e) \\
& x_f^b \geq 0 \\
& x_i^b \geq 0 \\
& x_e^b \geq 0
\end{aligned}$$

We confirm that when the ambiguity radius ε increases, the problem narrows down the scenario tree to focus on a few worst case scenarios.

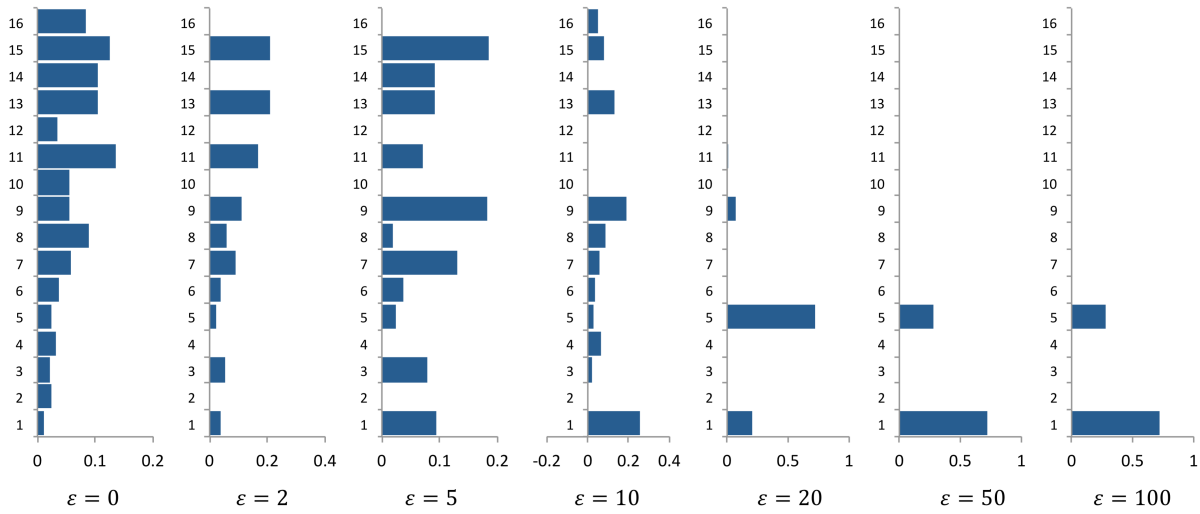


Figure A.2: When the ambiguity radius increases the scenario tree narrows down to focus on only a few scenarios.

We emphasize that these worst case probabilities differ from those presented in Analui and Pflug (2014). An explanation is that we believe the stochastic baseline problem presented in by Analui and Pflug (2014) is incorrect. The authors have implemented the stochastic optimization example in Chapter 17 of Bisschop (2006). However, where we confirm the optimal stochastic model solution of EUR 76 482.406 from this book, this does not correspond with the optimal value claimed by Analui and Pflug (2014) of EUR 7 688. Nor does this value correspond with the optimal initial stage profits.

We further confirm that the increased conservatism of the model decreases the objective function value. The ensured robustness against the worst case probabilities in Figure A.2 comes at a maximal cost of 7,94%.

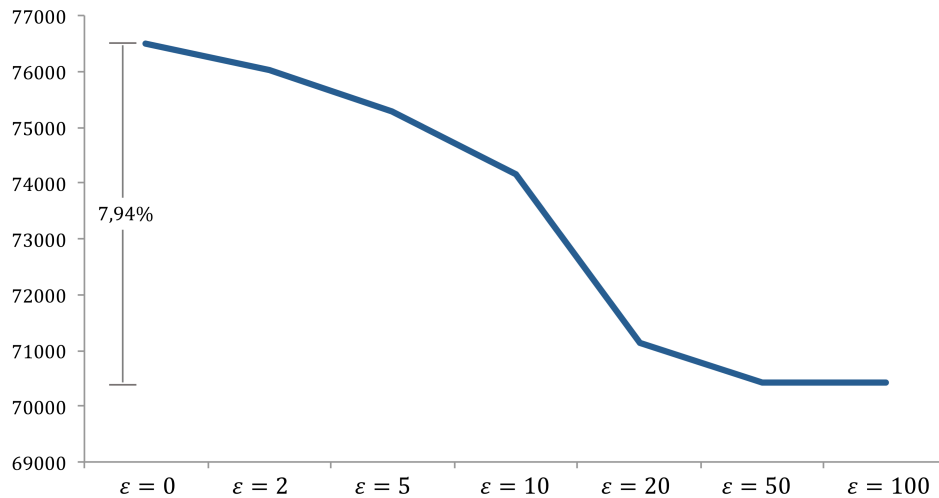


Figure A.3: Decline in objective function value when the ambiguity radius increases.

A.4 An accumulated formulation with Nested CVaR

As discussed in Section 5.3, the inner problem in the original framework of Analui and Pflug (2014) inputs only the optimal objective values at the leaf stage from the outer problem. An intuitive example is the inventory control model presented in Section A.3. Here, the scenario profits are the accumulated profits from sales at every stage in the planning period. In this section, we present an illustrative example of why an accumulated formulation of the stage-wise utilities does not fit well in our instance. This problem follows from the recursive structure we use to make CVaR time consistent. We observe that attempting to accumulate the utility from the root down to the leaf nodes would eventually nest up the constraints to one single constraint, where all objective function variables cancel each other out. In addition to make little sense, the problem then becomes infeasible.

Consider a tree of two stages where we attempt to accumulate the utilities by adding the root node utility to the leaf stage utilities. Exemplified for two scenarios with probability P and $1 - P$ and assuming $\lambda = 0$, we write the leaf node utilities (6.13) as

$$v_{01} = h_{01} + v_{00}$$

$$v_{11} = h_{11} + v_{10}$$

Since we know from (6.12) that the root node utility is written

$$v_{00} = h_{00} + Pv_{01} + (1 - P)v_{11} \tag{A.2}$$

By inserting the leaf node utilities we end up with

$$0 = h_{00} + Ph_{01} + (1 - P)h_{11} \tag{A.3}$$

which requires negative profits, and further makes the problem infeasible.

Appendix B

Numerical results

In this appendix we present the numerical results referred to in Chapter 7. Section B.1 contains the hedge ratios from running the distributionally robust and stochastic hedging models, Section B.2 presents the data of which the gain for distributional robustness and price of ambiguity is calculated, Section B.3 contains the results from simulating from an alternative worst case distribution, and Section B.4 contains complementary analysis for the backtesting.

B.1 Hedging strategies from the optimization models

The hedging strategies achieved from applying a rolling horizon approach on scenario trees from 2014 to 2017 follows in tables B.1, B.2, B.3 and B.4. The tables give the entrance period and maturity periods of the forward contracts recommended from both the stochastic- and distributionally robust hedging models with varying ambiguity radii, ε . The hedging activity is given as hedge-ratios, meaning that the hedged quantum is divided by the expected production for the relevant maturity period. Consider the following example that clearly illustrates how the expected production is obtained: As contracts maturing in February can be entered both in December and January, the expected production for February is found as the average of the expectations obtained from the scenario tree generated in December and January. Recall that a planning horizon in

the models is 4 months, and that the rolling horizon approach is applied over a window of 5 months.

			Forwards entered						
		ϵ	Des	Jan	Feb	Mar	Apr	Hedge-ratio	
Maturity	Jan	SO	23 %					23 %	
		2	23 %					23 %	
		5	19 %					19 %	
		10	24 %					24 %	
		25	20 %					20 %	
		50	-					0 %	
	Feb	SO	-	2 %					2 %
		2	-	9 %					9 %
		5	-	9 %					9 %
		10	-	13 %					13 %
		25	-	13 %					13 %
		50	-	-					0 %
	Mar	SO	-	16 %	62 %				78 %
		2	-	30 %	57 %				87 %
		5	-	-	93 %				93 %
		10	-	-	98 %				98 %
		25	-	-	98 %				98 %
		50	-	-	98 %				98 %
	Apr	SO	-	-	76 %	-			76 %
		2	-	-	82 %	-			82 %
		5	-	-	77 %	-			77 %
		10	-	69 %	1 %	-			70 %
		25	-	69 %	-	-			69 %
		50	-	93 %	-	-			93 %
	May	SO		-	-	-	32 %		32 %
		2		-	-	-	32 %		32 %
		5		-	-	-	39 %		39 %
		10		-	-	-	50 %		50 %
		25		-	-	-	51 %		51 %
		50		-	-	-	49 %		49 %

Table B.1: 2014: Detailed hedging strategies from the stochastic- and distributionally robust hedging models.

		Forwards entered							
	ϵ	Des	Jan	Feb	Mar	Apr	Hedge-ratio		
Maturity	Jan	SO	61 %					61 %	
		2	86 %					86 %	
		5	86 %					86 %	
		10	86 %					86 %	
		25	86 %					86 %	
		50	84 %					84 %	
	Feb	SO	35 %	-					35 %
		2	36 %	-					36 %
		5	36 %	-					36 %
		10	13 %	37 %					50 %
		25	13 %	44 %					57 %
		50	-	57 %					57 %
	Mar	SO	-	45 %	-				45 %
		2	-	46 %	-				46 %
		5	-	48 %	-				48 %
		10	-	45 %	-				45 %
		25	2 %	43 %	-				45 %
		50	-	43 %	-				43 %
	Apr	SO	-	-	35 %	-			35 %
		2	-	-	36 %	-			36 %
		5	-	-	38 %	-			38 %
		10	-	6 %	46 %	-			52 %
		25	-	-	46 %	-			46 %
		50	-	-	46 %	-			46 %
	May	SO		-	-	35 %	-		35 %
		2		-	-	43 %	-		43 %
		5		-	-	43 %	-		43 %
		10		-	-	35 %	-		35 %
		25		-	-	5 %	-		5 %
		50		-	-	4 %	-		4 %

Table B.2: 2015: Detailed hedging strategies from the stochastic- and distributionally robust hedging models.

		Forwards entered					Hedge-ratio		
	€	Des	Jan	Feb	Mar	Apr			
Maturity	Jan	SO	9 %					9 %	
		2	8 %					8 %	
		5	6 %					6 %	
		10	3 %					3 %	
		25	-					0 %	
		50	-					0 %	
	Feb	SO	36 %	-					36 %
		2	36 %	-					36 %
		5	38 %	-					38 %
		10	40 %	-					40 %
		25	40 %	-					40 %
		50	40 %	-					40 %
	Mar	SO	-	17 %	8 %				25 %
		2	-	17 %	8 %				25 %
		5	-	10 %	15 %				25 %
		10	-	-	24 %				24 %
		25	-	14 %	20 %				34 %
		50	-	12 %	21 %				34 %
	Apr	SO	-	-	51 %	-			51 %
		2	-	-	76 %	-			76 %
		5	-	-	76 %	-			76 %
		10	-	-	78 %	-			78 %
		25	-	-	76 %	-			76 %
		50	-	-	76 %	-			76 %
	May	SO		-	-	-	29 %		29 %
2			-	-	-	30 %		30 %	
5			-	-	-	29 %		29 %	
10			-	-	-	34 %		34 %	
25			-	-	-	-		0 %	
50			-	-	-	-		0 %	

Table B.3: 2016: Detailed hedging strategies from the stochastic- and distributionally robust hedging models.

		Forwards entered							
		€	Des	Jan	Feb	Mar	Apr	Hedge-ratio	
Maturity	Jan	SO	13 %					13 %	
		2	23 %					23 %	
		5	21 %					21 %	
		10	16 %					16 %	
		25	28 %					28 %	
		50	3 %					3 %	
	Feb	SO	24 %	14 %					38 %
		2	21 %	22 %					43 %
		5	27 %	15 %					42 %
		10	9 %	32 %					41 %
		25	-	44 %					44 %
		50	-	44 %					44 %
	Mar	SO	-	8 %	-				8 %
		2	-	9 %	-				9 %
		5	-	10 %	-				10 %
		10	-	9 %	-				9 %
		25	-	12 %	-				12 %
		50	-	12 %	-				12 %
	Apr	SO	-	20 %	-	-			20 %
		2	-	21 %	-	-			21 %
		5	-	19 %	-	-			19 %
		10	-	6 %	-	84 %			90 %
		25	-	20 %	-	70 %			90 %
		50	-	20 %	-	70 %			90 %
	May	SO		-	46 %	37 %	5 %		88 %
		2		-	53 %	29 %	7 %		90 %
		5		-	57 %	26 %	9 %		92 %
		10		-	26 %	57 %	9 %		92 %
		25		-	33 %	50 %	10 %		92 %
		50		-	32 %	51 %	10 %		93 %

Table B.4: 2017: Detailed hedging strategies from the stochastic- and distributionally robust hedging models.

B.2 Gain for distributional robustness and Price of ambiguity

We here present the data of which the gain for distributional robustness and price of ambiguity is calculated. The data are found by running the distributionally robust model for the January data for years 2014 to 2017, for $\varepsilon = 0, 5, 10, 15, 20, 25, 50$. Recall that

$\varepsilon = 0$ equals the stochastic model. The presented values are the average values over the four observations.

It should be noted that we here have simplified the recursive utility function. For a given scenario we have calculating the utility as the sum of all stagewise utilities for that scenario. This is slightly incorrect, since the utility at a stage in a given scenario already contains information about the utility at the succeeding stages. The reason for this adjustment is because it makes the computations easier. The effects of this simplification is that the utilities in the lower stages of the tree are assigned too much weight. We could alternatively have calculated gain for distributional robustness and price of ambiguity based on the profits. We however find this to be inadequate since the risk would not be considered, which indeed is an important component of the objective function.

The detailed results are presented below, where we first show the gain for distributional robustness (GFDR).

ε	DRO worst case utility	SO worst case utility	GFDR
0	29 889	29 889	0 %
5	29 036	28 274	2,6 %
10	28 406	27 163	4,4 %
15	26 516	23 570	11,1 %
20	25 701	22 337	13,1 %
25	25 206	21 558	14,5 %
50	25 140	21 341	15,1 %

Table B.5: Expected utility for the distributionally robust (DRO) and stochastic model (SO) if the realized probability model is the worst case model. This is presented for a set of ε . The difference, gain for distributional robustness (GFDR), is here represented in percent.

Further we present the price of ambiguity (POA).

B.2. GAIN FOR DISTRIBUTIONAL ROBUSTNESS AND PRICE OF AMBIGUITY 181

ε	DRO baseline utility	SO baseline utility	POA
0	29 889	29 889	0,0 %
5	29 145	29 889	2,5 %
10	29 079	29 889	2,7 %
15	27 508	29 889	8,0 %
20	27 131	29 889	9,2 %
25	26 869	29 889	10,1 %
50	26 743	29 889	10,5 %

Table B.6: Expected utility for the distributionally robust (DRO) and stochastic (SO) models if the realized probability model is the baseline model. The price of ambiguity (POA) is the difference between the stochastic and the distributionally robust expected utility, here presented in percent.

B.3 Simulations from alternative probability distribution II

In order to analyze the performance of the stochastic and distributionally robust hedging models under distributions that deviate from the assumption, we conduct simulations where the scenario-probabilities from the EMPS model, as explained in Chapter 7, are changed. As we are interested in potential distributions that can cause more harm for the producer, we here simulate from a distribution where the lower 25% percentile of the revenue scenarios are given twice as high probabilities to occur, see figure B.1. This complements the analysis in Chapter 7, where a simulation was conducted from an alternative worst case probability distribution in order to investigate the models robustness towards distributional uncertainty.

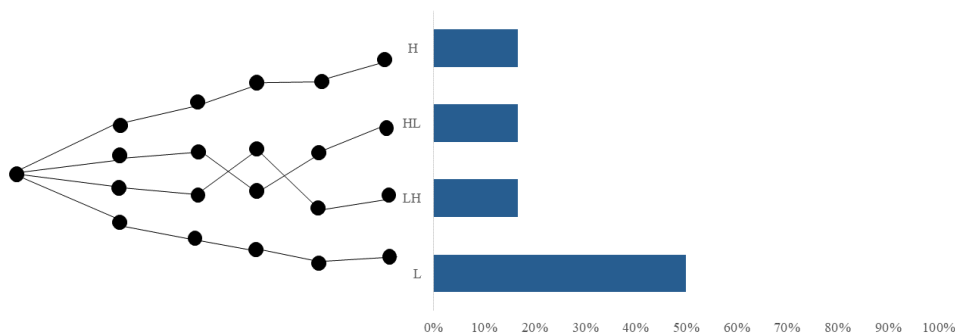


Figure B.1: Illustration figure: The potential scenarios from the EMPS model and their corresponding probabilities under the probability distribution the simulations are conducted from. Compared to the baseline distributions, the downside revenue scenarios are weighted with higher probabilities.

B.3. SIMULATIONS FROM ALTERNATIVE PROBABILITY DISTRIBUTION II 183

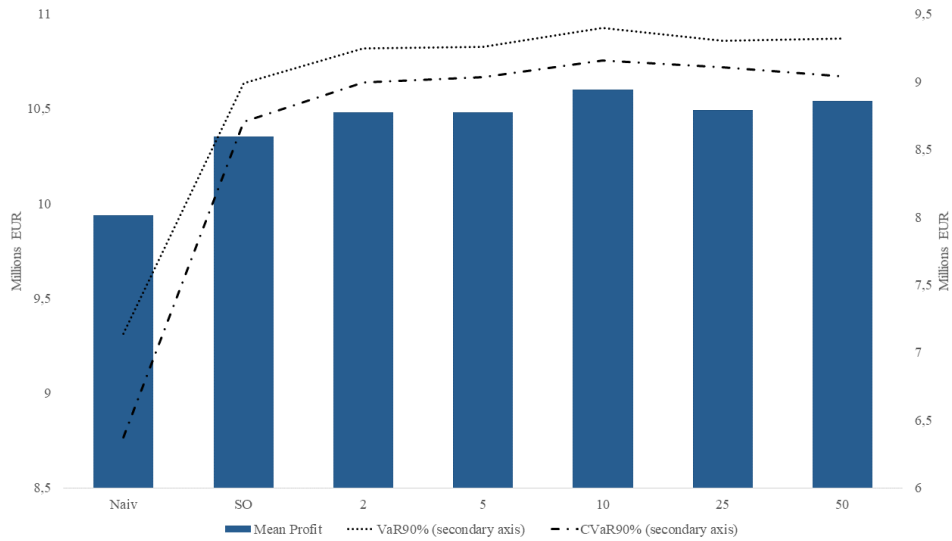


Figure B.2: Mean profits and tail risk under the simulations from the worst case probability distribution. The performance is shown for the naive-, stochastic- and distributionally robust hedging models. We observe that the mean profits and the tail risk is respectively the highest and lowest for the distributionally robust model with $\varepsilon = 10$

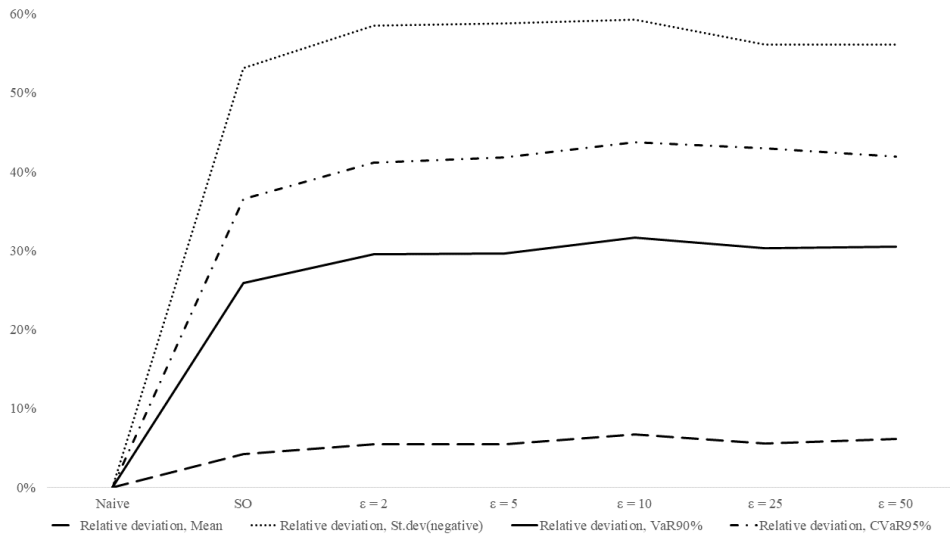


Figure B.3: Deviations in mean profits and tail risk under the simulations from the worst case probability distribution. Illustrated for the stochastic- and distributionally robust hedging models, relative to the naive strategy

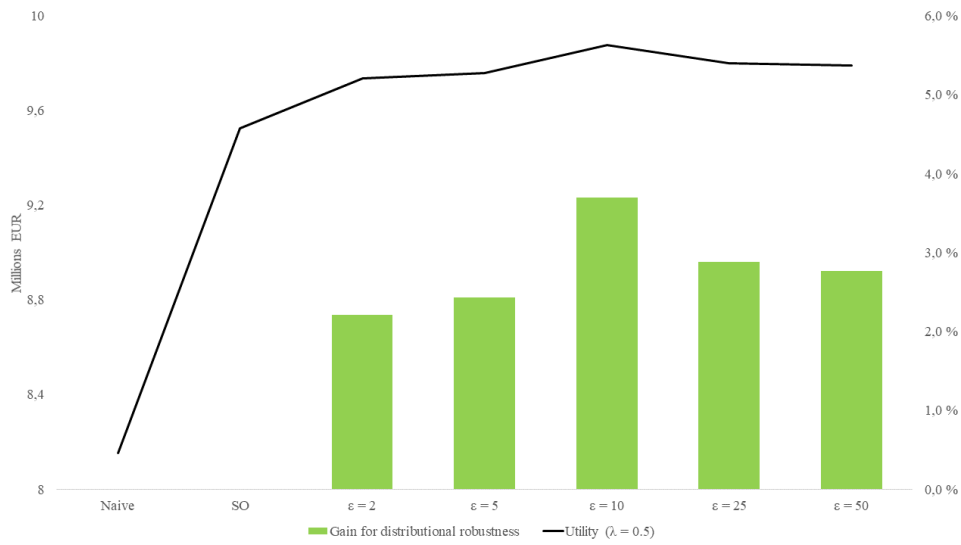


Figure B.4: The utilities for the naive-, stochastic- and distributionally robust hedging strategies under the worst case distribution are measured on the primary axis. The gain of distributionally robustness is measured on the secondary axis. We observe that the gain of distributionally robustness is highest for the distributionally robust model with $\varepsilon = 10$. Further, this gives a significantly higher gain of distributionally robustness than price of ambiguity, as is shown in Chapter 7.

B.4 Details from backtesting the hedging models

This section complements the backtesting of the hedging strategies on historical data from 2014 to 2017, conducted in Chapter 7. Each of the figures below show the following for the considered year: Realized profits at from the different hedging models (*left*). Realized hedge ratios from the different models split on months, as well as historical spot prices and the average achieved forward prices on the hedged quantum (*middle*). Relative deviations in historical production quantities compared to the expected production (*right*).

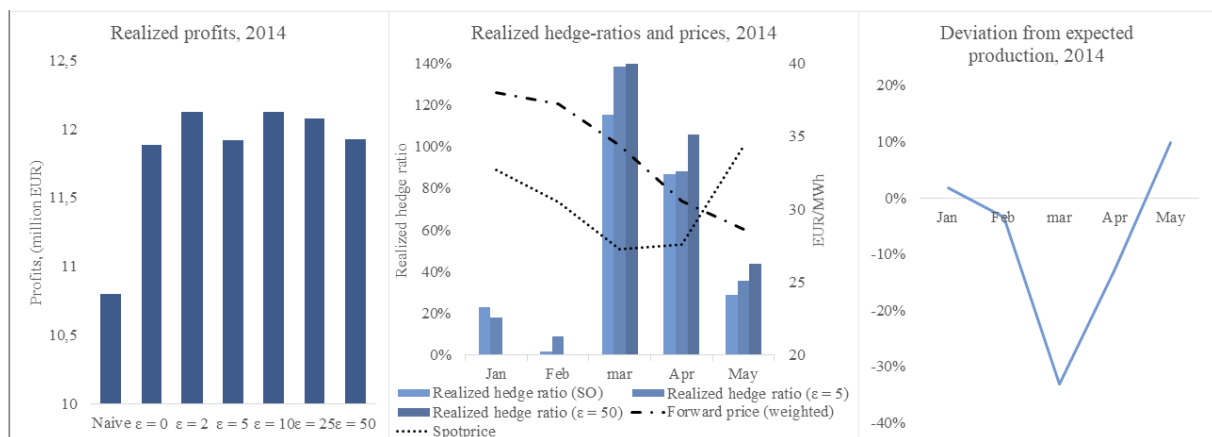


Figure B.5: 2014: We observe that the realized profits tend to increase with higher degrees of ϵ . This is because the achieved forward prices exceed the spot price for the first four months. The realized hedge ratios are over 100% in March following the negative deviation in realized production relative to the expectation for that month. The latter emphasizes the risk of high hedge ratios under production uncertainty.

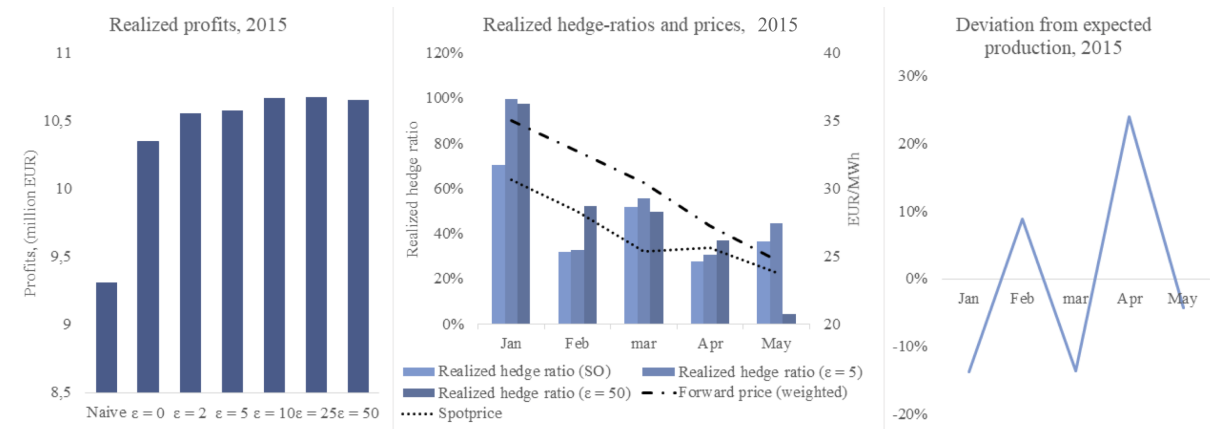


Figure B.6: 2015: We observe that high hedge ratios was favourable in terms of profits because the achieved forward price exceeded the spotprice for all months. We further observe low hedging for $\epsilon = 50$ in May, in line with the discussion in Chapter 7. In turn this gives slightly lower profits for this model.

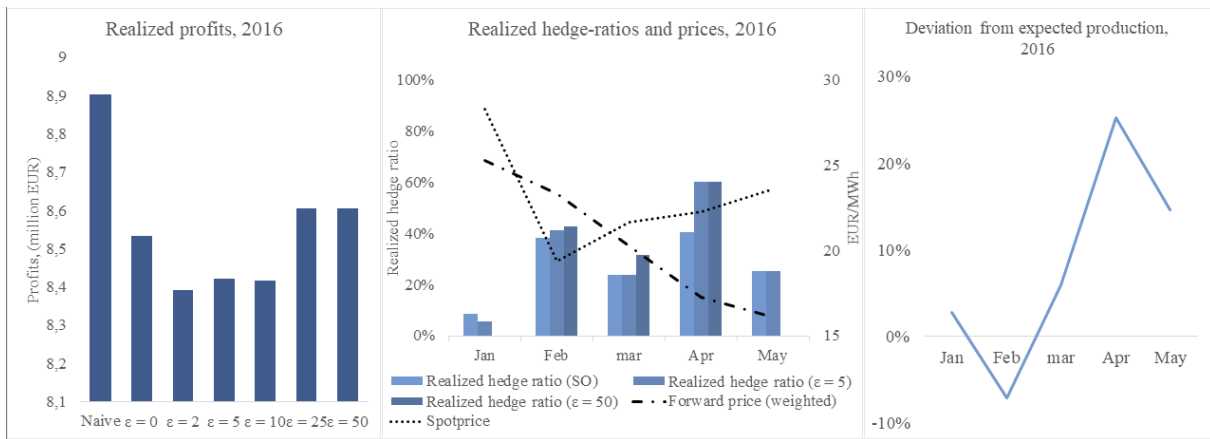


Figure B.7: 2016: We observe decreasing realized profits for strategies up to $\epsilon = 10$. For higher ambiguity radii however, profits are higher. This follows as $\epsilon = 25$ and $\epsilon = 50$ gave close to zero hedging in May, in line with the discussion in Chapter 7, and that the spotprice in this month was significantly higher than the achieved forward price.

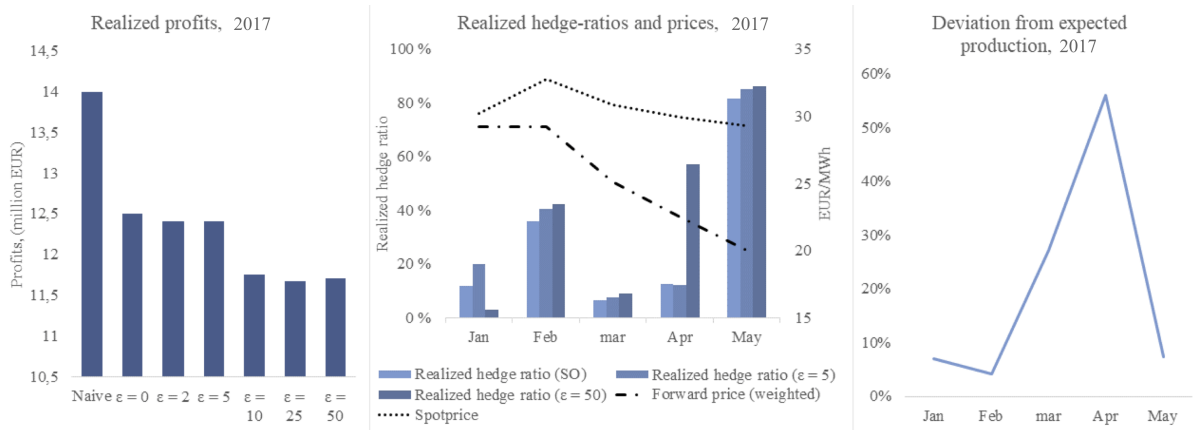


Figure B.8: 2017: We observe strictly decreasing revenues for the models with higher ambiguity radii as forward prices were exceeded by the spotprice in every month. Further, we observe a high deviation between the expected and realized production in April.