DISSERTATION

Titel der Dissertation

ĂMultistage Stochastic Optimization of Energy Portfolios under Model Ambiguityñ

Verfasserin

Bita Analui

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 094136
Dissertationsgebiet lt. Studienblatt:	Statistik und Operations Research
Betreuer:	o. Univ.-Prof. Mag. Dr. Georg Pflug

I hereby declare that this thesis entitled *"Multistage Stochastic Optimization of Energy Portfolios under Model Ambiguity"* is the result of my own research except as cited in the references. This thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Student : Bita Analui

Date : 8 May 2014

Supervisor : Prof. Georg. Ch. Pflug

Internal Referee: Ass.-Prof. Mag. Dr. Ivana Ljubic, University of Vienna.

External Referee: Prof. Dr. Karl Frauendorfer, University of St. Gallen.

For my beloved, most precious Mom and Dad

Acknowledgment

I would like to express myself using humble words of the Persian polymath Ibn Sīnā.[1][2]

دل گرچه در این بادیه بسیار شتافت   یک موی ندانست ولی موی شکافت

I propound these paragraphs as a token of my deep gratitude for all incredible individuals who made this journey possible.

First and for most, I am truly grateful to my PhD supervisor Prof. Georg Pflug, for his substantial guidance, supports and insights without which this thesis would not have been completed. I have been fortunate to benefit from his profound knowledge throughout my research.

I would also like to express my sincere respect to my colleagues at ISOR and IK Computational Optimization for providing a pleasant and dynamic working environment.

Special thank goes to my senior colleagues Dr. Raimund Kovacevic and Alois Pichler, PhD for their invaluable guides and feedbacks.

I would also like to thank all administrative and technical staff at ISOR.

Completion of this thesis would not have been possible without the financial support which was partly provided by Vienna Science and Technology Fund (WWTF) and partly by Initiativkolleg Computational Optimization University of Vienna.

[1] a.k.a Avicenna (980-1037 A.D.) in the West.

[2] Meaning: I have wandered far to learn and investigated the subjects deeply. Yet, the more I learned, the more I realized I know very little.

Furthermore, I would like to cordially appreciate my friends in Vienna with whose presence, I did not feel too far from home in the last four years.

Last but not least, I would like to express my profound gratitude to my family, Mom, Dad and my brothers: Behnam, Behrad and Behdad whose unconditional love has enlightened every single day of my life and their constant support has driven me to where I stand today. I am highly indebted to them.

Bita Analui,

8 May 2014.

Contents

List of Figures

List of Tables

Table 0.0.1: Nomenclature

Symbol	**Explanations**
n_+	Direct successor of node n on the tree $\mathbb{T}$
n_-	Direct predecessor of node n on the tree $\mathbb{T}$
$\mathrm{conv}(X)$	Convex hull of set X
$\mathrm{dl}_r(\mathbb{P}, \tilde{\mathbb{P}})$	Nested distance of order r between two nested distributions $\mathbb{P}$ and $\tilde{\mathbb{P}}$
$\mathbf{d}(P, \tilde{P})$	Elementary distance between two probability measures P and $\tilde{P}$
$\mathbf{d}_{KA}(P, \tilde{P})$	Kantorovich distance between two probability measures P and $\tilde{P}$
$\mathsf{d}(\cdot, \cdot)$	Distance between the elements of the state space
$\mathcal{F}$	$\sigma-$algebra on space Ω
$\mathcal{F}_t$	$\sigma-$algebra on space Ω generated by $(\xi_1, \xi_2, \cdots, \xi_t)$ or $\mathcal{F}_t = \sigma(\xi_1, \xi_2, \cdots, \xi_t)$
$x_t \lhd \mathcal{F}_t$	x_t is measurable w.r.t $\mathcal{F}_t$
$\mathfrak{F}$	Filtration; increasing sequence of $\sigma-$algebras, $\mathfrak{F} = (\mathcal{F}_t)_{t=0}^T$ and $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_T$
$\mathbb{E}[\cdot]$	Expected value function $\mathbb{E}[Y] = \int_\Omega Y(\omega)\mathrm{d}P(\omega)$ w.r.t. measure P
$(\Omega, \mathcal{F})$	Generic Measurable Space
$(\Omega, \mathcal{F}, P)$	Generic Probability Space
$(\Omega, \mathfrak{F}, P)$	Filtered Probability Space
$(\Omega, \mathfrak{F}, P; \xi) \sim \mathbb{P}$	Particular realization of nested distribution $\mathbb{P}$
$\mathcal{C}(\mathbb{P}, \tilde{\mathbb{P}}; \lambda)$	Compound convex combination of $\mathbb{P}$ and $\tilde{\mathbb{P}}$ with probability measure λ
$\mathbb{T}$	Generic tree
$\mathbb{P}(\mathbb{T}, P)$	Nested distribution with a finite tree $\mathbb{T}$ and scenario probabilities P
$\mathcal{P}$	Ambiguity set

Chapter 1

Introduction

1.1 Overview

Multistage stochastic optimization is a well studied, suitable framework where decisions have to be taken in a sequential manner only based on the available information and incorporating stochasticity of the relevant parameters. Since the first formulations, this approach has been applied to various fields such as portfolio optimization, energy production, transportation, telecommunication, etc., in which decision makers are faced with uncertainty about the model parameters. Literature has shown that solving these problems analytically is only possible in very rare cases. Hence, in solving the real world class of problems one critical question is how to model and approximate the reality such that the optimal solution to the approximate problem is as close as possible to the real solution. We put one step further and ask in addition, what if the modeled reality is subject to uncertainty itself, how are we going to robustify the solution against the worst case and for that, which price do we pay. These are the questions which we are going to answer throughout this thesis for multistage stochastic optimization problems.

For a moment, let us look at stochastic optimization problem in general. The assumption of having the complete knowledge about the probability model is already an archaic part of stochastic optimization *folklore*.[1] However, this assumption is unrealistic and often too optimistic. On the other hand solving the problem with respect to the worst case of classical robust optimization is too conservative.

[1] I borrowed this term from Römisch and Wets [43].

One approach which bridges these two extremes, has been studied since 1958 and was pioneered by Scarf [50]. This approach suggests that when information about the true underlying model is partial, the stochastic optimization problem should be solved for a set of models to which (it is believed) the true model belongs. Assume that $H(x, \xi)$ is a convex[2] cost function, and $\xi \sim P$ i.e., probability distribution P represents the uncertainty about ξ. Then minimizing the expected costs under $x \in X$ constraints reads:

$$\min_{x \in X} \mathbb{E}_P[H(x, \xi)] \tag{1.1.1}$$

and under model uncertainty, by defining a set of models $\mathcal{P}$ in which the true model is assumed to belong to, problem (1.1.1) is reformulated with respect to the *worst case*, i.e.:

$$\min_{x \in X} \max_{P \in \mathcal{P}} \mathbb{E}_P[H(x, \xi)]. \tag{1.1.2}$$

Problem (1.1.2) is called *Distributionally Robust* [3] (DR) counterpart of (1.1.1). As a result of extension (1.1.2), the minimization problem turns to minimax formulation covering a class of extremum problems which involve both minimization and maximization. This approach has received many attention and there exists a large number of literature looking at this problem parametrically, meaning that only some characteristics of probability distribution P is known, for instance, some first moments and/or its support. Accordingly, set $\mathcal{P}$ is constructed based on these information. Although the existing literature and prior works in this area are invaluable, we require to construct a new structural body which is compatible with the more involved structure of multistage stochastic optimization problems.

1.2 Contributions and Structure of the Thesis

In our study through the existing literature, very few non-parametric approaches toward explicit formulation of *Distributionally Robust Multistage Stochastic Optimization* (DRMSO) problems were found. Specifically, computational algorithms for finding minimax solutions for class of DRMSO problems are rare and have not been often elaborated in the literature.[4] To this end, we will review these prior works

[2] H is convex in x for every ξ.

[3] The first time term distribultionally robust was appeared in the literature dates back to 1999 by Barmish and Shcherbakov [4] in context of least squares.

[4] It is important to clarify our terminology of DRMSO here. We refer to the problem, the ambiguity problem and the minimax solution is, in fact, the distribitonally robust solution.

and add our further contribution to this area. However, our aim is not evaluating these approaches since responding to our research questions basically requires a new perspective and mathematical setting. Our contribution to this thesis is therefore, threefold:

1. **Introducing theoretical basis for distibutionally robust risk neutral multistage stochastic optimization problems and respective minimax theorem**. Chapter 2 is devoted to introducing an explicit framework for distributionally robust counterpart of risk neutral multistage stochastic optimization problems. Construction of ambiguity sets based on nested neighborhoods around the baseline model are proposed in order to facilitate adapting the assumptions of classical minimax theorem into multistage setting. Consequently, multistage minimax theorem is stated and proved. The proof implies the existence of saddle point solution for distributionally robust multistage stochastic optimization problems.
2. **Algorithms for computing saddle point of DRMSO problems and corresponding convergence results**. Chapter 3 concentrates on carefully designing an algorithmic approach which respects the information structure involved in the underlying uncertainty model, is computationally tractable and guaranties against falling into trap of non-linearity (since conditional probabilities are involved). To this end, we propose an iterative algorithm which for a fixed linear criterion moves toward the worst case stage-wise. Moreover, we prove that this iterative procedure converges to the saddle point.
3. **Numerical results regarding the ambiguity radius and sensitivity analysis of minimax decisions.** Our main computational results are presented and discussed in Chapters 4 and 5. Since we are basing the whole ambiguity problem on the nested neighborhoods of a baseline model, which in class of real problem are mainly history-data driven, it is very crucial to invest enough effort in modeling the baseline too. This step consists of first: scenario generation and second: tree construction. Moreover, in approximating a continuous stochastic process by a discrete one the quality of scenario generation strongly influences the quality of the approximation. In this regard, although we do not develop original results in modeling the baseline, we use novel existing approaches which provide us with a fine baseline model such that further results on ambiguity problem and minimax decisions are reliable. Therefore, in these chapters, we present all steps of formulating a real world decision

problem from mathematical modeling to ambiguity extension as follows:

- mathematical modeling of a short term hydro electricity production planning using a multistage stochastic optimization approach under weather and market risk;
- using existing model selection and scenario generation approaches to model multivariate price-inflow stochastic process;
- using a novel approximation algorithm to generate the baseline tree;
- employ our proposed algorithm to solve the ambiguity problem for hydro electricity production problem and quantify the cost of ambiguity for different ambiguity radii and analyze the sensitivity of distributionally robust solutions.

Finally Chapter 6, concludes this research and states some of the further steps that can enrich this proposed approach in order to make it more realistic.

Chapter 2

Distributionally Robust Multistage Stochastic Optimization

In this chapter we will first study the existing literature and prior works on the subject of distributionally robust optimization. Then we introduce the fundamental notations required to build the DRMSO and finally, we extend a risk-neutral MSO to its distributionally robust counterpart by defining the nested neighborhoods and present the statement and proof of minimax theorem in the multistage setting. Observing the fact that very few studies have been done on distributional robustness in multistage setting has motivated us to introduce an explicit framework for a distributionally robust counterpart of risk neutral multistage stochastic optimization problems. In addition to examining and adapting the assumptions of classical minimax theorems to the multistage setting, we present the proof and existence of a saddle point solution for DRMSO problems. The contents of this chapter have been initially presented in Analui and Pflug [2].

2.1 Model Uncertainty also known as Model Ambiguity in the Literature

The standard assumption in stochastic optimization is that at the time of decision making, the probability laws of the uncertain parameters are *known* and only the realizations are *unknown* . Experience with realistic applications has shown that the choice of an appropriate probability model is crucial for the quality of the solution. Typically, the structure of the parametric model is chosen in a more or less adhoc manner, e.g. by specifying that the data come from a Gaussian process, and the parameters of the model are estimated on the basis of a sequence of past observations. Not only that trusted results of parameter estimates are confidence regions and not point estimates, but also the model class itself can be chosen erroneously. On the basis of the available information a whole set of models could represent the real phenomenon equally well and this fact is called model ambiguity or model uncertainty. Model ambiguity embodies the class of problems in which the decision maker's knowledge about the true underlying probability model is partial or incomplete. A careful decision maker should then take all these equivalent models into account when looking for the robust decision strategy. The notion of ambiguity was introduced by Ellsberg [17]. In Ellsberg's words, ambiguity is

> *...a quality depending on the amount, type, reliability and unanimity of information and giving rise to one's "degree of confidence" in an estimate of relative likelihoods. ([17], p. 657).*

In stochastic programming, one way to deal with model ambiguity is to investigate the *stability* behavior of the optimal solution under model perturbations. In general the notion of stability refers to continuity properties of the optimal value and optimal solutions with respect to model perturbations, see e.g. Robinson and Wets [44], Römisch and Schultz [42], Rachev and Römisch [41] and Römisch and Wets [43]. However, the solution considered in these stability investigations is always with respect to *one* single model and its perturbed variants and the question of how to improve decisions under endogenous model uncertainty is not addressed. In Römisch and Wets [43] for instance, the authors give quantitative estimates of distance between pair (optimal value, optimal solution) of a two stage stochastic optimization problem with respect to a given original probability measure and its perturbation. These results are based on general perturbation results for optimization problems derived in Rockafellar and Wets [46], Chapter 7J.

The idea of optimal decisions under ambiguous stochastic models appeared in an early attempt by Scarf [50]. The author studied optimization of a single product inventory problem under an unknown demand distribution with known mean and variance. The problem was formulated as a linear inventory problem seeking the stockage policy which maximizes the minimum profit considering all demand distributions with given mean and variance. Whereas, more sophisticated approaches assume that true underlying probability model belongs to a given class of models and has motivated the utilization of general *minimax* decision rules. This paradigm was pioneered in the mid-1960s by Žáčková a.k.a. Dupačová [59]. This approach was applied for the class of stochastic linear programs with recourse, where results were formulated in terms of two person zero-sum games. The minimax solution was introduced as an optimal pure strategy of the first player in the game and developed further by Dupačová [13, 14]. In Jagannathan [25] the class of an ambiguity set consisting of all probabilities with given first two moments was studied for linear stochastic problems with simple recourse. Many parametric and non-parametric proposals for ambiguity sets for two-stage problems have been made and analyzed among which, the probability models are defined by certain properties such as the support and the moments of corresponding probability distributions or neighborhoods with respect to some appropriate distances. A list of popular classes of probability models is introduced by Dupačová [15, 16] and a very fast growing literature dealing with model uncertainty either from theoretical or applied viewpoint can be found in Chen and Epstein [9], Calafiore [7], Shapiro and Kleywegt [53], Shapiro and Ahmed [52], Pflug and Wozabal [39], Thiele [56], Delage and Ye [11], Goh and Sim [22] and the list extends. In the following few structures of ambiguity set are discussed in more detail. When it is required, a probability measure $P = (p_1, p_2, ..., p_n)$ is considered as the baseline measure. Shapiro and Kleywegt [53] propose an approach with a Bayesian flavor. Suppose $\mathcal{P}$ be a family of probability measures to which true P belongs. In other words, $\mathcal{P}$ is viewed as the set of conceivable probability measures relevant for the considered problem. Since $\mathcal{P}$ can be very large and generic, the authors consider the set $\mathcal{P} = \{\tilde{P}_1, \tilde{P}_2, ..., \tilde{P}_k\}$ to be finite. Then by assigning *a priori* probabilities $\lambda_1, ..., \lambda_k$ to the corresponding measures, a unique probability measure $\bar{P}$ is obtained:

$$\bar{P} = \sum_{i=1}^{k} \lambda_i \tilde{P}_i \text{ such that } \tilde{P}_i \in \mathcal{P},\ \sum_{i=1}^{k} \lambda_i = 1,\ \lambda_i \geq 0.$$

This approach simplifies the corresponding minimax problem to a minimization problem and can be interpreted as averaging over the set of possible models. One crucial point here is the choice of prior distribution λ.

In another attempt, Shapiro and Ahmed [52] propose a specific class of minimax stochastic programming, when the set $\mathcal{P}$ is defined in the following form: for positive measures P_1 and P_2[1], real valued measurable functions ϕ_i, on $(\Omega, \mathcal{F})$

[1] $P_1 \prec P_2$ means that $P_1(A) < P_2(A)$ for all measurable sets $A \in \mathcal{F}$.

and given real numbers $b_i \in \mathbb{R}$ for $i = 1, ..., n$,

$$\mathcal{P} = \{\tilde{P} : \ P_1 \prec \tilde{P} \prec P_2 \ , \int_\Omega \phi_i \mathrm{d}\tilde{P} = b_i \ i = 1, ..., k, \\ \int_\Omega \phi_i \mathrm{d}\tilde{P} \leq b_i \ i = k+1, ..., n, \ b_i \in \mathbb{R}\}.$$

Hence, $\tilde{P}$ being a probability measure (i.e., $\int_\Omega \mathrm{d}\tilde{P} = 1$), implies that P_1 is a positive measure with total mass smaller than one and P_2 is a positive measure with total mass at least one. If $\tilde{P} \in \mathcal{P}$ being bounded by positive measures P_1 and P_2, it would be replaced by $\tilde{P}$ being a non-negative measure, i.e. $\tilde{P} \succeq 0$, then the minimax problem would be simplified to the classical moments problem.

On the other hand, in Pflug and Wozabal [39] a non-parametric approach is proposed and the ambiguity set of the form

$$\mathcal{P} = \{\tilde{P} = (\tilde{p}_1, \tilde{p}_2, ... \tilde{p}_n) : \ \mathbf{d}_{KA}(\tilde{P}, P) \leq \epsilon\}$$

where $\mathbf{d}_{KA}$ represents the Kantorovich distance (cf. Appendix A.1.2), for a case of two-stage portfolio selection problem, is considered.

Besides, another notion of distance is proposed by Calafiore [7] using *Kullback-Leibler* divergence and the ambiguity set is defined by

$$\mathcal{P} = \{\tilde{P} = (\tilde{p}_1, \tilde{p}_2, ... \tilde{p}_n) : \ \sum_{i=1}^{n} \tilde{p}_i \log \frac{\tilde{p}_i}{p_i} \leq \epsilon\},$$

assuming that the true probability measure $\tilde{P}$ is known to belong into the $\epsilon-$Kullback-Leibler neighborhood of baseline measure P.

The list of parametric and non-parametric choices of ambiguity sets for two-stage stochastic optimization problems expands extensively. To our knowledge, and we constantly keep abreast of corresponding literature, there are few studies dealing with distributional robustness of multistage stochastic optimization with a non-parametric approach and an explicit formulation of multistage minimax. Since our approach constructs an ambiguity set in the form of nested neighborhoods, in the following paragraphs we only report two of the existing studies. No evaluation and comparison with our approach is presented.

In their paper, Goh and Sim [22] aim to solve a linear two stage stochastic optimization problem with expectation of recourse costs in the constraints. They propose an approximate solution algorithm for a class of DR problems and then extend this for nonanticipativity. In terms of uncertainty, their access to partial information contains precise knowledge about the support and covariance matrix while mean itself is subject to uncertainty and is assumed that it belongs to an interval. They start with a simple *Linear Decision Rule* (LDR) to tractably approximate the recourse, however, this a conservative approach and they proceed using more complex decision rules. Next, by remapping the uncertainties into higher dimensional spaces, e.g., a scalar uncertainty is segregated into its positive and negative parts, they prove that *segregated* LDR (SLDR) outperforms LDR while the linearity of the problem is retained. However, to generate more improved decision rules *deflected* LDR (DLDR), originally proposed by Chen et al. [8], is considered. DLDR approach is adapted to this context and a series of sub problems are solved in order to exploit the information structure within the model. The authors call this approach *bideflected* LDR (BLDR) and show that it also outperforms SLDR.

Xin et al. [61], on the other hand, extend the pioneering work of Scarf and consider the multistage news vendor problem. Their knowledge about the underlying uncertainty is the support and first two moments of the demand at each stage. They use the concept of time consistency to model the associated uncertainty in the joint distribution of the demand. Distributionally robust counterpart of multistage news vendor problem is formulated in two ways: static and dynamic. Finally, by comparing the optimal policies of two formulations, weak or strong time consistency of static formulation is discussed.

In general, minimax approach is regarded as a body which bridges the gap between the conservatism of robust optimization and the specificity of stochastic optimization. In minimax approach the optimal decisions are sought for the worst case probability models by obtaining the best possible decisions for the most adverse considered circumstances. There is no unique nomenclature for the ambiguity problem. Synonymous names *model uncertainty problem*, *minimax stochastic optimization* and *distributionally robust problem* are also used in the literature.

The new concept which is introduced in this thesis, investigates and basis founding for distributionally robust decision making for multistage stochastic optimization problems. Since we consider multistage decision models, information structure plays a crucial role. When time passes, the initially unknown uncertain scenario values can gradually be observed. Stage-by-stage, the amount of information increases and planning decisions have to be made at each time stage based on the available information, i.e., decisions are taken at times $t = 0, ..., T-1$ with typically different levels of information. We denote the stochastic scenario process by $\xi := (\xi_1, ..., \xi_T)$ and the pertaining multistage decision sequence by $x = (x_0, x_1, ..., x_{T-1})$. The indices of the random process and the decision process differ by one, since at time t, a decision is to be made but the realization of the random process will be observable only at time $t+1$. In addition, decisions at each time step t may depend only on the actual outcomes of the random variables up to time t, they must be *nonanticipative* with respect to the observations after time t, i.e., $x_t = x_t(\xi^t)$ with $\xi^t = (\xi_1, ..., \xi_t)$. Therefore, the peculiarity of the multistage situation lays in the fact that model uncertainty has to be defined in terms of the conditional distributions of the scenario process and not just of its multivariate distribution. In Section 2.2, we show that a filtered probability space $(\Omega, \mathfrak{F}, P)$ is used as the basis for multistage stochastic optimization. More specifically, we consider a discrete time (multivariate) scenario process ξ with finite time horizon which sits on Ω such that $\xi(\omega) := (\xi_1(\omega), ..., \xi_T(\omega))$. For computational purposes this filtered probability space and the scenario process is represented by a valuated tree. For a broad technical presentation of multistage stochastic programming refer to Birge and Louveaux [5], Pflug and Römisch [38] and Ruszczynski and Shapiro [48].

Since for multistage optimization problems, not only the marginal distributions of the scenario process but also the information structure should be taken into account, we argue here that it is quite natural to base the ambiguity set on the nearness of the nested distributions. To this end, we apply the concept of nested distances for the nested distributions. Neglecting the information structure and looking only at the multivariate distributions of the scenario processes leads to counter intuitive examples (cf. Pflug and Pichler [36], Example 1. and Heitsch et al. [23]). On the other hand, the nested distance initially introduced by Pflug and Pichler [36] is a suitable concept for dealing with the information structure in the multistage decision context. The following example would motivate the fact that nested distance is a natural choice when measuring nearness of nested distributions. When this is established, ideation of model uncertainty that leads to ambiguity problem in multistage stochastic optimization, despite the complicated structure, appears more intuitive.

Example 1. The trees in Figure 2.1.1, represent three similar 2-stage processes. The only difference are the slight changes in scenario probabilities. Consider P as

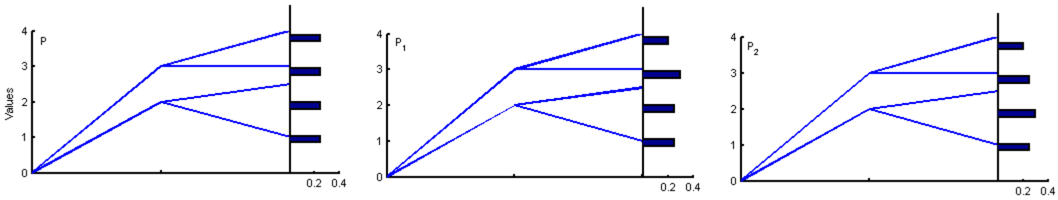


Figure 2.1.1: Motivational example: Nearness of finite trees

the baseline model, with variants P_1 and P_2 as presented below:

$$P = (0.25,\, 0.25,\, 0.25,\, 0.25), \quad P_1 = (0.2,\, 0.3,\, 0.25,\, 0.25), \quad P_2 = (0.2,\, 0.25,\, 0.3,\, 0.25)$$

The stochastic value process $\xi(\omega)$ for $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ reads:

$$\xi(\omega_1) = (0, 3, 4),\ \xi(\omega_2) = (0, 3, 3),\ \xi(\omega_3) = (0, 2, 2.5),\ \xi(\omega_4) = (0, 2, 1)$$

We calculate the distance between model P and its variants using the following distances:

- the naive distance : $\mathbf{d}(P, \cdot) = \sum_{i=1}^{4} |p_i - \tilde{p}_i|$ calculates the distance between P and its variants in the most primitive way,

- the Kantorovich distance: $\mathbf{d}_{KA}(P, \cdot)$ (see Appendix A.1.2),

- the nested distance:[2] $\mathrm{dl}(P, \cdot)$ (cf. Definition 2).

The results are shown in Table 2.1.1. Besides the simple norm $\mathbf{d}$ between the probabilities, it can be observed that even the Kantorovich distance, $\mathbf{d}_{KA}$, does not use any information about the structure of the tree. The nested distance, on the other hand, generates a much finer topology. It is proved by Pflug and Pichler [36] that $\mathbf{d}_{KA}$ is a lower bound for dl. Hence, $\mathbf{d}_{KA}(P, \tilde{P}) \le \epsilon$ does not necessarily imply $\mathrm{dl}(P, \tilde{P}) \le \epsilon$, because there might exist an ϵ such that $\mathbf{d}_{KA}(P, \tilde{P}) \le \epsilon \le \mathrm{dl}(P, \tilde{P})$. In our example,this fact is emerged in $\mathrm{dl}(P, P_2) = 0.193 \gg 0.05$. We wrap up the example by arguing that P_2 is not included in the set of alternative models which are in 0.05-nested neighborhood of P and in fact it is *far away* from this neighborhood.

[2] The distance between *conditional* probability measures vs. distance between probability measures a.k.a Kantorovich distance.

Table 2.1.1: Distances and Information Structure

	P_1	P_2
$\mathbf{d}(P,\cdot)$	0.1	0.1
$\mathbf{d}_{KA}(P,\cdot)$	0.05	0.07
$\mathrm{dl}(P,\cdot)$	0.05	0.193

2.2 Multistage Stochastic Optimization - Risk Neutral Formulation

Here, we briefly discuss the risk neutral formulation for the multistage linear stochastic optimization problem. Consider the compact formulation

$$\min_x \{\mathbb{E}_{\mathbb{P}}[H(x,\xi)] : \; x \in \mathbb{X}, \; x \lhd \mathfrak{F}; \; (\Omega, \mathfrak{F}, P; \xi) \sim \mathbb{P}\}, \tag{2.2.1}$$

where H is a real-valued convex cost function, depending on decision sequence $x = (x_0, ..., x_{T-1})$ and stochastic value process $\xi = (\xi_1, ..., \xi_T)$. The first stage data are deterministic, i.e., ξ_0 is known. At each stage t, decision x_t has to be made and its value is required to lie in a feasible set $\mathbb{X}_t$, a linear vector space. The whole decision space is denoted by $\mathbb{X} := (\mathbb{X}_t)_{t=0,\ldots,T-1}$. The stochastic process ξ describes the economic environment of the decisions (e.g. future prices, demands, external supplies,...) and is defined by its nested distribution. Assume for a moment that this process is defined on a given filtered probability space $(\Omega, \mathfrak{F}, P)$, where $\mathfrak{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_T)$ is a filtration such that ξ_t is measurable w.r.t. $\mathcal{F}_t$, which is denoted by $\xi_t \lhd \mathcal{F}_t$ (and for the whole process $\xi \lhd \mathfrak{F}$). The nested distribution is the collection of conditional distributions of ξ_t given $\mathcal{F}_{t-1}$, written as $\xi_t|\mathcal{F}_{t-1}$, more precisely of the nested structure $(((((\xi_T|\mathcal{F}_{T-1}), \xi_{T-1}|\mathcal{F}_{T-2}) \ldots), \xi_2|\mathcal{F}_1), \xi_1)$. It turns out that the nested distribution is the right concept to formulate the distribution of the scenario process and the information structure given by the filtration independent of a concrete probability space. That is, two processes which may be defined on different probability spaces, but can - together with the respective filtrations - be mapped to each other by a bijective transformation, share the same nested distribution. For a proof and more discussion about the concept of nested distribution see Pflug [35]. We denote the nested distribution by $\mathbb{P}$ and notice that it can be concretized to a process ξ defined on a filtered probability space $(\Omega, \mathfrak{F}, P)$ if a concrete model is needed. The notation $(\Omega, \mathfrak{F}, P; \xi) \sim \mathbb{P}$ symbolizes this. We assume that the decisions at stage t are nonanticipative, i.e. measurable w.r.t. $\mathcal{F}_t$.

In particular problem (2.2.1) can be considered as linear, if the objective function and the corresponding constraints are linear.

Recalling the Kantorovich distance (cf. Appendix A.1), the distance between two nested structures is closely inline with the concept of transportation distances. For nested distributions: baseline model $(\Omega, \mathfrak{F}, P; \xi) \sim \mathbb{P}$ and an alternative model $(\tilde{\Omega}, \tilde{\mathfrak{F}}, \tilde{P}; \tilde{\xi}) \sim \tilde{\mathbb{P}}$ a concept of distance for the nested distributions, nested (multistage) distance, has been introduced by Pflug and Pichler [36] which allows to quantify the model error. We recall this definition here:

Definition 2. The nested (multistage) distance of order $r \geq 1$ of two nested-structures $(\Omega, \mathfrak{F}, P; \xi) \sim \mathbb{P}$ and $(\tilde{\Omega}, \tilde{\mathfrak{F}}, \tilde{P}; \tilde{\xi}) \sim \tilde{\mathbb{P}}$ is the optimal value of the optimization problem

$$\begin{array}{lll} \min\limits_{\pi} & \left(\int \mathsf{d}(\omega, \tilde{\omega})^r \pi[\mathrm{d}\omega, \mathrm{d}\tilde{\omega}]\right)^{\frac{1}{r}} & \\ \text{s.t.} & \pi[A \times \tilde{\Omega} | \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t](\omega, \tilde{\omega}) = P[A|\mathcal{F}_t](\omega) & (A \in \mathcal{F}_T,\ 1 \leq t \leq T) \\ & \pi[\Omega \times B | \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t](\omega, \tilde{\omega}) = \tilde{P}[B|\tilde{\mathcal{F}}_t](\tilde{\omega}) & (B \in \tilde{\mathcal{F}}_T,\ 1 \leq t \leq T). \end{array} \tag{2.2.2}$$

Here the infimum in (2.2.2) is taken among all bivariate probability measures π defined on $\mathcal{F}_T \otimes \tilde{\mathcal{F}}_T$ and d is a distance for the realizations of the stochastic scenario processes (defined on the original sample space $\Omega \times \tilde{\Omega}$), for instance

$$\mathsf{d}(\omega, \tilde{\omega}) = \sum_{t=1}^{T} \sum_{m=1}^{M} w_t^m \left\| \xi_t^m(\omega) - \tilde{\xi}_t^m(\tilde{\omega}) \right\|_1 \tag{2.2.3}$$

where w_t^m are some weights, reflecting discounting in time and re-weighting (scaling) different dimensions of the M-dimensional process ξ. The optimal value of (2.2.2) is the nested distance of order r between two nested structures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ which is denoted by $\mathrm{dl}_r(\mathbb{P}, \tilde{\mathbb{P}})$.

The nested distance is defined for the nested distributions and is independent of the respective realizations on concrete probability spaces. Notice that in particular the processes ξ and $\tilde{\xi}$ can be defined on different probability spaces. It has been proven by Pflug and Pichler [36] that if the criterion function H is Lipschitz in ξ and convex in x, then the optimal value of the decision problem (2.2.1) is Lipschitz w.r.t. the nested distance. While (2.2.1) describes the general form of a multistage stochastic optimization problem, such problems are often formulated in a finite discrete setup, especially for making them tractable by numerical optimization. Finite nested distributions can be represented by *node*- and *arc* valuated *trees*, where the tree structure reflects the filtration, the node valuation represents the values of the stochastic scenario process ξ and the arc valuation encodes the conditional probability distributions. Again we refer to Pflug [35] for a thorough treatment of scenario tree (better: equivalence classes of scenario trees) as representations of nested distributions. In the following, we consider scenario trees as finite versions of nested distributions. Trees are characterized by the node sets $\mathcal{N}_t$ per stage t, hence $\mathcal{N} = \bigcup_{t=0}^{T} \mathcal{N}_t$, and the predecessor relations $\prec$. If $i \in \mathcal{N}_{t-1}$, $j \in \mathcal{N}_t$ and i is a direct predecessor of j, we write $i \prec j$.[3] The node set $\mathcal{N}_0$ consists only of the root and the node set $\mathcal{N}_T$ (the set of all leaf nodes) can be identified with the probability space Ω. If $j \in \mathcal{N}_t$ and $i \in \mathcal{N}_{t-1}$ with $i \prec j$, then probabilities $Q(j) = \frac{P(j)}{P(j_-)} = \frac{P(j)}{P(i)}$ sitting on the arcs represent the conditional probabilities of reaching node j from node i. The unconditional probabilities P_i sitting on the leaves of the tree represent the probability distribution P on $\Omega = \mathcal{N}_T$. The specialization of Definition 2, for the tree situation is given by Definition 3 (again we refer to Pflug and Pichler [36]). In the following, we only consider the nested distance of order $r = 1$ i.e. $\mathrm{dl}(\cdot,\cdot)$, however all results can be generalized for $r > 1$.

Definition 3. The nested distance of order $r = 1$ between two tree models $\mathbb{P}$ and $\tilde{\mathbb{P}}$ is given by the optimal value of the following large linear program

$$\begin{aligned}
\mathrm{dl}(\mathbb{P},\tilde{\mathbb{P}}) = \min_{\pi} \quad & \sum_{i',j' \in \mathcal{N}_T} \mathsf{d}(i',j')\, \pi(i',j') \\
\text{subject to} \quad & \sum_{j \in l+} \pi(i,j|k,l) = Q(i) \qquad (i \in k+,\ l) \\
& \sum_{i \in k+} \pi(i,j|k,l) = \tilde{Q}(j) \qquad (k,\ j \in l+) \\
& \sum_{i',j'} \pi(i',j') = 1 \\
& \pi(i',j') \geq 0.
\end{aligned} \tag{2.2.4}$$

Here $\mathsf{d}(i',j')$, the distance of the process at the leaves $i' \in \mathcal{N}_T$ and $j' \in \tilde{\mathcal{N}}_T$, are given by a distance between the paths leading to i' resp. j' similar to (2.2.3). $\pi(i',j')$ runs through all joint probability distributions on $\Omega \times \tilde{\Omega} = \mathcal{N}_T \times \tilde{\mathcal{N}}_T$, which we call *transportation plans.* The conditional probabilities in a transportation plan

[3] If k is any predecessor of j we write $k \precsim j$. Notation $pred_u(j)$ that denotes the predecessor of j in $\mathcal{N}_u$, with $u < t$ might also be used. If $u = t-1$ the notation is written as $pred_{t-1}(j)$ or j_-.

are given by $\pi(i,j|k,l) = \frac{\sum\limits_{i \precsim i',\ j \precsim j'} \pi(i',j')}{\sum\limits_{k \precsim i',\ l \precsim j'} \pi(i',j')}$[4] and are called *transportation subplans,* therefore the constraints in (2.2.4) can be reformulated for the unconditional probabilities P and $\tilde{P}$ as

$$\begin{aligned} P(i) \cdot \sum_{k \precsim i',\ l \precsim j'} \pi(i',j') &= P(k) \cdot \sum_{l \precsim j'} \pi(i,j') \qquad (i \in k+,\ l) \\ \tilde{P}(j) \cdot \sum_{k \precsim i',\ l \precsim j'} \pi(i',j') &= \tilde{P}(l) \cdot \sum_{k \precsim i'} \pi(i',j) \qquad (k,\ j \in l+). \end{aligned}$$

When P and $\tilde{P}$ are given, reformulated problem (2.2.4) is indeed a linear program (Pflug and Pichler [37]) and reads:

$$\begin{aligned} \mathrm{dl}(\mathbb{P},\tilde{\mathbb{P}}) = \min_{\pi} \quad & \sum_{i',j' \in \mathcal{N}_T} \mathsf{d}(i',j')\ \pi(i',j') \\ \text{subject to} \quad & P(i) \cdot \sum_{k \precsim i',\ l \precsim j'} \pi(i',j') = P(k) \cdot \sum_{l \precsim j'} \pi(i,j') \qquad (i \in k+,\ l) \\ & \tilde{P}(j) \cdot \sum_{k \precsim i',\ l \precsim j'} \pi(i',j') = \tilde{P}(l) \cdot \sum_{k \precsim i'} \pi(i',j) \qquad (k,\ j \in l+). \\ & \sum_{i',j'} \pi(i',j') = 1 \\ & \pi(i',j') \geq 0. \end{aligned}$$

In Figure 2.2.1, the nested structure of transportation matrix π (induced by two trees of the same height and structure) together with the schematic distance matrix d is depicted. In this thesis, we intend to construct the ambiguity set as a nested neighborhood around a given tree model $\mathbb{P}$. Albeit the concept of nested distance provides us with an appropriate tool for constructing these neighborhoods, this approach appears to be not as effortless as calculating the nested distance between two trees. In the next section the distributionally robust counterpart of model (2.2.1) is derived and its structural properties are discussed.

Now that the linear programing formulation of nested distance is established, I should declare that in the subsequent chapters and for simplicity of notations, we will switch to $i,\ j$ to represent the leaf nodes.

[4]This quotient necessitates inclusion of constraint $\sum\limits_{i',j'} \pi(i',j') = 1$ in (2.2.4), otherwise every multiplication of any feasible transportation plan π, would be feasible.

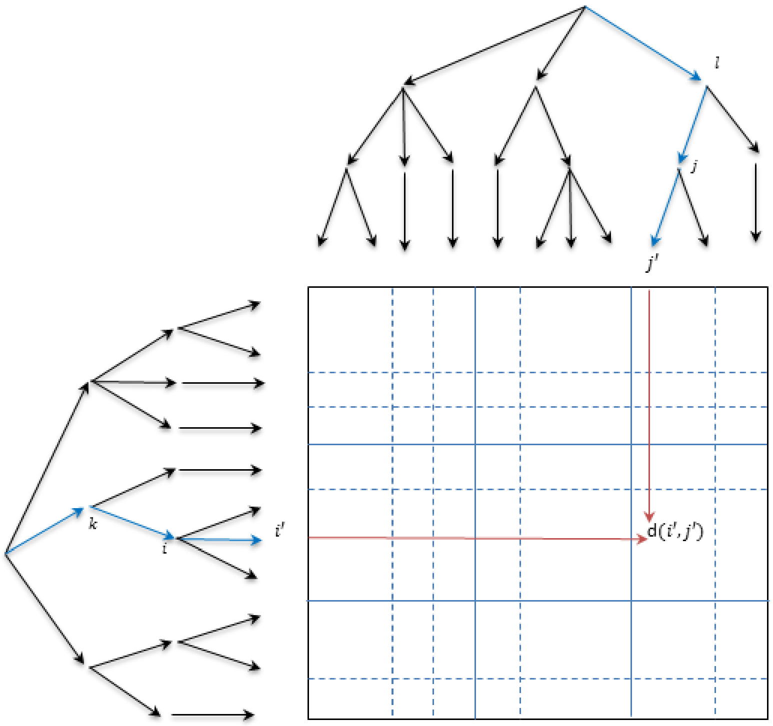


Figure 2.2.1: Visualization of transportation matrix π and distance matrix d for two trees. $i^{'}, j^{'}$ are leaf and k, l are generic intermediate nodes.

2.3 Distributionally Robust Multistage Stochastic Optimization[5]

As it was addressed before, one approach to deal with model uncertainty is to look for the equilibrium strategies among a set of possible models. The distributionally robust counterpart of (2.2.1) is given by

$$\min_{x} \max_{\mathbb{P} \in \mathcal{P}_\epsilon} \{\mathbb{E}_{\mathbb{P}}[H(x, \xi)] : \ x \in \mathbb{X}, \ x \lhd \mathfrak{F}\},$$

where $\mathcal{P}_\epsilon$ denotes an ambiguity set of probability models. In the present work we consider balls with radius ϵ[6] around a baseline model $\mathbb{P}$ w.r.t the nested distance $\mathbb{d}$

[5]The results which are presented in this section are elaborated version of corresponding section in Analui and Pflug [2].

[6]Preprocessing about the range of ϵ, being strongly problem dependent, is discussed in Chapters 3 and 5.

of order 1, i.e.:

$$\mathcal{P}_\epsilon = \{\tilde{\mathbb{P}} : \, \mathrm{dl}(\mathbb{P}, \tilde{\mathbb{P}}) \leq \epsilon\}. \tag{2.3.1}$$

The distributionally robust counterpart reads now

$$\min_x \{\max_{\tilde{\mathbb{P}}} \mathbb{E}_{\tilde{\mathbb{P}}}[H(x,\xi)] : \; x \in \mathbb{X}, \; x \lhd \tilde{\mathfrak{F}}, \; (\tilde{\Omega}, \tilde{\mathfrak{F}}, \tilde{P}; \tilde{\xi}) \sim \tilde{\mathbb{P}}, \; \mathrm{dl}(\mathbb{P}, \tilde{\mathbb{P}}) \leq \epsilon\}. \tag{2.3.2}$$

In its general form, problem (2.3.2) is quite difficult to solve. Even in the single-stage case, it requires algorithms for non-convex optimization such as DC-algorithms, see Wozabal [60]. For this reason, we will consider a smaller ambiguity set, where the filtration and the scenario process are fixed and only the arc probabilities vary, i.e., $(\Omega, \mathfrak{F}, \tilde{P}; \xi) \sim \tilde{\mathbb{P}}$.

In addition to the assumptions which were discussed about problem (2.2.1), it is assumed that $\mathcal{P}_\epsilon$ is a non empty set of probability models which w.l.o.g all share the same state space Ω and filtration $\mathfrak{F}$, function $H(x,\xi)$ is $\mathfrak{F}-$measurable and $\tilde{\mathbb{P}}-$integrable. By the assumptions above, and for every $\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon$, mapping $\tilde{\mathbb{P}} \mapsto \mathbb{E}_{\tilde{\mathbb{P}}}[H(\cdot,\xi)]$ is linear on a neighborhood of $\mathbb{X}$. Since $\mathbb{E}_{\tilde{\mathbb{P}}}[H(\cdot,\xi)]$ is well defined and linear in $\tilde{\mathbb{P}}$, the minimum in (2.3.2) does not change if $\mathcal{P}_\epsilon$ is replaced by its closed convex hull.

To this end, we introduce the following notation: Let $\mathbb{T}$ denote a tree with given structure (e.g., given number of the nodes per stage) valuated by the scenario process. The leaf set (scenarios) of $\mathbb{T}$ is denoted by $\Omega = \mathcal{N}_T$. The probability valuations are given by the scenario probabilities $P = (P_i)_{i \in \mathcal{N}_T}$. The fully valuated tree is denoted by $\mathbb{P}(\mathbb{T}, P)$. Even in cases that the structure and the values of the scenario process are fixed and only the scenario probabilities vary, it would be inconsistent to define simply ambiguity sets as neighborhoods of P, such as

$$\left\{ \tilde{P} : \sum_{i \in \mathcal{N}_T} |P_i - \tilde{P}_i|^r \leq \epsilon^r \right\}. \tag{2.3.3}$$

The reason is that an ambiguity set of the form (2.3.3) does not respect the tree structure (recall the naive distance in Example 1.)

As already mentioned, we restrict ourselves in the following to alternative models, which are defined on the same tree structure of the baseline model, but only with the varying probabilities. However, we keep the ambiguity set as a ball in the nested distance sense, i.e. we specify (2.3.1) to

$$\mathcal{B}_\epsilon = \{\tilde{P} : \, \mathrm{dl}(\mathbb{P}(\mathbb{T}, P), \mathbb{P}(\mathbb{T}, \tilde{P})) \leq \epsilon\} \tag{2.3.4}$$

and set

$$\mathcal{P}_\epsilon = \left\{\mathbb{P}(\mathbb{T}, \tilde{P}) : \, \tilde{P} \in \mathcal{B}_\epsilon\right\}.$$

The final formulation of the ambiguity extension problem is now

$$\min_x \max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} \{\mathbb{E}_{\tilde{\mathbb{P}}}[H(x, \xi)] : \, x \in \mathbb{X}, \, x \lhd \mathfrak{F}\}. \tag{2.3.5}$$

After a brief review on classical minimax theorems, in the next subsection, we amplify the ambiguity set to its closed convex hull i.e., $\bar{\mathcal{P}}_\epsilon$, in order to apply a minimax theorem and identify a saddle point. In addition, we show that the *worst case* model is contained in the original ambiguity set $\mathcal{P}_\epsilon$.

2.3.1 Minimax Theorems

A brief anecdote regarding the history of minimax theorem, takes us to von Neumann's initiation on game of strategy in [58], which can be stated as follows:

Theorem 4. *If X and Y are finite dimensional simplices, e.g., sets of row and column vectors with unit sum of matrix A and $f(x, y) = xAy$ a jointly continuous, bilinear function on $X \times Y$, then f has a saddle point (x^*, y^*) for which the following equality holds,*

$$\max_{x \in X} \min_{y \in Y} f(x, y) = \min_{y \in Y} \max_{x \in X} f(x, y). \tag{2.3.6}$$

Since von Neumann's 1928 paper [58], there have been several extensions and generalizations of the minimax theorem. For instance, allowing spaces X and Y be subsets of infinite dimensional linear spaces. The generalized minimax theorems by Fan [19] and Sion [54] and all the references therein, assert that the min and the max can be interchanged in (2.3.6). The validity of such theorems is strongly related to algebraic (convexity/concavity/linearity) and topological properties of the criterion function and feasible sets. According to Sion [54], there are essentially two types of arguments in proving minimax theorems. First, using separation of disjoint convex sets in Euclidean spaces by a hyperplane which yields Kneser, Fan and Berge Theorems for concave-convex and semi-continuous functions. Second class of arguments, use Brouwer's fixed-point theorem which yield Nikaido's results and function f is allowed to be quasi-concave-convex, but satisfying continuity which certainly is a stronger topological assumption compared to the first argument. Sion [54] however, unifies these results and proves the minimax theorem for quasi-convex-concave functions which are appropriately semi-continuous in each variable. In the following Theorem 5, the result of Sion [54] together with a sketch of the proof, is stated.

Theorem 5. *[54] Let X and Y be convex compact spaces and f a function on $X \times Y$, quasi concave-convex and upper semicontinuous-lower semicontinuous, then*

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) = \inf_{y \in Y} \sup_{x \in X} f(x, y).$$

The idea of the proof stems from convex geometry of level sets which is a valid argument for quasi-concave-convex functions. The method itself is proof by contradiction, meaning that finite number of level sets (open sets) of function $f(x, y)$ with respect to each variable and for given real number c such that $\sup \inf \; f < c < \inf \sup \; f$, cover X resp. Y , and therefore by compactness any finite subsets of which,

$$\begin{cases} \forall\; x \in \operatorname{conv}(X) \;\; \exists\; y \in Y \;\; f(x, y) > c \\ \forall\; y \in \operatorname{conv}(Y) \;\; \exists\; x \in X \;\; f(x, y) < c. \end{cases}$$

In a finite number of steps $X_n \subset X_{n-1} \subset \cdots \subset X$, finite subsets of X and Y arrive to minimal sets X_n and Y_n such that

$$\begin{cases} \exists\; x^* \in \operatorname{conv}(X_n) \;\; f(x^*, y) > c \;\; \forall\; y \in Y \\ \exists\; y^* \in \operatorname{conv}(Y_n) \;\; f(x, y^*) < c \;\; \forall\; x \in X \end{cases} \implies f(x^*, y^*) < c < f(x^*, y^*) \;,$$

which is contradiction.

2.3.2 Minimax Theorem in Multistage Setting

Being adjusted to multistage setting, the question that must be answered is: *In what respect nested distributions allow convex combinations?* It would be incorrect to just form convex combinations of the scenario probabilities, since such combinations are not invariant w.r.t. equivalent permutations of the leaves, i.e., they cannot be formulated in terms of the nested distributions. According to Pflug and Pichler [37], the correct notion of convex combinations of the nested distributions, is in terms of *compounding*. This notion is recalled in Definition 6.

Definition 6. If $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are nested distributions, then their compound distribution with probability λ is given by

$$\mathcal{C}(\mathbb{P}, \tilde{\mathbb{P}}; \lambda) = \begin{cases} \mathbb{P} & \text{with prob. } \lambda \\ \tilde{\mathbb{P}} & \text{with prob. } 1-\lambda. \end{cases}$$

If $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are tree models, then $\mathcal{C}(\mathbb{P}, \tilde{\mathbb{P}}; \lambda)$ is also a tree model, where from a new root subtree $\mathbb{P}$ can be reached with probability λ and subtree $\tilde{\mathbb{P}}$ can be reached with probability $1-\lambda$. Denote by $\mathbb{P}_+$ the degenerated compound model, where the baseline model $\mathbb{P}$ is chosen with probability 1. It is equivalent to $\mathbb{P}$, but has an additional root, from which subtree $\mathbb{P}$ can be reached with probability 1.[7]

The counter Example 7 illustrates that ambiguity set $\mathcal{P}_\epsilon$ is not convex in terms of convex combination of probability vector P and its variants. Therefore we need to consider the closed convex hull of $\mathcal{P}_\epsilon$, i.e., $\bar{\mathcal{P}}_\epsilon$ to be able to state the multistage minimax theorem.

Example 7. Recall the probability models of Example 1 : Consider P as the baseline model, with its variants P_1 and P_2 as presented below:

$$P = (0.25,\, 0.25,\, 0.25,\, 0.25), \quad P_1 = (0.2,\, 0.3,\, 0.25,\, 0.25), \quad P_2 = (0.2,\, 0.25,\, 0.3,\, 0.25).$$

In Example 1, we saw that $\mathrm{dl}(P, P_1) = 0.0500$ and $\mathrm{dl}(P, P_2) = 0.1930$. Hence, we can claim that both P_1 and $P_2 \in \mathcal{P}_{0.1930}$. However, there exists a convex combination of P_1 and P_2, namely $P_3 = (0.2,\, 0.2502,\, 0.2998,\, 0.25)$ such that $\mathrm{dl}(P, P_3) = 0.1947 > 0.1930$ and $P_3 \notin \mathcal{P}_{0.1930}$. Since convexity of nested neighborhoods is a fundamental assumption in minimax theorem, $\mathcal{P}_\epsilon$ is replaced with its closed convex hull, $\bar{\mathcal{P}}_\epsilon$.

It was mentioned before that $\mathbb{E}_{\tilde{\mathbb{P}}}[H(x, \cdot)]$ is well defined and linear with respect to $\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon$, therefore replacing $\mathcal{P}_\epsilon$ with its closed convex hull does not affect the optimal solution. In the following, the multistage minimax theorem for the

[7] This convention green-lights further comparisons between $\mathcal{C}(\mathbb{P}, \tilde{\mathbb{P}}; \lambda)$ and $\mathbb{P}_+$ as nested distributions of the same depth.

extended ambiguity set $\bar{\mathcal{P}}_\epsilon$ is stated and proved accordingly. The theorem follows from general minimax theorems cited above.

Theorem 8. *Let $H(x,\xi)$ be convex in x with a convex and compact decision set $\mathbb{X}$. Then*

$$\min_{x\in\mathbb{X}} \max_{\tilde{\mathbb{P}}\in\bar{\mathcal{P}}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[H(x,\xi)] = \max_{\tilde{\mathbb{P}}\in\bar{\mathcal{P}}_\epsilon} \min_{x\in\mathbb{X}} \mathbb{E}_{\tilde{\mathbb{P}}}[H(x,\xi)]$$

and a saddle point $(x^*,\tilde{\mathbb{P}}^*)$ exists, i.e.

$$\mathbb{E}_{\mathbb{P}}[H(x^*,\xi)] \leq \mathbb{E}_{\tilde{\mathbb{P}}^*}[H(x^*,\xi)] \leq \mathbb{E}_{\tilde{\mathbb{P}}^*}[H(x,\xi)].$$

Moreover, $\tilde{\mathbb{P}}^* \in \mathcal{P}_\epsilon$ and not just in $\bar{\mathcal{P}}_\epsilon$.

Proof. Recalling the assumption of (2.3.5), throughout this proof, a finite $\mathbb{T}$ is fixed with the given structure and scenario values sitting on the node. When the scenario probabilities $P=(P_i)_{i\in\mathcal{N}_T}$ are determined, the corresponding baseline tree $\mathbb{P}(\mathbb{T},P)$ is representable. In addition ambiguity set

$$\mathcal{P}_\epsilon = \left\{\mathbb{P}(\mathbb{T},\tilde{P}) : \ \tilde{P}\in\mathcal{B}_\epsilon\right\},$$

determines that alternative models are $\mathbb{P}(\mathbb{T},\tilde{P})$ with a variant $\tilde{P}$ of only the scenario probabilities, on the same tree $\mathbb{T}$. By generalizing the notion of compounding to allow for infinity many $\tilde{\mathbb{P}}$ and introducing

- family $\mathfrak{P}$ of all probability measures on $\Omega=\mathcal{N}_T$, which is a simplex since $\mathcal{N}_T$ is a finite set,
- a probability measure Λ from $\mathfrak{P}$,
- and Definition 6, the compound $\mathcal{C}(\mathbb{P}(\mathbb{T},\tilde{P}),\Lambda)$ can be defined as

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

meaning that the compound is obtained by first sampling a distribution $\tilde{P}$ according to Λ and then taking the model $\mathbb{P}(\mathbb{T},\tilde{P})$. For intuition, refer to Figure 2.3.1 in which $\mathcal{C}(\mathbb{P}(\mathbb{T},\cdot),\Lambda)$ for probability measure Λ with finite support is illustrated. Notice that the tree of $\mathcal{C}(\mathbb{P}(\mathbb{T},\tilde{P}_\lambda),\Lambda)$ is of height $T+1$. Thus for the original tree $\mathbb{P}(\mathbb{T},P)$ to be comparable with $\mathcal{C}(\mathbb{P}(\mathbb{T},\tilde{P}_\lambda),\Lambda)$, we assume that a further root (with probability *one*) is appended to the tree of $\mathbb{P}(\mathbb{T},P)$ and denote this extended tree by $\mathbb{P}_+(\mathbb{T},P)$. In the following, we write $\mathbb{P}(\mathbb{T},\Lambda)$ for $\mathcal{C}(\mathbb{P}(\mathbb{T},\cdot),\Lambda)$. If Λ sits on $\tilde{P}^{(1)},\tilde{P}^{(2)},..,\tilde{P}^{(\iota)}$

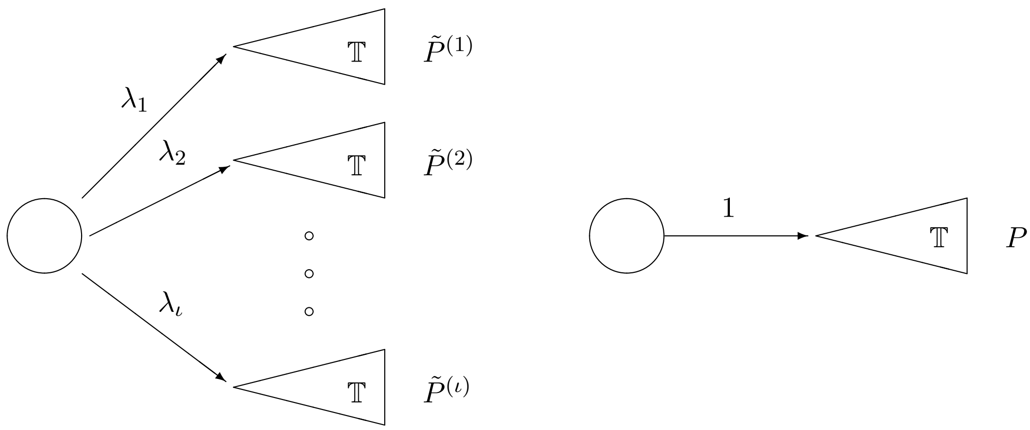


Figure 2.3.1: The compound convex structure of trees $\mathbb{P}(\mathbb{T}, \tilde{P}^{(\iota)})$ and augmented tree $\mathbb{P}_+(\mathbb{T}, P)$

with probabilities λ_n for $1 \leq n \leq \iota$, then compound model has ι nodes at stage 1 and to the ι-th node of stage 1 the subtree $\mathbb{P}(\mathbb{T}, \tilde{P}^{(\iota)})$ is associated, i.e.

$$\mathcal{C}(\mathbb{P}(\mathbb{T}, \cdot), \Lambda) = \sum_{n=1}^{\iota} \lambda_n \mathbb{P}(\mathbb{T}, \tilde{P}^{(n)})$$

where the convex combination $\sum_{n=1}^{\iota} \lambda_n \mathbb{P}(\mathbb{T}, P^{(n)})$ is in the sense of compounding. The convex hull of the set $\mathcal{P}_\epsilon$, is therefore set

$$\bar{\mathcal{P}}_\epsilon = \{\mathcal{C}(\mathbb{P}(\mathbb{T}, \cdot), \Lambda) : \ \Lambda \text{ is a probability measure on } \mathcal{B}_\epsilon\}. \tag{2.3.7}$$

This convention, convexifies problem (2.3.2) which is rewritten to:

$$\min_{x \in \mathbb{X}} \max_{\tilde{\mathbb{P}} \in \bar{\mathcal{P}}_\epsilon} \{\mathbb{E}_{\tilde{\mathbb{P}}}[H(x, \xi)] \text{ s.t. } x \lhd \mathfrak{F}, \ (\Omega, \mathfrak{F}, \tilde{P}; \xi) \sim \tilde{\mathbb{P}}\}. \tag{2.3.8}$$

Notice that in the formulation (2.3.8) the decision variables x must coincide in all randomly sampled subproblems, cf. Figure 2.3.1. Hence by safeguarding ourselves against any random selection of elements of $\mathcal{B}_\epsilon$, we automatically safeguard ourselves against the worst case in $\mathcal{B}_\epsilon$. Now we show that set $\bar{\mathcal{P}}_\epsilon$ is compact which implies the fulfillment of the assumptions of minimax theorem and hence the existence of saddle point solution. In this regard, we show that the nested distance between elements of $\bar{\mathcal{P}}_\epsilon$ namely $\mathrm{dl}(\mathbb{P}(\mathbb{T},P),\mathbb{P}(\mathbb{T},\tilde{P}))$ are bounded from below and above.

For two leaves i resp. j of the tree $\mathbb{T}$ the distance is defined as the distance of the corresponding paths leading to i resp. j, i.e.,

$$\mathsf{d}(i,j)=\sum_{t=1}^{T}\sum_{m=1}^{M} w_t^m|\xi_t^m(i)-\xi_t^m(j)|$$

Assume that no two distinct scenario paths $\xi(i)$ and $\xi(j)$ are identical, [8] i.e., for all $i\neq j$, there exist constants c, $C>0$ such that $c\leq \mathsf{d}(i,j)\leq C$. Let

$$\left\|P-\tilde{P}\right\|=\sum_{i\in\mathcal{N}_T}\left|P_i-\tilde{P}_i\right|=2-2\sum_{i\in\mathcal{N}_T}\min(P_i,\tilde{P}_i).$$

Since an optimal transportation plan can transport a mass of $\min(P_i,\tilde{P}_i)$ from i to i with distance 0. Thus only the masses $1-\sum_{i\in\mathcal{N}_T}\min(P_i,\tilde{P}_i)$ have to be transported, over distances which lie between c and C, hence

$$\frac{c}{2}\cdot\left\|P-\tilde{P}\right\|\leq \mathrm{dl}(\mathbb{P}(\mathbb{T},P),\mathbb{P}(\mathbb{T},\tilde{P}))\leq\frac{C}{2}\cdot\left\|P-\tilde{P}\right\|. \tag{2.3.9}$$

Notice well that the use of the distance $\left\|P-\tilde{P}\right\|$ is only to facilitate the demonstration of compactness. While relation (2.3.9) implies that the topologies generated by the two metrics $\left\|P-\tilde{P}\right\|$ and $\mathrm{dl}(\mathbb{P}(\mathbb{T},P),\mathbb{P}(\mathbb{T},\tilde{P}))$ are the same, balls are quite different in the two metrics and only the latter metric, nested distance, is appropriate for nested distributions. Since $\bar{\mathcal{P}}_\epsilon$ is the continuous image of the set of all probability measures on $\mathcal{B}_\epsilon$, it is compact. Because $\mathcal{B}_\epsilon$ itself is compact and set of all probability measures on $\mathcal{B}_\epsilon$ is compact too. Thus all conditions for the validity of the minimax theorem are fulfilled and a saddle point $(x^*,\mathbb{P}(\mathbb{T},\Lambda^*))$ must exist. Now we prove the equation

$$\mathrm{dl}(\mathbb{P}(\mathbb{T},\Lambda),\mathbb{P}_+(\mathbb{T},P))=\int \mathrm{dl}(\mathbb{P}(\mathbb{T},\tilde{P}),\mathbb{P}(\mathbb{T},P))\ \Lambda(\mathrm{d}\tilde{P}). \tag{2.3.10}$$

In order to see this, assume first that Λ is finite, say $\mathbb{P}(\mathbb{T},\Lambda)=\sum_{n=1}^{\iota}\lambda_n\mathbb{P}(\mathbb{T},\tilde{P}^{(n)})$.

[8]This assumption is only required in our setting and not necessarily in general.

Then

$$\begin{aligned}\mathrm{dl}(\mathbb{P}(\mathbb{T},\Lambda),\mathbb{P}_+(\mathbb{T},P)) \quad &= \mathrm{dl}(\sum_{n=1}^{\iota}\lambda_n\mathbb{P}(\mathbb{T},\tilde{P}^{(n)}),\mathbb{P}_+(\mathbb{T},P))\\ &= \sum_{n=1}^{\iota}\lambda_n[\mathrm{dl}(\mathbb{P}(\mathbb{T},\tilde{P}^{(n)}),\mathbb{P}(\mathbb{T},P))].\end{aligned}$$

If Λ is not finite, it can be approximated by finite measures and therefore the relation (2.3.10) holds in general. The last argument if the theorem is to prove that at saddle point solution, worst case model $\tilde{\mathbb{P}}^*$ happens at a single tree and not a mixture of trees, which means that $\tilde{\mathbb{P}}^*$ belongs to $\mathcal{P}_\epsilon$ and not only to $\bar{\mathcal{P}}_\epsilon$. Let x^* be the minimax decision, i.e.

$$\mathbb{E}_{\tilde{\mathbb{P}}}[H(x^*,\xi)] \leq \mathbb{E}_{\tilde{\mathbb{P}}^*}[H(x^*,\xi)] \leq \mathbb{E}_{\tilde{\mathbb{P}}^*}[H(x,\xi)].$$

Let the saddle point model be $\tilde{\mathbb{P}}^* = \mathbb{P}(\mathbb{T},\Lambda^*)$. The support of Λ^* is closed (hence compact) and the continuous function $\tilde{P} \mapsto \mathbb{E}_{\mathbb{P}(\mathbb{T},\tilde{P})}[H(x^*,\xi)]$ takes its maximum at some distribution $\tilde{P}^*$. Since $\mathrm{dl}(\mathbb{P}(\mathbb{T},\tilde{P}^*),\mathbb{P}(\mathbb{T},P)) \leq \epsilon$ by construction, $\mathbb{P}(\mathbb{T},\tilde{P}^*) \in \mathcal{P}_\epsilon$ and therefore $\mathbb{E}_{\mathbb{P}(\mathbb{T},\tilde{P}^*)}[H(x^*,\xi)] \leq \mathbb{E}_{\tilde{\mathbb{P}}^*}[H(x^*,\xi)]$. On the other hand,

$$\mathbb{E}_{\tilde{\mathbb{P}}^*}[H(x^*,\xi)] = \int \mathbb{E}_{\mathbb{P}(\mathbb{T},\tilde{P})}[H(x^*,\xi)]\,\mathrm{d}\Lambda(\tilde{P}) \leq \mathbb{E}_{\mathbb{P}(\mathbb{T},\tilde{P}^*)}[H(x^*,\xi)].$$

Consequently, $\mathbb{E}_{\mathbb{P}(\mathbb{T},\tilde{P}^*)}[H(x^*,\xi)] = \mathbb{E}_{\tilde{\mathbb{P}}^*}[H(x^*,\xi)]$, which shows that the saddle point model can be chosen from $\mathcal{P}_\epsilon$. This concludes the proof. □

Summary

In decision making under uncertainty, the significance of incomplete knowledge about the underlying probability model and its corresponding impacts specifically on class of real world problems is inevitable. To this end, in this chapter, prior works and existing literature on how to treat this problem were studied. Interestingly, this type of problems has attract attention since more than half a century ago. It was seen that majority of research in this area has been concentrated on two-stage stochastic optimization problems and parametric or non-parametric approaches to construct ambiguity neighborhoods of probability measures have been proposed.

Our approach, however, is to develop a new theoretical foundation and algorithmic approach to deal with multistage stochastic optimization problems under ambiguity. We started with the general structure of risk neutral multistage stochastic optimization problems and elaborated its ambiguity extension. It was seen that the ambiguity problem in its general formulation is computationally intractable, therefore, by fixing the tree structure and scenario values, we presented a non-parametric approach for construction of *smaller* ambiguity neighborhoods. These neighborhoods are constructed around the baseline tree by using the notion of nested distance. The main theoretical result of this chapter is our version of multistage minimax theorem and hence the existence of the saddle point solution.

The agenda in the next chapter is to look at the problem from the algorithmic perspective and to construct nested neighborhoods for dynamic trees together with obtaining the convergence results for the proposed algorithms.

Chapter 3

Algorithms to Compute the Saddle Point of DRSMO

In this chapter we first provide an overview of existing classical saddle point computations on the basis of gradient algorithms. Next, having proved the existence of a saddle point for the distributionally robust multistage stochastic optimization, in Chapter 2, we introduce the corresponding algorithmic approach and prove the convergence. For illustration, the algorithm is performed on a multiperiod production-inventory control problem with uncertain demand and numerical results and sensitivity of decision scenarios under model ambiguity are presented. It should be mentioned that partial content of this chapter has been initially presented in Analui and Pflug [2].

3.1 Classical Saddle Point Algorithms

To begin with, we only look at the general saddle point problem rather than any specific. Consider $f(x, y)$, where f is convex in $x \in X$ and concave in $y \in Y$, a saddle point solution (x^*, y^*) is a tool for decision makers to evaluate the optimal response to the worst strategy. It can be interpreted as an equilibrium, deviation from which is beneficial neither for the decision maker nor for her opponent which is usually the nature. Classical methods based on the gradient or sub gradient for solving the saddle point problems have been of great interest since the seminal work of Arrow et al. [3]. In its classical setting, several algorithms have been proposed. Consider the unconstrained problem

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} \{f(x, y)\}. \tag{3.1.1}$$

Under convex-concavity assumptions of f in problem (3.1.1), for a given x, $f(x,y)$ has an unconstrained maximizer with respect to y and for a given y, an unconstrained minimizer with respect to x. A necessary and sufficient condition for a joint optimum is satisfied by $\zeta^* = (x^*, y^*)$ which solves the the simultaneous system of equations: $\mathcal{E}(\zeta) \equiv \begin{bmatrix} \nabla_x f(x,y) \\ -\nabla_y f(x,y) \end{bmatrix} = 0$. Sometimes it is even more convenient to solve the problem

$$\min_{\zeta} \{\frac{1}{2} \|\mathcal{E}(\zeta)\|_2^2\} \tag{3.1.2}$$

rather than $\mathcal{E}(\zeta) = 0$, Rustem and Howe [47]. For solving (3.1.2), an iterative algorithm follows the sequence:

$$\zeta_{\kappa+1} = \zeta_\kappa + \alpha_\kappa d_\kappa, \tag{3.1.3}$$

where d_κ is the direction of the progress and α_κ is the stepsize. However, exact arguments about the starting point ζ_0 are not addressed.

Demynov and Pevnyi [12] and Danilin and Panin [10] proposed a gradient based algorithm for unconstrained problem (3.1.1) based on direction d_κ and step size strategy α_κ such that sufficient progress at each iteration is ensured. Besides, we refer to Rustem and Howe [47] in which more saddle point computation algorithms, based on gradient information: $\nabla_\zeta \mathcal{E}_\kappa^{\mathrm{T}} \mathcal{E}_\kappa$ or an approximation to it, are presented and discussed. For example: quadratic approximation algorithm for constrained problems based on Qi and Sun [40] approach, interior point saddle point algorithm for constrained problems as elaborated by Sasai [49] and finally a quasi-Newton algorithm for non-linear systems.

3.2 Stage-wise Construction of Ambiguity Neighborhoods

Regarding the DPMSO problems, the aim of section is to elaborate an algorithm for constructing the ambiguity neighborhood around a baseline model $\mathbb{P}(\mathbb{T}, P)$ in the form of nested balls. We have seen that in its general form, problem (2.3.5) has a complex structure. Therefore in construction of models $\mathbb{P}(\mathbb{T}, \tilde{P})$ only scenario probabilities differ from the baseline model $\mathbb{P}(\mathbb{T}, P)$ as long as the respective nested distance remains small. However, the measurability of decisions x w.r.t $\mathfrak{F}$, i.e., $x \lhd \mathfrak{F}$ ensures the comparability of the decisions of both models (2.2.1) and (2.3.5).

In order to describe the nested distance formulation (2.2.4) in a recursive form, we recall the notion of transportation subplans from Section 2.2. A transportation subplan indexed with pair node (k,l) transports the elements of k_+, the set of direct successors of k, into l_+, the set of direct successors of l and must satisfy the following marginal constraints, $i,j \in \mathcal{N}_T$,

$$\begin{aligned} \sum_{j\in l+} \pi(i,j|k,l) &= Q(i) \\ \sum_{i\in k+} \pi(i,j|k,l) &= \tilde{Q}(j) \end{aligned} \tag{3.2.1}$$

From $P_i = Q(i) \cdot Q(pred_{T-1}(i)) \cdots Q(pred_1(i))$, $\tilde{P}_j = \tilde{Q}(j) \cdot \tilde{Q}(pred_{T-1}(j)) \cdots \tilde{Q}(pred_1(j))$ and satisfying constraints (3.2.1), all these subplans are *concatenated* to the full transportation plan $\pi(i,j)$:

$$\begin{aligned} \pi(i,j) \quad = \quad & \pi(pred_1(i), pred_1(j)|1,1) \cdots \\ & \cdot\pi(pred_{T-1}(i), pred_{T-1}(j)|pred_{T-2}(i), pred_{T-2}(j)) \\ & \cdot\pi(i,j|pred_{T-1}(i), pred_{T-1}(j)). \end{aligned} \tag{3.2.2}$$

In Figure 3.2.1 the schematic form of concatenation is shown. Emphasizing again that considering *only* the cases where the alternative models are based on the same tree $\mathbb{T}$, implies that only the probabilities vary between these close (in the nested distance sense) models. In this case, we would rather use the notion of transportation *subkernels* instead of transportation subplans. For arbitrary nodes $k,l \in \mathcal{N}_t$ and $k \prec i$ the subkernel is the probability distribution in the set $l_+(j \succ l)$ such that

$$K_t(j|i;k,l) \geq 0, \qquad \sum_{j\succ l} K_t(j|i;k,l) = 1, \;\; (\forall \; (i,j) \in \mathcal{N}_{t+1} \; i \succ k, \; l)$$

Where $K_t(j|i;k,l) = \frac{\pi(i,j|k,l)}{\sum\limits_j \pi(i,j|k,l)}$. The relation between transportation subkernels and transportation subplans is given by:

$$\begin{aligned} \pi(i,j) \quad = \quad & K_1(pred_1(j)|pred_1(i);1,1) \cdots \\ & \cdot K_{T-2}(pred_{T-1}(j)|pred_{T-1}(i); pred_{T-2}(i), pred_{T-2}(j)) \\ & \cdot K_{T-1}(j|i; pred_{T-1}(i), pred_{T-1}(j)) \times Q(i) \cdot Q(pred_{T-1}(i)) \times \cdots Q(pred_1(i)) \end{aligned} \tag{3.2.3}$$

Therefore transportation kernel $K(j|i)$ is the *composition* of subkernels K_t, for $t =$

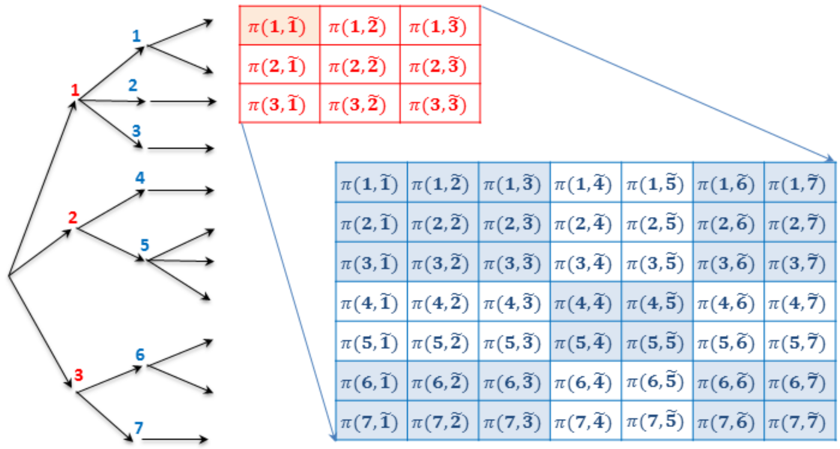


Figure 3.2.1: Concatenation structure between stages 1 and 2. Conditional on stage t (red), eligible transportation subplans of stage $t+1$ (blue) are composed. Notice that $\tilde{\cdot}$ only symbolizes the eligible transportation destinations. Since in our setting, the filtration (tree structure) remains unchanged, $\tilde{\cdot}$ is eliminated form the notation of the proofs in this chapter.

$1, ..., T-1$:

$$\begin{aligned} K(j|i) &= K_1 \circ \cdots \circ K_{T-1}(j|i) \\ &= K_1(pred_1(j)|pred_1(i); 1, 1) \\ &\quad \cdot K_{T-2}(pred_{T-1}(j)|pred_{T-1}(i); pred_{T-2}(i), pred_{T-2}(j)) \\ &\quad \cdot K_{T-1}(j|i; pred_{T-1}(i), pred_{T-1}(j)). \end{aligned} \tag{3.2.4}$$

For a given baseline probability distribution $P = (P_i)_{i \in \mathcal{N}_T}$ we shall define the new probability distribution $\tilde{P}$ by $\tilde{P}_j = \sum\limits_{i,j \in \mathcal{N}_T} P_i \cdot K(j|i)$, hence $\tilde{P} = P \circ K = P \circ (K_1 \circ \cdots \circ K_{T-1})$. Then problem (2.3.2) can be written in the form

$$\min_{x \in \mathbb{X}} \max_{K} \{\mathbb{E}_{P \circ K}[H(x, \xi)] \text{ s.t. } K = K_1 \circ ... \circ K_{T-1}, \sum_{i,j \in \mathcal{N}_T} \mathsf{d}(i,j) \cdot P_i \cdot K(j|i) \le \epsilon\}. \tag{3.2.5}$$

It is noticeable that expression $\sum\limits_{i,j \in \mathcal{N}_T} \mathsf{d}(i,j) \cdot P_i \cdot K(j|i) \le \epsilon$ in (3.2.5) is multi-linear in transportation subkernels $K_1, ..., K_{T-1}$, hence non-convex. Recalling the notations in Section 2.2, If $j \in \mathcal{N}_t$ and $l \in \mathcal{N}_{t-1}$ with $i \prec j$, then probabilities $Q(j) = \frac{P(j)}{P(j_-)} = \frac{P(j)}{P(l)}$ sitting on the arcs represent the conditional probabilities of reaching node j from node i. In applications for ambiguity problem, P and Q are fixed and $\tilde{P}$ and $\tilde{Q}$, regarded as *worst tree* candidates, are constructed with the feasible corresponding transportation subkernels such that $\mathbb{dl}(P, P \circ K) \leq \epsilon$. Optimization over these subkernels is done stage-by-stage repeatedly by optimizing (for fixed decisions) the subkernels at each stage. Algorithmically, this procedure is done by successive linear optimization. This algorithm for multistage stochastic optimization problems has been implemented and the results are being analyzed in the subsequent sections.

3.3 Stochastic Successive Programing

In distributionally robust multiperiod stochastic setting, the algorithm should be tailored in order to fit the complex structure of ambiguity sets and at the same time guarantee the convergence to the equilibrium strategy. Due to the dissimilarity between decision space $\mathbb{X}$ and model space $\mathcal{P}_\epsilon$ in our setting, gradient based algorithms are avoided. In addition, a direct coordinate-wise approach i.e., moving in one step from decisions' space to models' space is subject to oscillation.[1]

For the problem at hand, the criterion function is $F(x, \mathbb{P}) = \mathbb{E}_{\mathbb{P}}[H(x, \xi)]$. We iteratively find the saddle point by stage-wise approximating the ambiguity set $\mathcal{P}_\epsilon$ by a finite set. In particular, the following variant (3.3.1) is proposed and the proof of convergence is discussed.

$$\begin{cases} x^{\kappa+1} \in \arg\min_{x \in \mathbb{X}} \max_{1 \leq l \leq \kappa} \ F(x,\ \tilde{\mathbb{P}}^l) \\ \tilde{\mathbb{P}}^{\kappa+1} \in \arg\max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} F(x^{\kappa+1}, \tilde{\mathbb{P}}) \\ \text{where } \mathcal{P}_\epsilon^{\kappa+1} = \mathcal{P}_\epsilon^{\kappa} \cup \{\tilde{\mathbb{P}}^{\kappa}\} \text{ and } \mathcal{P}_\epsilon^0 = \{\mathbb{P}\} \end{cases} \tag{3.3.1}$$

In the following, the iterative manner of solution procedure is shown in Algorithm 3.1. At each iteration a new model $\mathbb{P}(\mathbb{T}, \tilde{P})$, (in short: $\tilde{\mathbb{P}}$), which is in ϵ nested distance of baseline model $\mathbb{P}(\mathbb{T}, P)$, (in short: $\mathbb{P}$), is included in the ambiguity set and therefore the size of the problem increases at each iteration. Notice that *Step 4.* of Algorithm 3.1 itself encompasses a non-convex optimization and for this reason constructing $\mathbb{P}(\mathbb{T}, \tilde{P})$ which was discussed in Section 3.2 is approximated by a stage-

[1]Notice that even under strict convex-concavity and compactness of X and Y the convergence of $\begin{cases} x^{\kappa+1} = \arg\min_{x \in X} f(x, y^\kappa) \\ y^{\kappa+1} = \arg\max_{y \in Y} f(x^{\kappa+1}, y) \end{cases}$ is not guaranteed. (See Appendix A.2 for a counter example.)

Algorithm 3.1 Successive Programming

1. Let $\kappa = 0$ and determine the value of ϵ.
2. Start with the „baseline” model, i.e. $\mathcal{P}_\epsilon^\kappa = \{\mathbb{P}\}$
3. Solve the outer optimization problem:

$$\left\| \begin{array}{ll} \min & u \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[H(x,\xi)] \leq u \quad \text{for all } \mathbb{P} \in \mathcal{P}_\epsilon^\kappa \\ & x \in \mathbb{X}, \\ & x \lhd \mathfrak{F} \end{array} \right. \longmapsto (x^\kappa, u^\kappa)$$

4. Fix x^κ and solve the inner optimization problem:

$$\left\| \begin{array}{ll} \max & \mathbb{E}_{\tilde{\mathbb{P}}}[H(x^\kappa,\xi)] \\ \text{s.t.} & \tilde{\mathbb{P}} \in \mathcal{P}_\epsilon \end{array} \right. \longmapsto (\tilde{\mathbb{P}}^\kappa) \text{ and set } \mathcal{P}_\epsilon^{\kappa+1} = \mathcal{P}_\epsilon^\kappa \cup \{\tilde{\mathbb{P}}^\kappa\}$$

5. Stop if there is no improvement in u^κ, otherwise go to 2.

Note : In practical implementation we might

- choose a stopping criteria θ s.t. $u^{\kappa+1} - u^\kappa \geq \theta$, or
- specify in advance the number of iterations κ, which means that the number of models included in set $\mathcal{P}_\epsilon$ has to be determined at the beginning.

---

wise procedure. This step itself is expanded and shown in Algorithm 3.2. This algorithm shows how the worst tree is constructed by updating the transportation kernel K at each stage t subject to the constraint that the nested distance remains less than ϵ. Recalling the concatenation structure depicted in Figure 3.2.1 and the assumption of fixed tree structure, K is a collection of stage-wise transportation kernels.

In the following proposition we prove the convergence of the iterative procedure (3.3.1). Before stating the proposition, it is required to provide the definition of *cluster point*.

Definition 9. [Schramm [51], Chapter 7, Def. 3] A point s of the set S is a cluster point if $\forall\delta > 0$, $(s-\delta, s+\delta) \cap S$ is infinite. This is equivalent to say that the intersection of every neighborhood of s with S to be infinite. Sometimes it is also convenient to show that $\forall\delta > 0$, $(s-\delta, s+\delta) \setminus \{s\} \cap S \neq \emptyset$.

Proposition 10. *Let $\mathbb{X}$ and $\bar{\mathcal{P}}_\epsilon$ be compact sets and $(x,\mathbb{P}) \mapsto F(x,\mathbb{P}) = \mathbb{E}_{\mathbb{P}}[H(x,\xi)]$ be jointly continuous, then every cluster point of the iteration given by (3.3.1) is a*

Algorithm 3.2 Stage-wise construction of the worst tree

- Calculate $n \times n$ distance matrix $\mathsf{d}(\omega, \tilde{\omega}) = \sum_{t=1}^{T} \sum_{m=1}^{M} w_t^m \, \|\xi_t^m(\omega) - \xi_t^m(\tilde{\omega})\|_1$. Recalling the problem (3.2.5), we seek to find $P \circ K = \tilde{P} = (\tilde{P}_1, ..., \ \tilde{P}_n)$ such that $\mathrm{dl}(P, \tilde{P}) \leq \epsilon$,
  - $\tilde{P}^{\kappa,[T-1]} = (\tilde{P}_1, ..., \ \tilde{P}_n)^{\kappa,[T-1]}$ is the worst tree corresponding to the κ^{th} iteration of Algorithm 3.1.
- Suppose that we are at κ^{th} iteration of Algorithm 3.1, i.e., x^{κ} is known and fixed.

1. for $t = 1, ..., T-1$
2. $$\left\| \begin{array}{ll} \max & (H(x^{\kappa}, \cdot))^{\mathsf{T}} \tilde{P}^{\kappa,[t]} \\ \text{s.t.} & \mathrm{dl}(\tilde{P}^{\kappa,[t]}, P) \leq \frac{\epsilon}{T-1} \\ \text{where} & \tilde{P}^{\kappa,[t]} = P \circ K_1^{new} \circ \cdots \circ K_{t-1}^{new} \circ K_t \circ K_{t+1}^{old} \circ \cdots \circ K_{T-1}^{old} \end{array} \right.$$
3. Call the solution K_t^{new}
4. end

---

minimax solution.

Proof. To prove the convergence of iterative procedure

$$\begin{cases} x^{\kappa+1} \in \arg\min_{x \in \mathbb{X}} \max_{1 \leq l \leq \kappa} \ F(x, \tilde{\mathbb{P}}^l) \\ \tilde{\mathbb{P}}^{\kappa+1} \in \arg\max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} F(x^{\kappa+1}, \tilde{\mathbb{P}}) \end{cases},$$

denote by $F^{\kappa} = \max_{1 \leq l \leq \kappa} F(x^{\kappa+1}, \tilde{\mathbb{P}}^l)$, then $F^{\kappa+1} = \max_{1 \leq l \leq \kappa+1} F(x^{\kappa+2}, \tilde{\mathbb{P}}^l)$ and by monotonicity $F^{\kappa+1} \geq F^{\kappa}$. Since the function F is bounded, F^{κ} converges to $F^* := \sup F^{\kappa}$. Moreover, by $\mathbb{X}$ being compact, the sequence x^{κ} has one or several cluster points. Let x^* be such a cluster point. We show that $F^* = \max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} F(x^*, \tilde{\mathbb{P}})$. Since always $F^* \leq \max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} F(x^*, \tilde{\mathbb{P}})$, suppose that $F^* < \max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} F(x^*, \tilde{\mathbb{P}})$. Then there must exist a $\tilde{\mathbb{P}}^+$ such that $F(x^*, \tilde{\mathbb{P}}^+) > F^*$. By continuity this inequality must then hold in a neighborhood of x^* and therefore there must exist a x^{κ} for which the same inequality holds. However, this contradicts the construction of the iterative procedure. Finally, we show that $x^* \in \arg\min_{x \in \mathbb{X}} \max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} \ F(x, \tilde{\mathbb{P}})$. If not, there must exist a x^+ such that $\max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} F(x^+, \tilde{\mathbb{P}}) < \max_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} F(x^*, \tilde{\mathbb{P}})$. By construction, we have $\max_{1\leq l\leq\kappa}\ F(x^+,\,\tilde{\mathbb{P}}^l)\geq\max_{1\leq l\leq\kappa}\ F(x^{\kappa+1},\,\tilde{\mathbb{P}}^l)=F^\kappa$ and letting κ tend to infinity, one sees that $\max_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} F(x^+,\,\tilde{\mathbb{P}})\geq F^*=\max_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} F(x^*,\,\tilde{\mathbb{P}})$ and this is a contradiction which shows that x^* is the cluster point and thus every cluster point is a solution of the minimax problem. □

In the following section the saddle point computation algorithms which were introduced in Section 3.3 are implemented and corresponding results for a classical multiperiod production-inventory control problem are presented.

3.4 Multiperiod Production-Inventory Control with Uncertain Demand

In this section first we formulate a small multistage stochastic optimization problem[2] then present and discuss the numerical results for the optimal and distributionally robust solutions from Algorithms 3.1 and 3.2. In the considered problem the production volume of two products, we name them product_1 and product_2, has to be decided while *maximizing*[3] the expected net profit. The net profit is derived from selling the products under a two-dimensional stochastic demand with fixed selling prices and production, inventory and external supply costs. Deciding on how much of each product types to produce during a particular week forms the decision variables. The production machine is designed to produce both product types and there is an overall production capacity. The full mathematical formulation of this problem is composed in (3.4.1). Stochastic demand is characterized in terms of scenarios and a tree terminology is used to describe event probabilities and multistage scenarios. The demand scenarios are represented on a *binary* tree with not necessarily equal event probabilities. In Figure 3.4.1, the tree structure and demand requirements of both products is depicted. In Figure 3.4.2, to emphasize on distinct demand levels and showing that demand of product_2 exhibits a higher level, demand scenarios for both products and the corresponding scenario probabilities are shown in separate graphics.

[2] The numerical example is taken from AIMMS optimization modeling (Bisschop [6], Chapter 17). However, all computational procedures, solution algorithms and results analysis are implemented in MATLAB R2012a.

[3] In this example, DR counterpart is therefore formulated as a *maximin* problem.

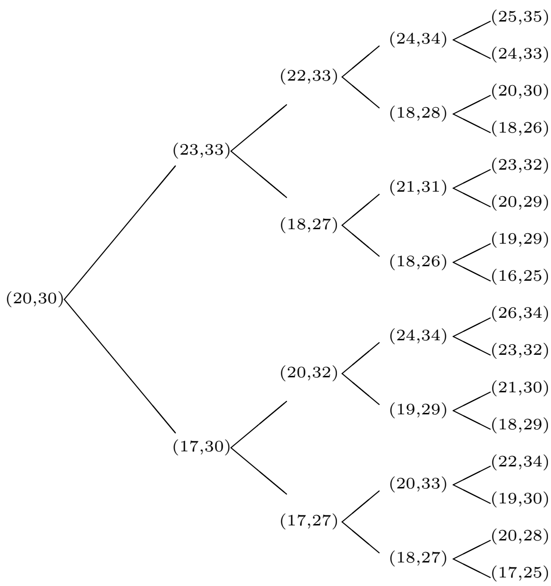


Figure 3.4.1: Demand scenarios (product_1,product_2) and the binary tree structure

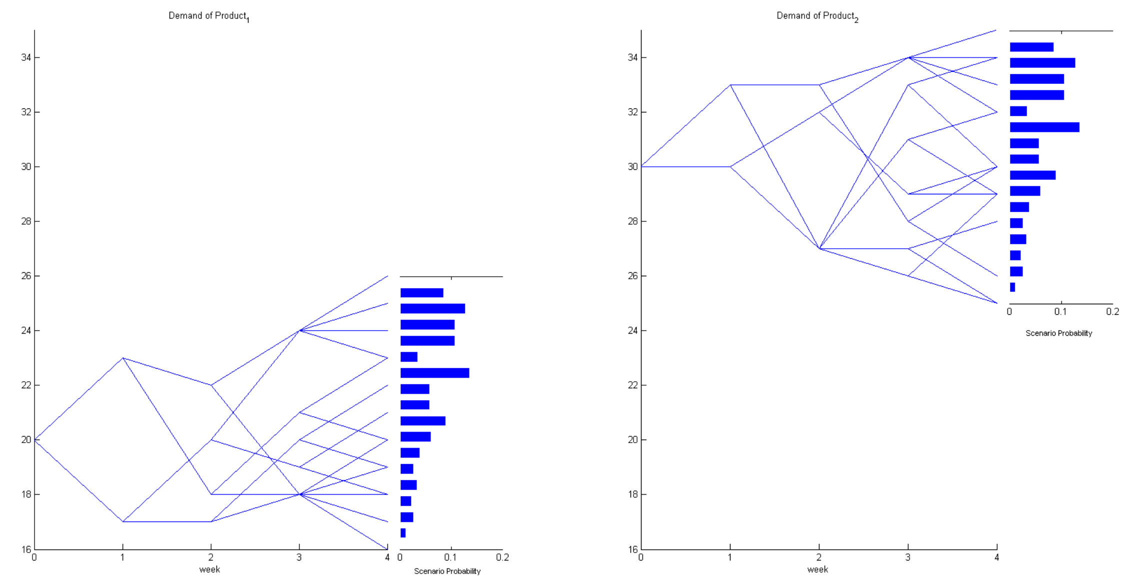


Figure 3.4.2: Demand scenarios of product_1 and product_2 and the corresponding scenario probabilities.

3.4.1 Modeling and Implementation

In Tables 3.4.1 and 3.4.2, the symbols defining the parameters, decisions and decision dependent variables of problem (3.4.1) are introduced.

Table 3.4.1: Production-inventory model nomenclature

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 $b = 1, 2$
Random Process	
d^b	Demand for product $b = 1, 2$
Decision Variable	
x_f^b	Production volume of product b for $b = 1, 2$
Decision Expressions	
x_i^b	Inventory level of each product $b = 1, 2$
x_e^b	External supply of each product $b = 1, 2$
v	Profit

Table 3.4.2: Parameters : revenues, costs and capacities

Product	pr^b(€/unit)	pc^b(€/unit)	ic^b(€/unit)	ec^b(€/unit)	$init_b$	$\bar{x}_i$	c
product_1	300	12	5	195	17	52	46
product_2	400	10	5	200	35		

Mathematical modeling

Objective Function

The objective of this production-inventory control model is to maximize the total expected net profit : $\sum\limits_{n} P(n)v(n)$, $P(n)$ is the unconditional probability of reaching node $n \in \mathcal{N} = \bigcup_{t=0}^{T} \mathcal{N}_t$, and $\mathcal{N}_t$ is the set of nodes at stage t, under the following constraints:

Constraints

Production capacity constraint ensures that the total production volume is bounded above with the overall capacity c.

$$\sum_b x_f^b(n_-) \leq c \qquad \forall n \in \mathcal{N} \setminus \mathcal{N}_0$$

Inventory determination constraint states that the inventory is determined at each reachable node by the inventory at the predecessor node plus the production volume at the predecessor node plus the external supply at that not minus the demand pertaining to that node for $b = 1, 2$.

$$x_i^b(n_-) + x_f^b(n_-) + x_e^b(n) - d^b(n) = x_i^b(n) \qquad \forall n \in \mathcal{N} \setminus \mathcal{N}_0$$

Inventory capacity constraint ensures that the inventory level for both products is bounded above with the maximum inventory capacity $\bar{x}_i$.

$$\sum_b x_i^b(n) \leq \bar{x}_i \qquad \forall n \in \mathcal{N}$$

Demand requirement constraint determines that the stochastic demand of both products is met at each node via this inequality since because of technicality issues, the produced product of the current node can not be used to satisfy the demand at that node, for $b = 1, 2$.

$$x_i^b(n_-) + x_e^b(n) \geq d^b(n) \qquad \forall n \in \mathcal{N} \setminus \mathcal{N}_0$$

Profit determination constraint is an accounting equation for the net profit position at each node which is derived from the sales revenue minus the total costs consisting of production, inventory and external supply.

$$\sum_b pr^b d^b(n) - \sum_b [pc^b x_f^b(n_-) + ic^b x_i^b(n) + ec^b x_e^b(n)] = v(n) \qquad \forall n \in \mathcal{N} \setminus \mathcal{N}_0$$

The full mathematical model in node representation is found in (3.4.1). Note that decisions are only defined for emanating nodes and thus not for leaf (terminating) nodes.

In the next section the maximin solution of distributionally robust extension of (3.4.1) is presented and discussed. Distributionally robust extension of this example seeks for equilibrium strategies that ensure the maximum expected net profit under the most adverse demand scenarios.

$$
\begin{aligned}
\max\ & \sum_n P(n)v(n) && && (3.4.1)\\
\text{s.t. } & \sum_b x_f^b(n_-) \leq c && \forall n \in \mathcal{N} \setminus \mathcal{N}_0\\
& x_i^b(n_-) + x_f^b(n_-) + x_e^b(n) - d^b(n) = x_i^b(n) && \forall b\ \forall n \in \mathcal{N} \setminus \mathcal{N}_0\\
& \sum_b x_i^b(n) \leq \bar{x}_i && \forall n \in \mathcal{N}\\
& x_i^b(n_-) + x_e^b(n) \geq d^b(n) && \forall b\ \forall n \in \mathcal{N} \setminus \mathcal{N}_0\\
& \sum_b pr^b d^b(n) - \sum_b [pc^b x_f^b(n_-) + ic^b x_i^b(n) + ec^b x_e^b(n)] = v(n) && \forall n \in \mathcal{N}\\
& x_e^b,\ x_i^b \in \mathbb{R}_+^{|\mathcal{N}|} && \forall b\\
& x_f^b \in \mathbb{R}_+^{|\mathcal{N} \setminus \mathcal{N}_T|} && \forall b\\
& v \in \mathbb{R}^{|\mathcal{N}|}
\end{aligned}
$$

3.4.2 Maximin Solution Summary

Problem (3.4.1) is solved using Algorithms 3.1 and 3.2 for an increasing sequences of ϵ ambiguity radii. Moreover for each ϵ, $\kappa = 4$, which means the ambiguity set $\mathcal{P}_\epsilon$ includes four varied models.

Worst Tree Construction

As it is expected by increasing the ambiguity radius the largest probability would be associated to a scenario which for given decisions (cf. Algorithm 3.2), has the worst outcome. As ϵ increases the worst tree turns out to be less and less complex. The ambiguity sets are constructed for $\epsilon = 1, 6, 11, 16$ (see Figure 3.4.3). The reason for choosing the values of ϵ is simply ranging it between $[\min \mathsf{d}(i,j),\ \max \mathsf{d}(i,j)]$, where $\mathsf{d}(i,j)$, as it was defined before, is the distance between demand scenarios i and j. Regardless of demand levels for product_1 and product_2 , in Figure 3.4.3, the scenario probabilities of the respective tree structure for increasing ambiguity radii are depicted. It is observed that at the largest radius i.e., $\epsilon = 16$, the remaining scenarios 3 and 4 form the worst tree. More over, it is observed that the stage wise

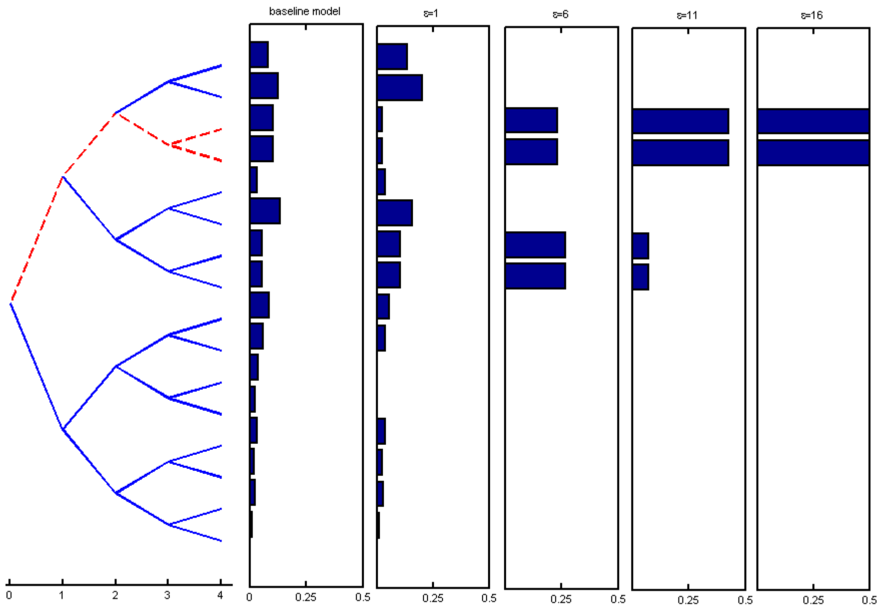


Figure 3.4.3: Tree structure and diminishing worst trees for increasing ambiguity radii

procedure of constructing the worst tree is not monotonic.

Sensitivity of Decision Scenarios

In Figure 3.4.5, the maximin solution of production scenarios for product$_1$ and its sensitivity with respect to increasing ambiguity radius is shown. It is noticeable that including more than just a baseline demand model, results in more diverse production scenarios. This consequence is observed for product$_1$ by comparing scenario paths in Figures 3.4.4 and 3.4.5. At first this results might seem quite controversial, since worst scenarios are getting a more simple structure as ϵ increases and on the other hand, the decision scenarios are revealed to be more complex. This might be seen as incorporating more models in the neighborhood of baseline, those decisions are taken which are good for all included models. This phenomenon also has an impact on all decision dependent scenarios which in this multiperiod inventory-production control problem are the external purchase x_e^b and inventory level x_i^b for $b = 1, 2$ and consequently on profit scenarios.

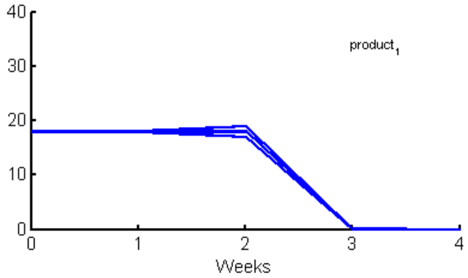


Figure 3.4.4: Optimal solution scenarios under the baseline - product$_1$

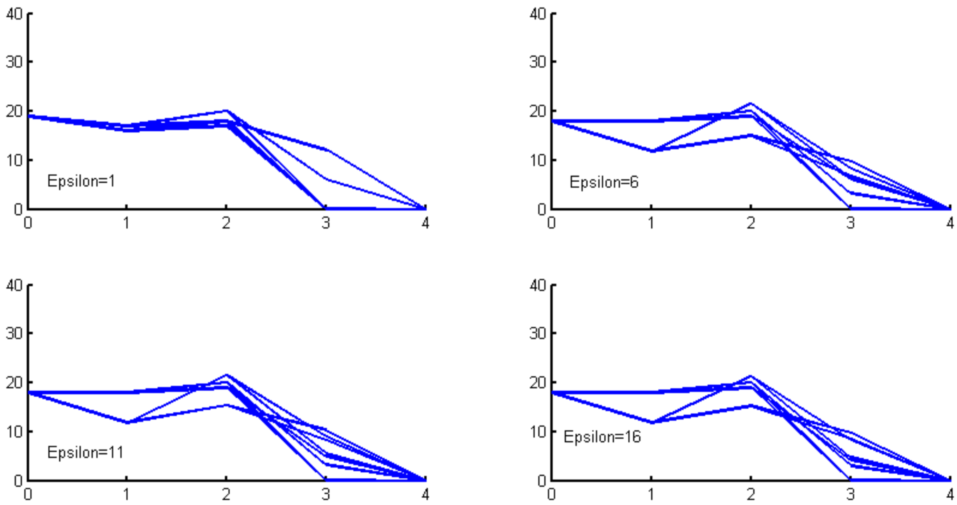


Figure 3.4.5: Sensitivity of decisions under model ambiguity - product$_1$

***Price* for *Gaining* Distributionally Robust Solution**

It is important to make a remark here, since in this example the objective function is to maximize the expected net profit, and to avoid confusion with the cost function H in the minimax setting in previous chapters, we consider $\bar{H}$ as the respective of concave profit function.

A fundamental result in maximin setting of this example is that at solution $(x^*, \tilde{\mathbb{P}}^*)$ we have the following well known saddle point inequality (cf. Rockafellar [45]):

$$\mathbb{E}_{\tilde{\mathbb{P}}^*}[\bar{H}(x,\xi)] \leq \mathbb{E}_{\tilde{\mathbb{P}}^*}[\bar{H}(x^*,\xi)] \leq \mathbb{E}_{\mathbb{P}}[\bar{H}(x^*,\xi)]. \tag{3.4.2}$$

This inequality indicates the robust nature of maximin equilibrium in presence of worst cases. On the right hand side, this inequality ensures the improvement of the optimal solution if worst case does not occur, whereas on the left hand side, it shows the potential losses if worst case is realized. To quantify the cost of ambiguity and the reward of robustness we adapt (3.4.2) to the distributionaly robust counterpart of optimization problem (3.4.1). Since problem (3.4.1) is a maximin, it reads:

$$\min_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x^*(\mathbb{P}),\xi)] \leq \max_x \min_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x,\xi)] \leq \mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathcal{P}_\epsilon),\xi)] \leq \mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathbb{P}),\xi)],$$

where $x^*(\mathbb{P})$ is the optimal solution under the baseline model and $x^*(\mathcal{P}_\epsilon)$ is the maximin solution. Figure 3.4.6 delineates this inequality against the increasing ambiguity radii. In Figure 3.4.6, the solid line shows $\mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathbb{P}),\xi)]$ the optimal solution to the original problem i.e., under the baseline model. The dashed line shows $\mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathcal{P}_\epsilon),\xi)]$ i.e, potential decreases in the expected net profit if maximin decisions are taken under the baseline model. The dotted-dashed line represents the distributionally robust or actual maximin solution $\max_x \min_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x,\xi)]$ and finally the dotted line shows the worst case solution which is the optimal solution, $x^*(\mathbb{P})$, under the baseline for the worst model. It practically shows :

- *cost of ambiguity* which is the price that decision maker pays to obtain the distributionally robust solution is calculated by $\mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathbb{P}),\xi)] - \mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathcal{P}_\epsilon),\xi)]$ and causes a 2.26% decrease in the expected net profit at the largest ambiguity neighborhood, and
- *reward for robustness* as $\max_x \min_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x,\xi)] - \min_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x^*(\mathbb{P}),\xi)]$ indicates that decision maker gains 7.03% by cautiously robustifying herself against the

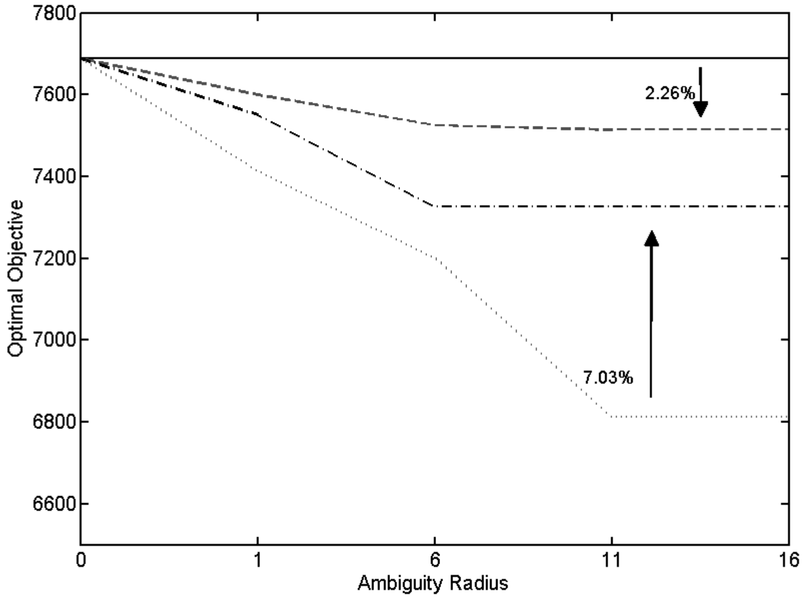


Figure 3.4.6: The price to ***pay*** for ***gaining*** robust solution. The pay is associated to the price for ambiguity and the gain is associated to the reward for robustness.

worst odds.

A closer look shows that for a specific ϵ radius, no more improvement in the maximin solution occurs i.e., there is an upper bound for the *price* to *gain* the robust solution. It becomes evident that eventually we look for the price the decision maker is willing to pay to robustify her optimal decisions against the worst case.

In the subsequent chapters, our central attention remains on finding such an equilibrium for real world class of problems and corresponding behavior of ambiguity radii.

Summary

In this chapter, an iterative algorithm to solve distributionally robust multistage stochastic optimization problem with ambiguity about the underlying probability model is presented. The distributionally robust counterpart of a risk neutral MSO is constructed by the worst case approach with respect to the probability models which are in an ϵ nested neighborhood of a baseline model. Considering fixed scenario values and tree structure, only the underlying probabilities are subject to changes. Proof of convergence of the solution algorithm was also presented. It was observed that when we consider minimax w.r.t the worst case (a bunch of them when $\kappa > 1$), worst case is getting a simpler structure as epsilon increases. Maybe at the largest radius even sits on a single scenario. The minimax decisions, however, have shown to be more complex. The reason for this phenomena might be the inclusion of more models and decisions should be taken which are optimal for all models. Hence it turns out that we have more bushy decision scenarios when the ambiguity radius increases. Moreover, we have quantified the cost of ambiguity and the reward of robustness around the maximin solutions for a small instance of a classical multistage production-inventory control problem. In addition, it is observed that there is a threshold for ϵ range, at which no improvements appear in minimax decisions and corresponding optimal value.

In the next two chapters, this approach is implemented for a real world optimization problem, namely *short term hydroelectric production planning*. Particularly, in the next chapter, mathematical modeling and the topology of a multi reservoir pumped-storage system is introduced. In addition, with the aim for a credible approximation of the baseline model, scenario simulation and tree generation of *electricity spot prices*, *pumping prices* and *natural inflows* are discussed. Further, in chapter 5, we lay the foundations for analysis of minimax solutions by first solving the optimization problem under the baseline, subsequently its distributionally robust counterpart and finally, sensitivity of decisions under model ambiguity for a real world decision problem.

Chapter 4

Target Class of Problems and Preprocessings

"Hydro power could double its contribution by 2050, reaching 2,000 GW of global capacity and over 7,000 TWh. This achievement, driven primarily by the quest of clean electricity, could prevent annual emissions of up to 3 billion tonnes of CO_2 from fossil-fuel plants. The bulk of this growth would come from large plants in emerging economies and developing countries." International Energy Agency 26 (Mo)vember 2013.

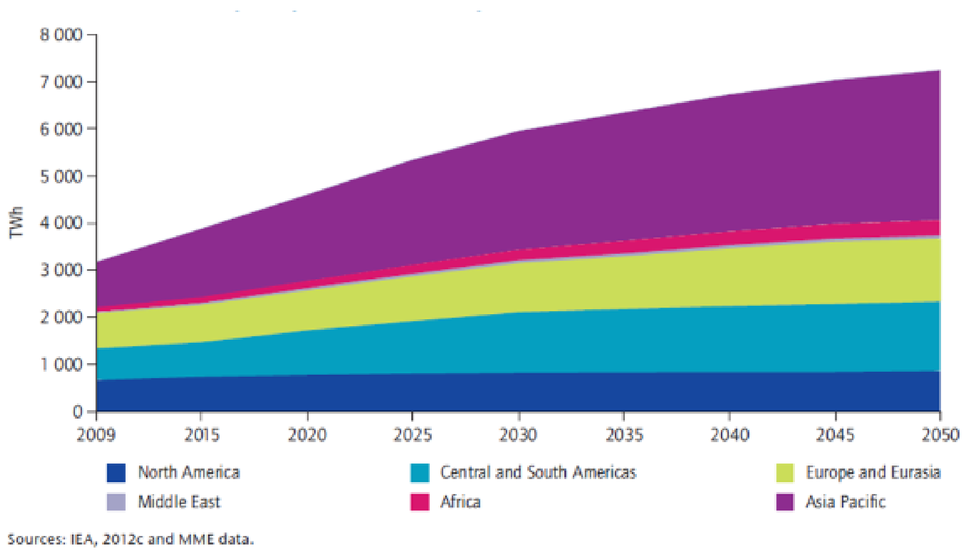


Figure 4.0.1: Hydroelectricity Generation (TWh) till 2050 in the Hydropower Roadmap vision- *International Energy Agency, 26.11.2013.*

The purpose of this chapter is to present a case study for the application of our distributionally robust approach to a real world decision problem. In this problem, the optimal solution and corresponding optimal decisions are vital and important enough to worth a step further and accept that complete knowledge about the underlying probability model is far too unrealistic. Although we start from crude (history) data and apply the modeling of uncertainty in a very fine manner, we would like to emphasize that relying on only one model, even if it is carefully chosen, does not suffice and decision makers should accept to pay a reasonable price to obtain distributionally robust solution and safeguard against worse models. The decision model of which we study the mathematical formulation in this chapter is a short term hydroelectric production planning. Such decision making problems are motivated by increasing number of stochastic factors as a direct result of liberalization in electricity industry. The problem under our investigation is decision making under *market* and *weather* uncertainty and modeling these uncertainties itself is highly subject to ambiguity. To begin with, we introduce a short term hydro electricity operation planning problem. The aggregate system we consider is hypothetical from the production network in Austria.[1] One section of this chapter is also dedicated to introducing stochastic parameters and discussing our scenario simulation strategy. In addition, a novel scenario tree generation algorithm via optimal quantization of stochastic processes is referred and briefly explained. Consequently, the baseline model which is constructed by this algorithm and will be used in Chapter 5 is visualized.

4.1 Hydro Electricity Production

There is simply no efficient, low cost technology for storing electricity in large quantities to meet temporary peaks in demand or unplanned outages of other power plants for a duration of a few hours, days or months. Until the day that such a technology emerges, conventional hydro power pumped storage is one solution. A single pumped-storage station is usually built consisting of an upper reservoir which has little or no inflow and a lower reservoir. Electricity generation in the upper reservoir depends on the potential power stored by pumping up the water from the lower reservoir. Since pumping up the water is costly, pump turbines are planned

[1]In this regard, I would like to thank Mr. Wilfried Grubauer, Mr. Willi Kritscha and Mr. Walter Reinisch from Siemens AG Österreich, for many fruitful discussions and their kind act in providing us with the natural inflow data.

to transfer water during off-peak hours when the price for electricity is at its lowest level.[2] Hydro electricity production/operation planning is a decision problem under uncertainty because of corresponding unknown parameters such as natural inflows to the reservoirs, volatile electricity prices, future demands and even sometimes availability of equipments. It is, in addition, a multistage optimization problem since the decisions have to be taken in a sequential manner. Early attempts to incorporate multistage stochastic optimization for hydro/hydrothermal operation planning were discussed by Pereira and Pinto [34] and Terry et al. [55]. The purpose of the hydro system operation planning is to define an operation strategy which, for each stage of the planning period, given the system state at the beginning of the stage, produces generation targets for each production unit (power plant). Usually the objective function is to minimize the expected total cost along the planning horizon or to maximize the expected total revenue of incorporation in the market so as to meet requirements on the continuity of energy supply and subject to feasibility constraints. In this section for mathematical modeling and implementations we refer to Analui and Kovacevic [1].

A multi-reservoir hydro generation system consists of a cascade of interconnected reservoirs along with corresponding turbines, pumps and spillways. Such a system can be represented by a directed graph (see Figure 4.1.1) where the nodes represent the reservoirs and the arcs represent water flows. Water flows can be either related to power generation by turbines, to pumped water for increasing reservoir storage, or to spillage.[3] Let $\mathcal{J}$ denote the set of reservoirs (nodes) and $\mathcal{I}$ denote the set of water flows (arcs), then the arc-node incidence matrix, whose $ij-$entry, denoted by A_{ij}, represents the interconnections among reservoirs and water flows as follows:

$$A_{ij} = \begin{cases} 1 & \text{water flows into reservoir } j \text{ over are } i \\ -1 & \text{water flows out of reservoir } j \text{ over are i} \\ 0 & \text{arc } i \text{ is not conected into reservoir } j \end{cases}$$

Incidence matrix for the interconnected system shown in Figure 4.1.1, can be found in Appendix A.3.1. The potential decisions are given by the water flows over arcs –

[2]The energy used for pumping the water might also come from other energy sources, including nuclear, fossil and intermittent renewables, that are less efficient to adapt to load fluctuations [Alstom Group].

[3]Spillage is operated via spillways. Spillways are structures built to provide safe release of flood waters from a dam to a downstream (normally the river on which the dam is constructed). [encyclopedia of civil-engineering]

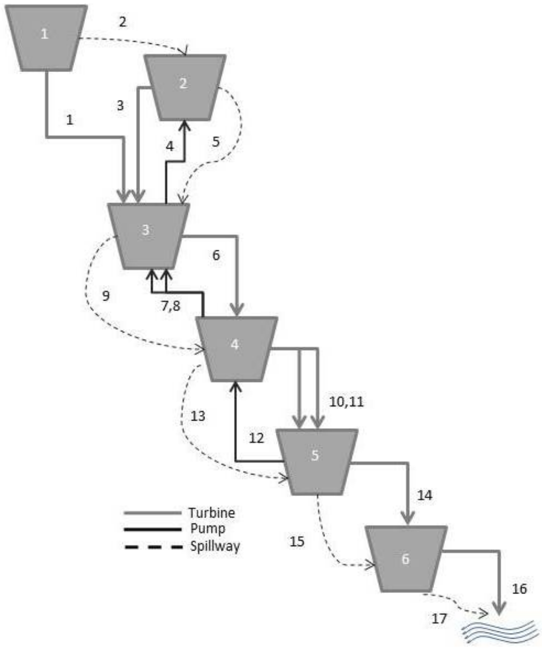


Figure 4.1.1: Hydroelectricity system topology

resulting in electric energy production to be sold at a spot market – and the amount of water stored in reservoirs, in order to maximize expected terminal revenue. All decisions on the amount of water flows and related energy production are made at discrete periods in time $t = 0, ..., T-1$. This leads to a multistage stochastic nonlinear programing problem, where the produced energy within time period of length $\triangle t$ depends on the elevation difference between head-water and tail-water. For small systems this problem can be solved by dynamic programing, but for large systems with many reservoirs (state variables) dynamic programming is not applicable. Often this problem is simplified (maintained convex) by keeping the height, difference between head-water and tail-water, constant. This way the model turns to be a large scale but linear multistage stochastic optimization problem. Throughout this chapter this simplification is applied. In the following, the decision problem of a generator using a hydrosystem with pre specified topology in a multistage stochastic optimization setting is mathematically formulated. In Table 4.1.1 the explanation of symbols and the corresponding units used to represent the data, random process, decision and expression variables are presented. The sources of stochasticity are

Table 4.1.1: Hydroelectricity model nomenclature

Data	
$k^i\,[MWh/10^3m^3]$	Energy coefficient.
$k^i > 0$	- arc i represents generation;
$k^i < 0$	- arc i represents pumping;
$k^i = 0$	- arc i represents spillage;
$\bar{q}^i[10^3m^3/\Delta t]$	Maximum water flow in arc i.
$\bar{v}^j[10^3m^3]$	Maximum storage volume in reservoir j.
$v^j_{end}[10^3m^3]$	Minimum storage volume required in reservoir j at the end of planning horizon
$\underline{v}^j[10^3m^3]$	Minimum storage volume in reservoir j .
$\mathbf{P}_{max}[MW]$	Maximum production capacity for each arc i (= 0 if i represents spillage)
Random Process	
$f^j_t\,[10^3m^3/\Delta t]$	Natural inflow in reservoir j.
$\xi^e_t[Euro/MWh]$	Electricity Spot prices.
$\xi^p_t[Euro/MWh]$	Electricity Pumping price.
Decision Variable	
$q^i_t\,[10^3m^3/\Delta t]$	The water flow on arc i - turbined volume if arc i represents generation; - pumped volume if arc i represents pumping; - spilled volume if arc i represents spillage
Decision Expressions	
$v^j_t[10^3m^3]$	Storage volume in reservoir j.
$x^{ei}_t[MWh]$	Energy produced during the next stage for each arc i.
$x^c_t[Euro]$	Cash

given by:

- The electricity price ξ_t^e at which electricity is sold to the spot market. In a model with weekly decisions this will usually be an average over hourly spot prices, using a typical production schedule;
- The pumping price ξ_t^p i.e. price the decision maker pays to pump up the water into the upper reservoir. With weekly decisions we use an average over off-peak electricity prices;
- Natural inflows to reservoirs f_t^j.

In Section 4.2, statistical analysis and modeling procedure of these scenarios are explained and a co-generated finite scenario tree based on Pflug and Pichler [37] approach [cf. [37], Chapter 4, Algorithms 4.5 & 4.10] as our baseline model to distributionally robust multistage stochastic optimization algorithm is presented.

In line with the notation introduced in Table 4.1.1, in the multistage stochastic optimization problem (4.1.1), f_t^j and prices ξ_t^e, ξ_t^p are random processes while flows through the arcs q_t^i form the decision process and energy produced x_t^{ei}, accumulated cash x_t^c and reservoir's storage level v_t^j are implied by the decision process which are named decision dependent processes. The mathematical model is formulated in the following.

Mathematical modeling

Objective Function

A hydropower producer seeks to maximize the expected terminal revenue $\mathbb{E}[x_T^c]$ under the following constraints:

Constraints

Arc capacity constraints: The lower bounds for water flows through each arc i at time t is set to zero, while the upper bound is equal to the plants' maximum (discharge) capacity.

$$0 \leq q_t^i \leq \overline{q}^i \qquad \forall i \in \mathcal{I},\ t = 0, ..., T-1.$$

Storage level constraints: Lower and upper bounds for storage level of reservoir j at time t is set. These explicit bounds on storage must be assigned for recreation, providing flood control space and assuring minimum levels for dead storage and power plant operation,

$$\underline{v}^j \leq v_t^j \leq \overline{v}^j \quad \forall j \in \mathcal{J},\ t = 0, ..., T-1.$$

Final reservoir content: Minimum content for each reservoir j at the final stage T needs to be fulfilled,

$$v_{end}^j \leq v_T^j \quad \forall j \in \mathcal{J}.$$

Storage balance equation: Water balance for each reservoir j : At each stage, the water level at the end of the period t depends on the water level at the beginning of period t, and the inflows and discharges during the period based on the system topology,

$$v_t^j = v_{t-1}^j + f_t^j + \sum_{\{i \in \mathcal{I} | \mathbf{P}_{max} > 0\}} A_{ij} \cdot q_{t-1}^i + \sum_{\{i \in \mathcal{I} | \mathbf{P}_{max} = 0\}} A_{ij} \cdot q_t^i \quad \forall j \in \mathcal{J},\ \forall t.$$

Energy production equation: The energy produced for Δt, in this problem Δt are equally lengthened as $168h$, is calculated as,

$$x_t^{ei} = q_{t-1}^i \cdot k^i \cdot \triangle t_{(t-1)} \quad \forall i \in \mathcal{I},\ t = 0, ..., T-1.$$

Revenue equation: The accounting equation for the cash position over time. It considers the interest rate r, and depends on the gain from selling electricity and the cost of pumping water to the upstream reservoirs. Note that the usage of x_{t-1}^{ei} vs. (ξ_t^e, ξ_t^p) in this equation reflects the fact that decisions have to be made before knowing the actual spot prices, at which electricity is bought and sold,

$$x_t^c = x_{t-1}^c \cdot (1+r)^{\triangle t_{(t-1)}} + \sum_{\{i \in \mathcal{I} | k^i > 0\}} x_{t-1}^{ei} \cdot \xi_t^e + \sum_{\{i \in \mathcal{I} | k^i < 0\}} x_{t-1}^{ei} \cdot \xi_t^p \quad \forall t.$$

The full mathematical model is found in (4.1.1). Note that decisions are only defined for emanating stages and thus not for terminal stage. However, this time oriented formulation does not refer to any specific discrete structure and is typically seen as an infinite variational problem and can not be solved unless it is approximated.

$$
\begin{aligned}
&\max\ \mathbb{E}[x_T^c] && (4.1.1)\\
&\quad \text{s.t. } 0 \le q_t^i \le \overline{q}^i && \forall i \in \mathcal{I},\ t = 0, ..., T-1\\
&\qquad \underline{v}^j \le v_t^j \le \overline{v}^j && \forall j \in \mathcal{J},\ t = 0, ..., T-1\\
&\qquad v_{end}^j \le v_T^j && \forall j \in \mathcal{J}\\
&\qquad v_t^j = v_{t-1}^j + f_t^j + \sum_{\{i \in \mathcal{I} | \mathbf{P}_{max} > 0\}} A_{ij} \cdot q_{t-1}^i +\\
&\qquad \sum_{\{i \in \mathcal{I} | \mathbf{P}_{max} = 0\}} A_{ij} \cdot q_t^i && \forall j \in \mathcal{J},\ \forall t\\
&\qquad x_t^{ei} = q_{t-1}^i \cdot k^i \cdot \triangle t_{(t-1)} && \forall i \in \mathcal{I},\ t = 0, ..., T-1\\
&\qquad x_t^c = x_{t-1}^c \cdot (1+r)^{\triangle t_{(t-1)}} + \sum_{\{i \in \mathcal{I} | k^i > 0\}} x_{t-1}^{ei} \cdot \xi_t^e +\\
&\qquad \sum_{\{i \in \mathcal{I} | k^i < 0\}} x_{t-1}^{ei} \cdot \xi_t^p && \forall t\\
&\qquad \xi_t \lhd \mathcal{F}_t && \forall t\\
&\qquad x \lhd \mathfrak{F}
\end{aligned}
$$

We would like to mention a state of the art review of optimal operation of multi-reservoir systems by Labadie [30] in which several solution strategies including implicit and explicit stochastic optimization and heuristic methods are explored. See also Giacometti et al. [21] for decision models with deterministic inflows and forward electricity prices.

We will use finite scenario trees (finite approximation) as the basic structure for application of the above model.[4] In the following the node- based version of the problem is represented in (4.1.2). In the node- based (tree) formulation non-anticipativity constraint $x \lhd \mathfrak{F}$, is enforced by the tree structure, which represents the relevant filtration. Similar to our discussion in Section 2.2, finite trees are characterized by the node sets $\mathcal{N}_t$ per stage t and the predecessor relations which uses the notation $pred_u(n)$ i.e., the predecessor of node n in $\mathcal{N}_u$, with $u < t$. If $u = t-1$ the notation is written as $pred_{t-1}(n)$ or n_-. In the formulation of problem (4.1.2), n_- denotes the direct predecessor relation.

[4]In stochastic optimization literature formulation (4.1.2), is often called the "deterministic equivalent" of (4.1.1). This terminology in not correct though, since the relation between finite scenario tree formulation and the time-oriented problem is an approximation and not an equivalence. [Prof. G. Pflug in one of our group discussions]

$$
\begin{aligned}
\max \; & \sum_{n \in \mathcal{N}_T} P(n) \cdot x^c(n) && (4.1.2)\\
\text{s.t. } & 0 \le q^i(n) \le \overline{q}^i && \forall j \in \mathcal{I},\ \forall n \in \mathcal{N} \setminus \mathcal{N}_T\\
& \underline{v}^j \le v^j(n) \le \overline{v}^j && \forall j \in \mathcal{J},\ \forall n \in \mathcal{N} \setminus \mathcal{N}_T\\
& v^j_{end} \le v^j(n) && \forall j \in \mathcal{J},\ \forall n \in \mathcal{N}_T\\
& v^j(n) = v^j(n_-) + f^j(n) + \sum_{\{i \in \mathcal{I} | \mathbf{P}_{max} > 0\}} A_{ij} \cdot q^i(n_-) + \\
& \sum_{\{i \in \mathcal{I} | \mathbf{P}_{max} = 0\}} A_{ij} \cdot q^i(n) && \forall j \in \mathcal{J},\ \forall n \in \mathcal{N}\\
& x^{ei}(n) = q^i(n_-) \cdot k^i \cdot \triangle t(n_-) && \forall i \in \mathcal{I},\ \forall n \in \mathcal{N} \setminus \mathcal{N}_T\\
& x^c(n) = x^c(n_-) \cdot (1+r)^{\triangle t(n_-)} + \sum_{\{i \in \mathcal{I} | k^i > 0\}} x^{ei}(n_-) \cdot \xi^e(n) + \\
& \sum_{\{i \in \mathcal{I} | k^i < 0\}} x^{ei}(n_-) \cdot \xi^p(n) && \forall n \in \mathcal{N}\\
& x^{ei}, q^i \in \mathbb{R}_+^{|\mathcal{N} \setminus \mathcal{N}_T|} && \forall i \in \mathcal{I}\\
& v^j \in \mathbb{R}_+^{|\mathcal{N}|} && \forall j \in \mathcal{J}\\
& x^c \in \mathbb{R}^{|\mathcal{N}|}
\end{aligned}
$$

To summarize, our model is a specification of the short term hydropower scheduling problem, based on simulated spot price, pumping price and natural inflow scenarios and leads to a large scale multistage linear stochastic optimization problem, which can be solved by available standard software. However, after the solving this problem, our aim is to solve the distributionally robust counterpart of (4.1.2) using Algorithms 3.1 and 3.2. In the next section a scenario tree generation algorithm for co-generation of spot prices, pumping prices and natural inflows to reservoirs is discussed.

4.2 Price-Inflow Modeling

Prices-Inflow scenario tree generation is done in a two-step procedure. First, simulation of spot prices and natural inflow scenarios using an explicit representation of spot price distribution and natural inflows model selection. Secondly, construction of a tree out of these scenarios using Pflug and Pichler [37] approach. In the following subsections each of these steps are described accordingly.

Table 4.2.1: PC Analysis of monthly inflows to the Reservoirs 1,2,6

	PC_1	PC_2	PC_3
January	0.0954	−0.2781	0.1724
February	0.0895	−0.3036	0.2489
March	0.1107	−3.3145	0.1792
April	0.1739	−0.3786	0.0937
May	0.4239	−0.2552	−0.5581
June	0.5036	0.2322	−0.4183
July	0.4669	0.3574	0.2157
August	0.3651	0.3489	0.5273
September	0.2805	−0.1117	0.1586
October	0.2095	−0.216	0.0277
November	0.1478	−0.2855	0.0863
December	0.1063	−0.2812	0.1518
Var.	3182	49.0	7.5
	95.25%	1.52%	0.23%

4.2.1 Natural Inflow Modeling - A Model Selection Procedure

To identify the statistical model for the inflows to the reservoirs, we use a 13- year daily profile for a period between 01.01.1995 and 31.12.2007. The time series plots of the the historical inflows ($\frac{m^3}{sec}$) to reservoirs 1, 2 and 6 from the system topology observed in Figure 4.1.1 are displayed in Figure 4.2.1. This historical record, shows a periodic pattern which is completed within a calendar year and again is repeated on annual basis. A more in depth analysis of the inflows to the reservoir via *Principal Component Analysis* (PCA) has shown environmental and climate change patterns when applying the analysis for all reservoirs together and each individual reservoir respectively. Table 4.2.1 reflects the PC Analysis for measurements in ($\frac{m^3}{sec}$) of monthly natural inflows to reservoirs 1, 2 and 6. The first component accounts for 95.25% of the total variance and reflects the variation in overall inflows, however, the second component is dominated by variations in summer months *June*, *July* and *August*. As a conclusion, reservoirs vary most in the overall and next in the seasonal inflows. Moreover, all three reservoirs seem to have a clearly pronounced seasonal pattern. Looking at the total annual inflows to the reservoirs, we observe stability over time which does not show any trend indication in the amount of inflows.

In the present study, as in Papamichail and Georgiou [33], a $SARIMA$ model is selected to simulate the time series of historical data of natural inflows into

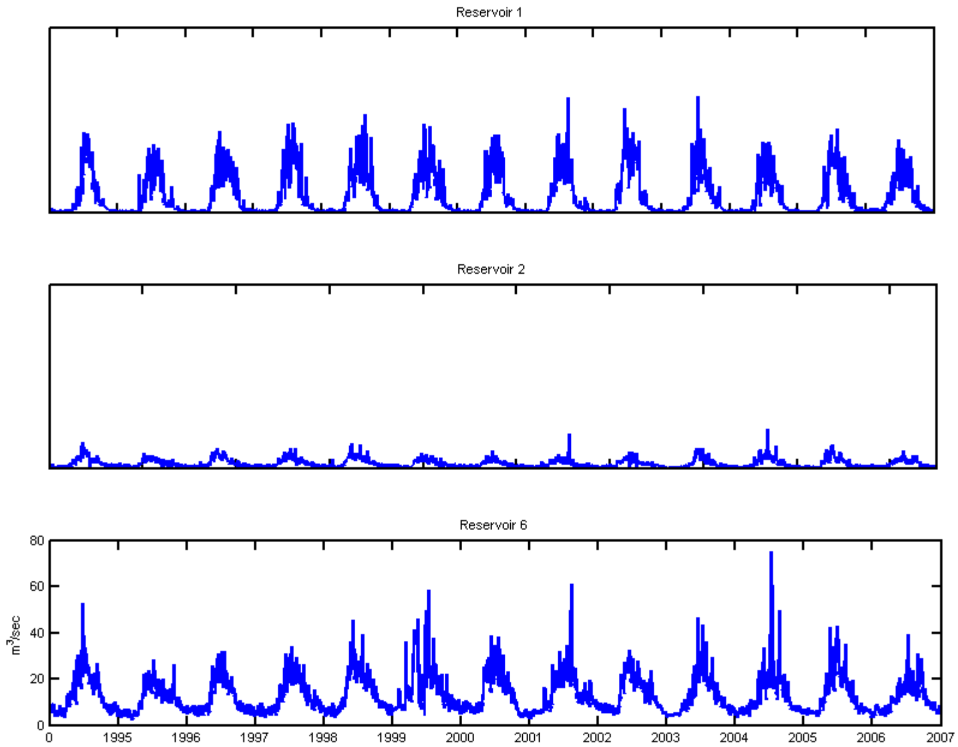


Figure 4.2.1: Historical profile of natural inflows to the reservoirs (1995-2007)

a chain of reservoirs over a one year period. In the model selection stage, both *Auto Correlation Function* (ACF) and the *Partial Auto Correlation Function* (PACF) of the logarithmically[5] transformed time series are considered. By implementing the "Hyndman-Khandakar" algorithm and by minimizing the Akaike's Information Criterion (AIC) factor, $SARIMA(1,0,2)(2,0,2)_{52}$ seems to be the appropriate candidate model. In Appendix A.3.2 the details of Hyndman-Khandakar algorithm and model coefficients are explained. This model will be used as an input for tree generation algorithm.

[5]To stabilize the variance, the logarithmic transformation is applied into the inflow time series Ledolter [31].

4.2.2 Spot Price and Pumping Price modeling: A Regime Switching Approach

It is well observed that electricity is an expensive sortable good. This property of electricity together with volatile and inelastic corresponding demand, specially in short term horizon, have differentiated electricity price modeling from all other financial assets. According to Eydeland and Wolyniec [18], one of these unique characteristics is that electricity prices exhibit pervasive spikes of extraordinary magnitude. Hence, it is very important to be able to model the spike behavior of electricity prices. Spike behavior in spot price evolution process can be interpreted, as in Keles et al. [27], that at hour h, spot price: SP_h may with probability p_h jump (up/down) to another price level, "spike regime", and afterward jumps back to "base regime". Regime switchings are models in which a number of regimes, or states, among which the price process moves are specified. This allows the price process to behave in any desirable way in any given regime, therefore introducing probability distribution of switches between regimes is required too, Eydeland and Wolyniec [18].

In this study we are using a similar approach to Kovacevic and Parachiv [28], in which the authors assume upward and downward spike dynamics are exponentially distributed and model dynamics for spot prices fluctuate around observed *Hourly Price Forward Curve* (HPFC). In Figure 4.2.2 the hourly forward price together with its upper and lower limits of Gaussian dynamics is shown. In addition, to observe the significance of including jumps and stochastic volatility in electricity spot price modeling due to the characteristics mentioned above, Figure 4.2.3 is providing the comparison between observed hourly forward and spot prices generated by regime switching approach. The details of spot price process and its corresponding parameters are explained in Appendix A.4. In addition to modeling spot price for trading electricity in the market, in pumped storage hydroelectric plants, pumps which transfer water to upper reservoirs, are themselves energy consumers therefore operate during off-peak hours where the electricity price is lower. Off-peak hours start from midnight to 6:00 a.m., however, it may vary in durations with respect to seasonal factors and holidays. We implement spot price dynamics according to (A.4.1) and simulate hourly spot prices from which the mean weekly spot prices are constructed. For this study, we simply consider the mean of off-peak, 01:00-06:00 a.m., spot prices to generate weekly pumping price. Finally, the simulation process provides us with two-dimensional vector of weekly spot and pumping price at each

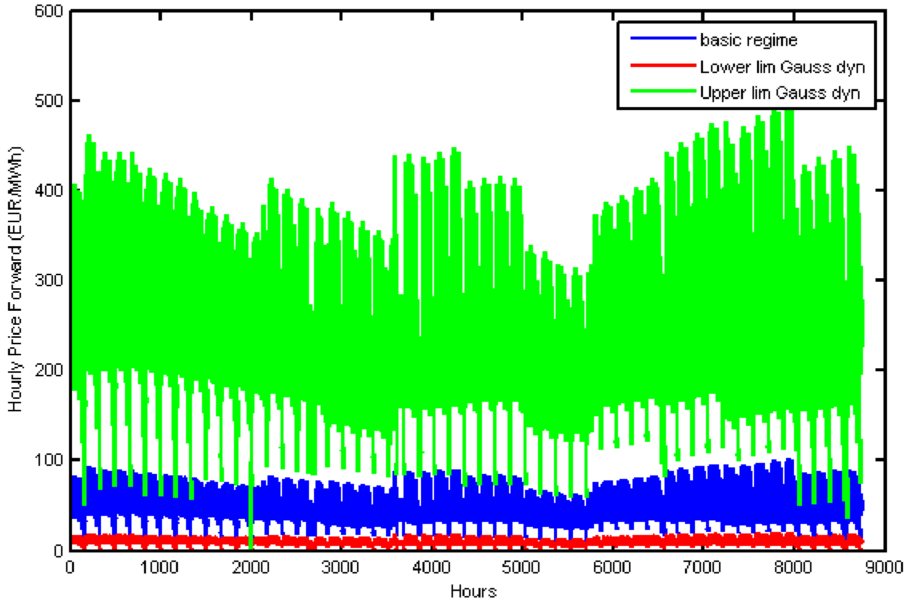


Figure 4.2.2: Hourly Price Forward Curve -year 2011

discrete time instance.

4.3 Scenario Tree Generation using Optimal Quantization

For the generation of scenario tree, suppose that large number of multivariate (five-vriate, in this case study) stochastic process $\xi_t = (\xi_t^e,\ \xi_t^p,\ f_t^j)$ consisting of weekly electricity spot and pumping prices and natural inflows to the reservoirs 1,2 and 6 are simulated at time instance t. These samples are used as inputs to the *optimal descretization* and *dynamic tree generation* algorithms from Pflug and Pichler [37]. These mentioned algorithms, aim to produce a finite discrete approximation of multivariate stochastic process ξ for $t = 1, ..., T$ given the maximal transportation distance vector $(\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_T)$ and the minimal bushiness vector $(\mathrm{b}_1, \mathrm{b}_2, \cdots, \mathrm{b}_T)$. Minimum bushiness at each stage t determines minimum number of direct successors for each node at that stage. In multivariate stochastic process, which is the case here, it is crucial to allocate appropriate weights to different components of ξ such that in the calculation of distance (cf. (2.2.3)), scaling problem does not occur. For our case study, the very first step is adjusting weights w_t^m for $m = 3, 4, 5$ which represent the corresponding weights to reservoirs' inflow scenarios to be comparable

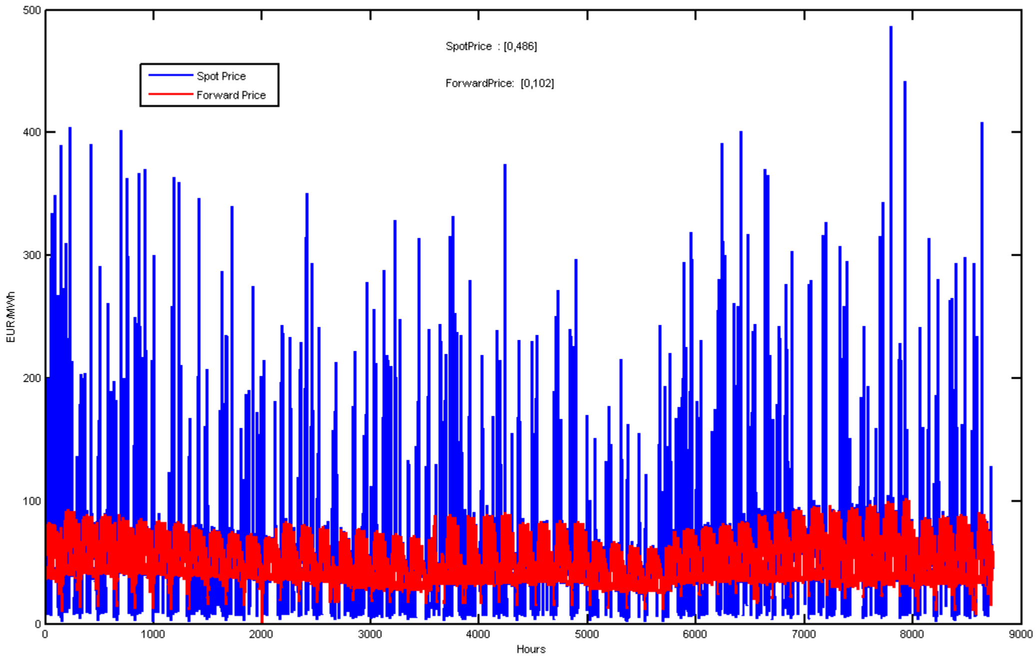


Figure 4.2.3: Simulated hourly spot prices (Regime Switching approach) vs. corresponding forward prices -year 2011

Table 4.3.1: Algorithmic statistics

$(\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_8) = (5, 5, 5, 8, 8, 8, 12)$
$(\mathrm{b}_1, \mathrm{b}_2, \cdots, \mathrm{b}_8) = (2, 2, 2, 1, 1, 1, 1)$
$\xi_0 = (\xi_0^e,\ \xi_0^p,\ f_0^j) = (55, 41.29, 82.3, 98.1, 1239)$
Number of simulation per node: 250

with spot and pumping price scenarios. Figure A.4.2 shows a glance of immediate requirement for scaling.

Although we are highly interested to model the uncertainties for a full year, for computational purposes and tractability of DRMSO algorithm with available technology, we sufficiently choose a short term horizon of two months length with weekly time intervals. In addition, to capture the seasonal effect of spot price and natural inflows, we shift the scenario simulation window and start from week 17 of the year. Figure 4.3.1 shows trajectories of weekly forward price for years 2011-2016 between weeks 10-30. And Figure 4.3.2 depicts the historical weekly trajectories (13-year profile) of natural inflows to reservoirs 1,2 and 6 based on the same time window. Following Subsections 4.2.1 and 4.2.2, our simulation function generates weekly inflows based on $SARIMA(1, 0, 2)(2, 0, 2)_{52}$ for reservoirs 1,2 and 6 and weekly spot and pumping prices using regime switching approach as inputs to the tree generation algorithm. In this regard, some algorithmic and computational statistics in Tables 4.3.1 and 4.3.2 are reported for the constructed tree. From computational point of view, size of the generated tree is satisfying, its quality, however, will be examined through the DRMSO result analysis in the next chapter. The visualizations of the generated price-inflow tree in three formats: structure, spot price component and natural inflows to Reservoir 1 component are also shown in Figures 4.3.3, 4.3.4 and 4.3.5 respectively. Three other components, namely: pumping price component, inflows to reservoir 2 and 6 components are shown in Figure A.4.3 in the appendix.

Having the baseline model available, in the next step the computational results of Algorithms 3.1 and 3.2 for problem (4.1.2) are presented and discussed.

Table 4.3.2: Computational performance

RunTime = 7265 sec $\sim 2h$ Intel Core i7 CPU @ 3.40GHz, 8GB RAM
Total number of nodes : 421
Total number of leaf nodes (scenarios) : 119
Number of nodes per stage : (1 3 11 35 51 82 119 119)

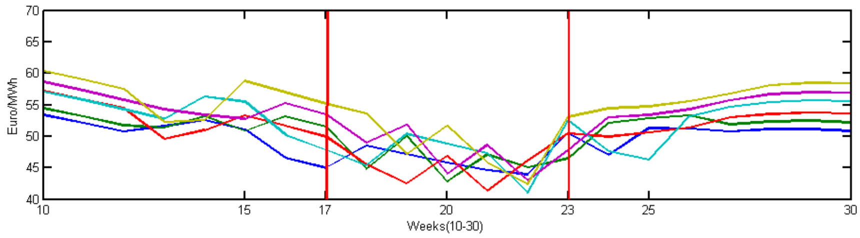


Figure 4.3.1: Weekly electricity forward curve paths (year 2011-2016, weeks 10-30)

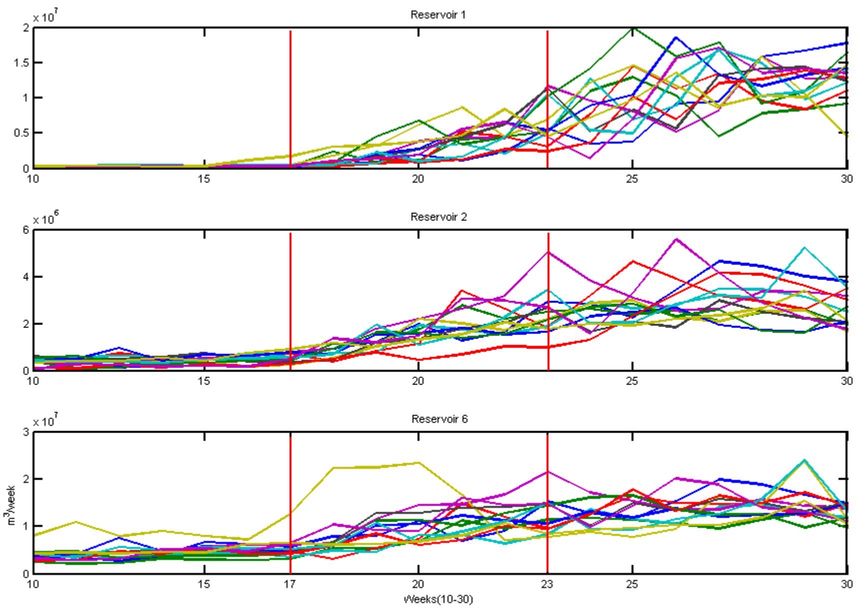


Figure 4.3.2: Weekly natural inflows paths (year 1995-2007, weeks 10-30)

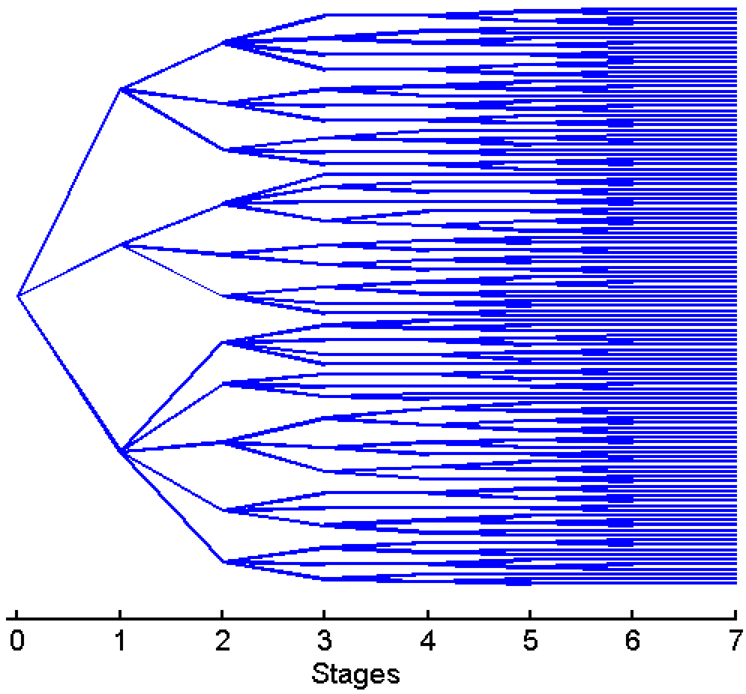


Figure 4.3.3: Tree structure, # nodes per stage [1 3 11 35 51 82 119 119]

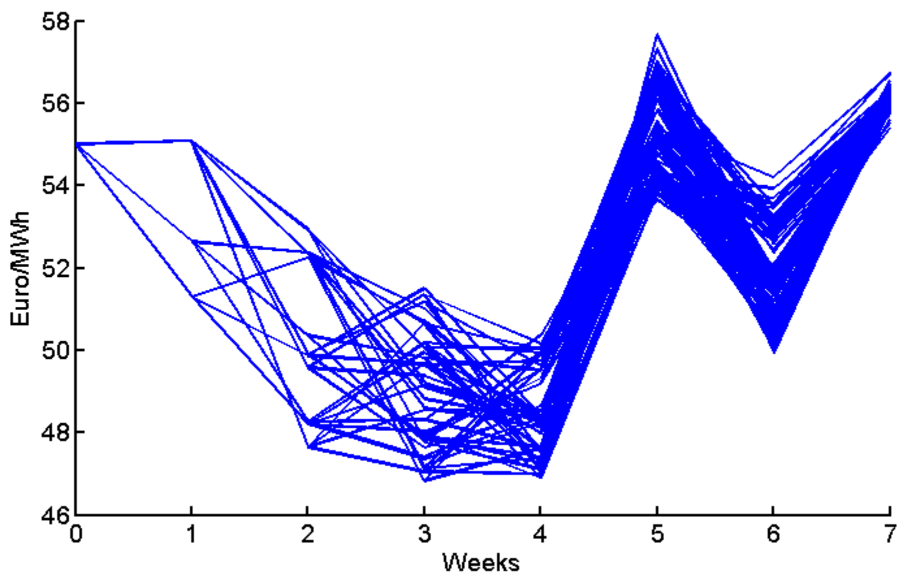


Figure 4.3.4: Spot price component of the scenario tree

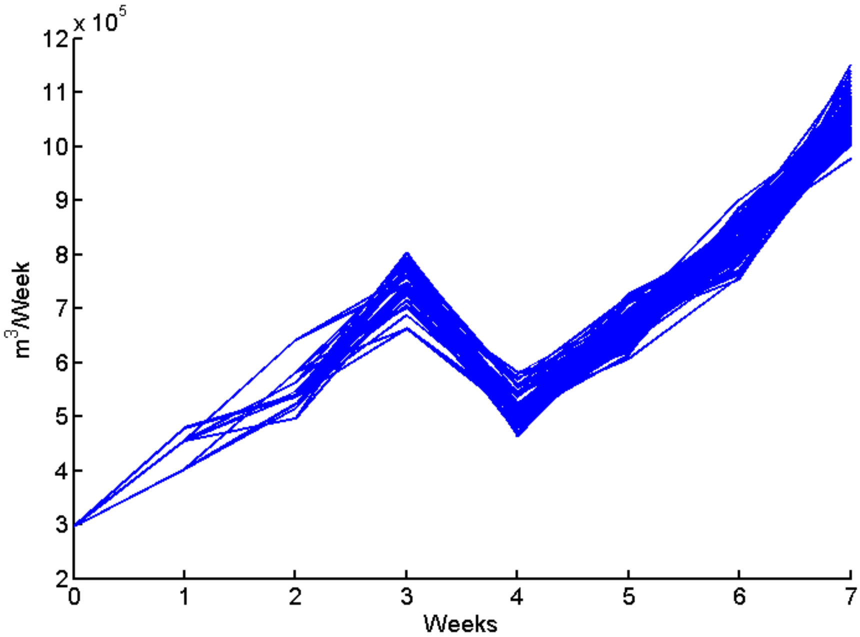


Figure 4.3.5: Reservoir 1 - natural inflow component of the scenario tree

Summary

As a case study for our distributionally robust approach to a real world decision problem a hydro electricity production planning under *market* and *weather* uncertainty was introduced and the corresponding topology and mathematical structure was explained. In addition model selection approach on the time series of inflow data and regime switching approach for modeling spot price and pumping price resp. provided us with a simulation procedure for Price-Inflow scenario co-generation. Using tree generation algorithms in [37, Chapter 4, Algorithms 4.5 and 4.10], our baseline model is generated. This is the tree around which the nested neighborhoods for various ambiguity radii are constructed. With this regard, in the next chapter first the short-term hydroelectricity optimization problem is solved with respect to the baseline model and then the distributionally robust computational results from the corresponding Algorithms 3.1 and 3.2 are discussed and analyzed.

Chapter 5

Computational Results and Maximin Decisions

The purpose of this chapter is to study the impacts of implementing distributionally robust approach discussed in Chapter 3 to a real world decision problem. Therefore, our aim is first to solve problem (4.1.2) with respect to the baseline tree, generated in Section 4.3, for the system topology shown in Figure 4.1.1 and further derive the maximin solutions of its distributionally robust counterpart for different ambiguity radii.

5.1 Optimal Short-term Hydroelectricity Generation

The hydro chain is considered to be participating in the market for a short term period of two months. The study horizon is split into 8 stages, each representing one week. We used AIMMS 3.12[1] for for formulating and solving (under the baseline model) the linear multistage stochastic problem (4.1.2) for a chain of 6 reservoirs with 17 arcs. In the following Table 5.1.1, some computational aspects of the corresponding instance is reported. Recalling problem (4.1.2), we have seen that it has a linear structure and there are no integrality constraints. In particular, the objective function consists only of the expected value of the terminal revenues which is linear

[1] *Algebraic Modeling Languages* (AMLs) have been developed since 1980s. They use notation closely tied to the algebraic formulation of optimization problems and often transcribing mathematical symbols into character strings that can be represented on an ordinary keyboard. Examples of AMLs are GAMS, AMPL, MPL, AIMMS, Modler and LPL [20]. *Advanced Interactive Multidimensional Modeling System* (AIMMS) - by Johannes J. Bisschop and Marcel Roelofs- is a software system designed for modeling and solving large-scale optimization and scheduling-type problems.

Table 5.1.1: Computational statistics

# Constraints	14892
# Variables	17193
# Non zeros	40346
Solver	GUROBI 5.0
Phase	Dual Simplex
Iterations	9934
Solving Time	0.23 sce

too. By employing this, the hydro electricity production planning is conducted in a risk neutral manner.

Figure 5.1.1 depicts density of accumulated cash over time together with the expected cash per stage . Accumulated cash density has a clear increasing trend over time and exhibits small variances at each stage. Experience has shown that, this phenomena might become less extreme by introducing interest rates r and risk-averse objective function. Since our main objective is analyzing the DRMSO results, we would keep the structure of the problem as it is i.e., linear and risk neutral.

In Figure 5.1.2, reservoir storage level scenarios for reservoirs 1 and 2 and 3 (top three reservoirs of topology 4.1.1) is shown. Although reservoir 2 has natural inflows, the interconnection between reservoirs 2 and 3 can be observed from the storage level patterns. Recalling the system topology, reservoirs 2 and 3 are connected via *turbine-arc3* and *pump-arc4*. The pattern clearly reflects that when pumping up decisions show a decrease in storage level of lower reservoir, simultaneously the storage lever in the upper reservoir is showing an increase. Some times the size differences between reservoirs may restrict the flexibility of the system. Meaning that, water release from the top reservoir forces the releases and sometimes even spillage in the lower reservoir. This is not the case here, since the lower reservoir has a size $\sim$ 10 times larger than the upper reservoir and the pumping decisions are with respect to the state of the system and not forced. Besides, singular and sparse scenario paths enfold the fact that mostly identical optimal decisions are taken. This observation, however, decays in the ambiguity problem and will be discussed in the next section where we will solve the ambiguity extension of problem (4.1.2) for various ambiguity radii.

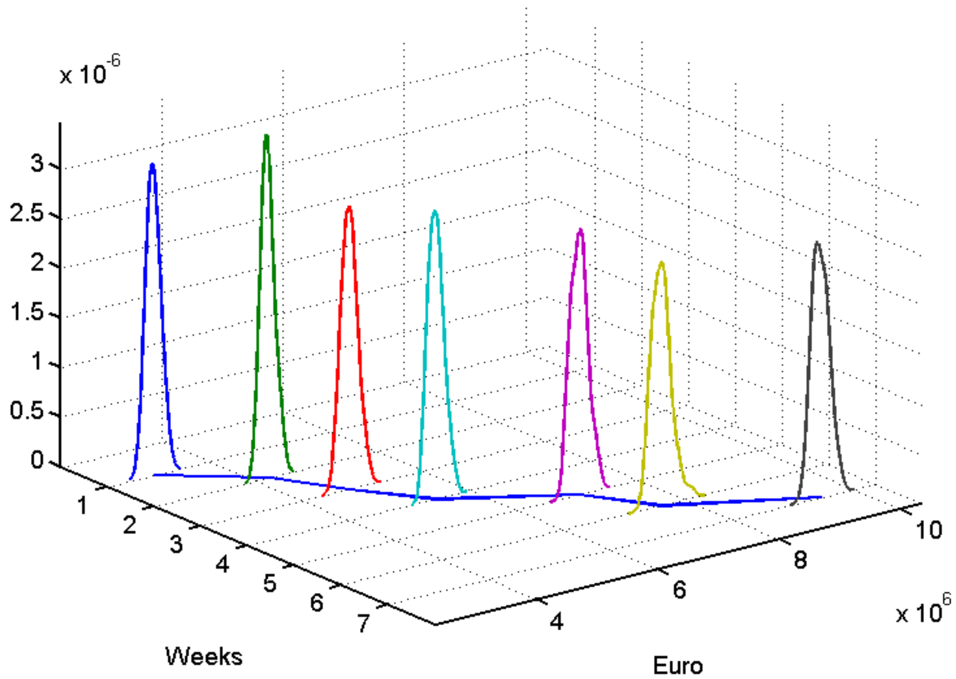


Figure 5.1.1: Accumulated cash density

5.2 Distributionally Robust Decisions

5.2.1 Pre-processing for Ambiguity Radii and Worst Tree Construction

In Section 3.4, we let that ambiguity radius range within $[\min \mathsf{d}(i,j),\ \max \mathsf{d}(i,j)]$, with $\mathsf{d}(i,j)$ representing the distance between demand scenarios i and j. In addition, uncertain parameters, demands of two products, were consistent and scaling was not required. In hydro problem, on the other hand, the uncertainties are prices and inflows. Therefore, not only scaling factors are required but also the interpretation of ϵ itself is more delicate. For the constructed tree in Section 4.3, after scaling $\max \mathsf{d}(i,j) = 256$, therefore ϵ is ranged between $[20,\ 200]$ at equal step size 60 and in addition for $\epsilon = 400$ (intuitively) too. Hence, we extend the original problem (4.1.2) to its distributionally robust counterpart for different ϵ and solve maximin problem using Algorithm 3.1 and Algorithm 3.2 therein. In the following, we first demonstrate the results of the worst cases and in the next subsections the maximin solutions and corresponding efficient frontiers. In Figure 5.2.1, the scenario probabilities of the respective tree structure for increasing ambiguity radii are depicted. It is observed

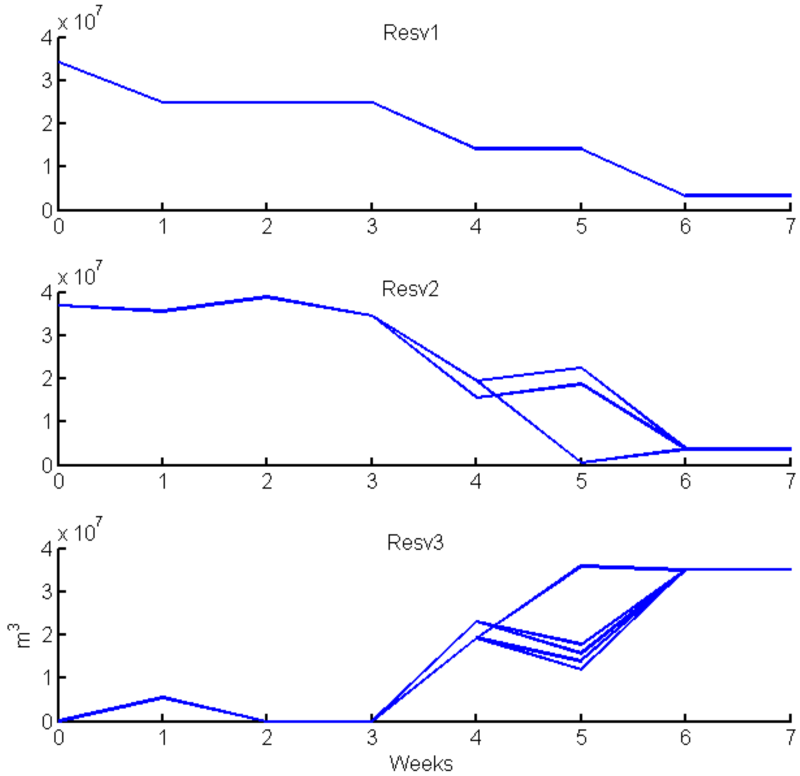


Figure 5.1.2: Reservoir interconnection

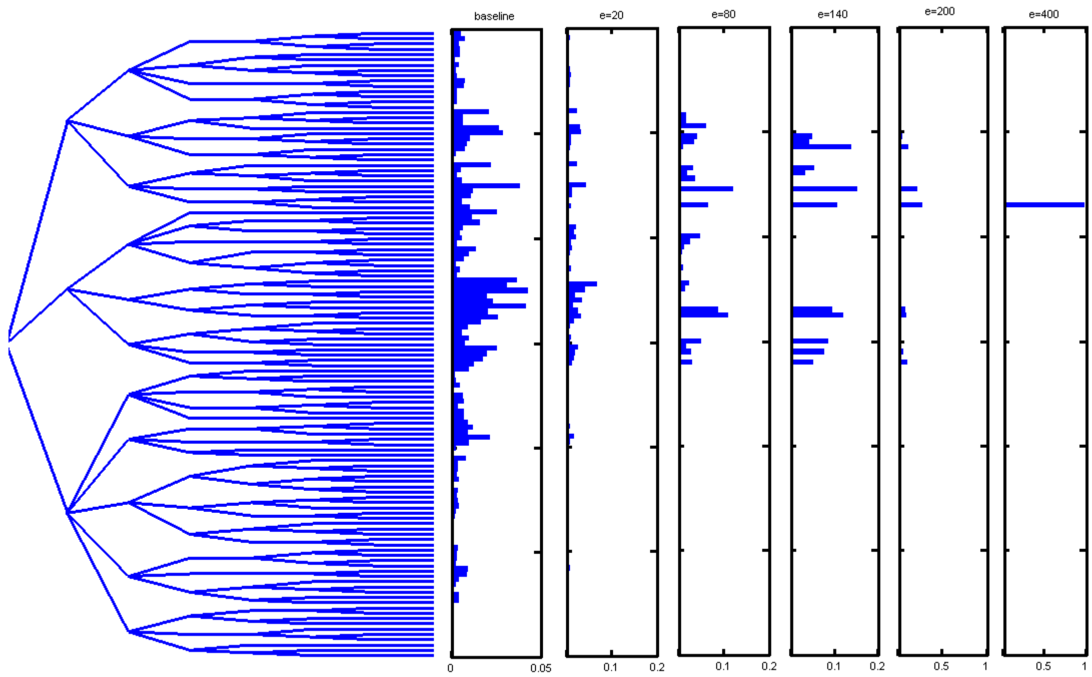


Figure 5.2.1: Tree structure and diminishing scenarios (e corresponds to ϵ)

that at the largest radius i.e., $\epsilon = 400$, only few scenarios remain among which one scenario has the probability very close to one. Figure 5.2.1, irrespective of tree components, only depicts the position and probabilities of the remaining worst scenarios at each ambiguity radius. In this regard and to observe how the scenario tree is actually modified with respect to increasing ambiguity radius, Figure 5.2.2 is presented for spot price component. It is clear that there are only four worst scenarios remaining at the largest ϵ, form which only one scenario has a significant weight.

5.2.2 Sensitivity of Maximin Decisions - More Dynamic Scenarios

We refer again to the system topology 4.1.1 and analyze the sensitivity of decisions i.e., water flows through arcs, in the ambiguity neighborhoods of the baseline model, from two perspective. First, we look at decision scenarios of *turbine-arc6*, which connects reservoirs 3 and 4. Second, we study reservoir storage level scenarios for reservoirs 1, 2 and 3 which implicitly represents the activities of incorporated *turbine-*

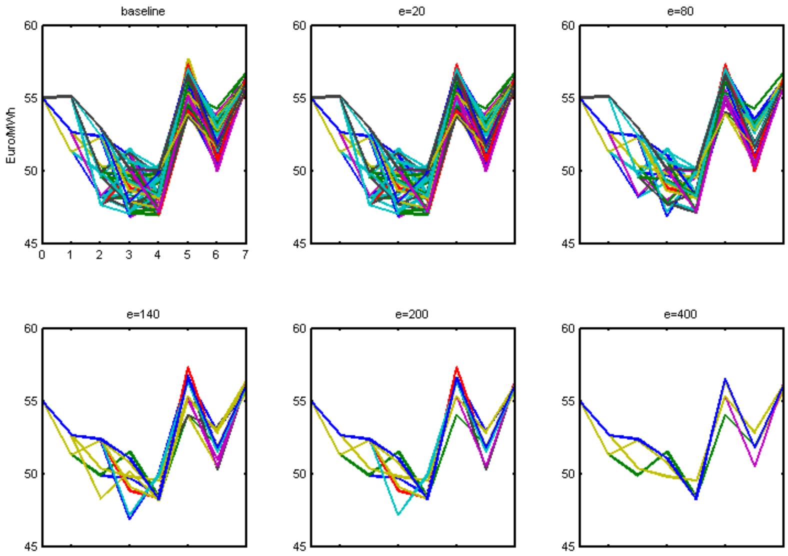
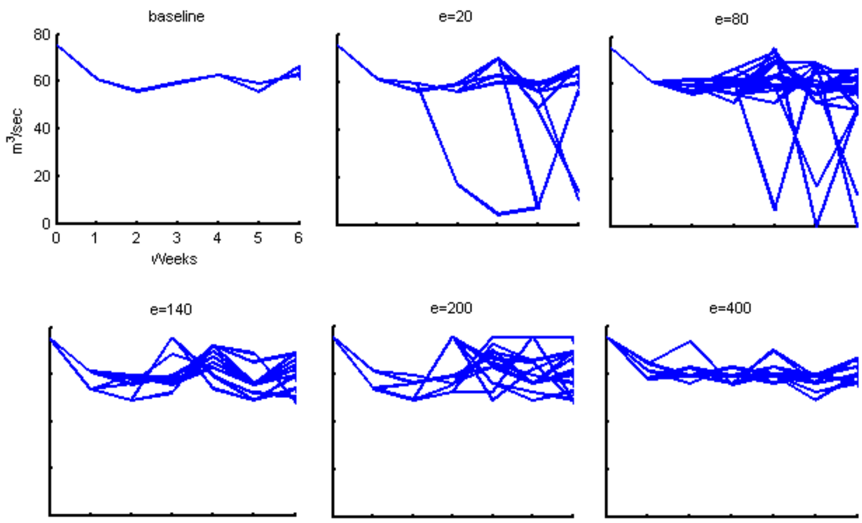


Figure 5.2.2: Spot price component and diminishing scenarios

arc1, *turbine-arc3* and *pump-arc4*.

In Figure 5.2.3, first we observe an increase in the complexity of maximin decisions along with the increasing ambiguity radius. This phenomenon has been previously seen for distributionally robust solutions in this thesis and all other instances we have solved. It is argued that more complex scenarios are caused by the fact that decisions are taken with respect to several models and not only the baseline model. In addition to that, state of the system when baseline model is distorted also affects the decision scenarios. Since there is no natural inflows to reservoir 3 and it is relatively large in size, we only look at Figure 5.2.2 with a magnifying glass for worst tree with $\epsilon = 20$, more specifically to the scenarios at stage 4 where lower spot prices are expected, extreme decision scenarios at stage 3 are observed. However, when ambiguity radius increases, decision scenarios in stages 1 and 2 tend to be more complex because the corresponding worst trees (spot price component) at neighborhoods $\epsilon \geq 140$ are more variable at the beginning stages.

In Figure 5.2.4, again the complexity of decision scenarios with respect to increasing ambiguity radius is conformed. Also, in reservoirs 2 and 3, more scenarios at which the reservoirs are nearly or completely empty are detected. This obviously implies the appearance of mixed effects between future spot and pumping prices and natural inflow to reservoir 2, as long as all three components are involved in the



$\bar{q}^6$	$25,200\ \frac{m^3}{sec}$

Figure 5.2.3: Water flow through arc 6 (decision variable : q^6) scenarios

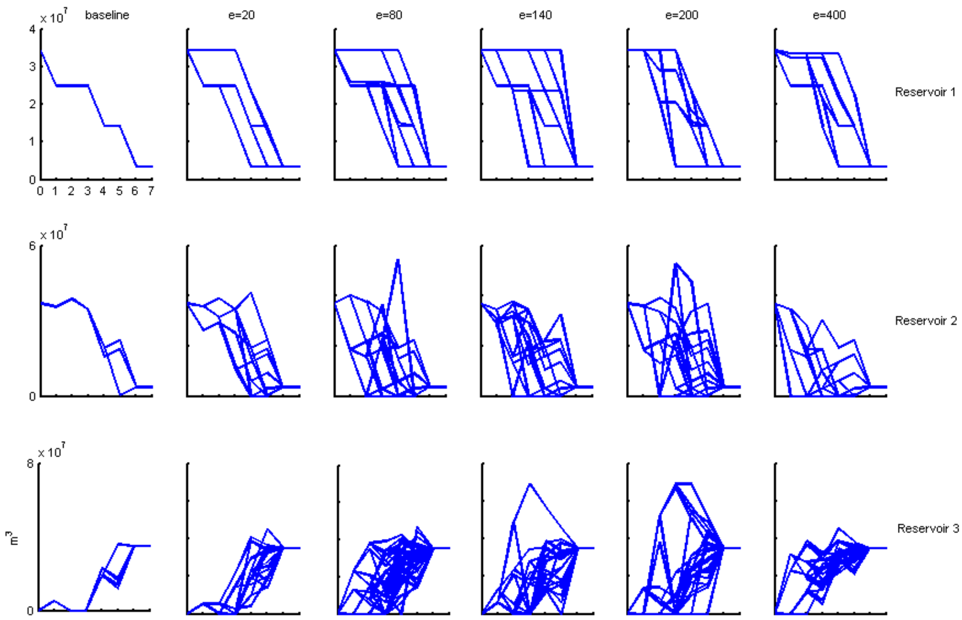
operation between these two reservoirs.

5.2.3 Cost of Ambiguity and Reward for Robustness

Similar to our analysis in Section 3.4.2, we consider the following inequality relation.

$$\min_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x^*(\mathbb{P}),\xi)] \leq \max_{x} \min_{\tilde{\mathbb{P}}\in\mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x,\xi)] \leq \mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathcal{P}_\epsilon),\xi)] \leq \mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathbb{P}),\xi)]. \tag{5.2.1}$$

The optimal expected revenue is seen as a function of robustness. Recall that H was introduced as a convex cost function in 2.3.5. Since in the problem of this chapter we are maximizing the expected terminal revenue, similar to Subsection 3.4.2, we consider $\bar{H}$ as the concave revenue function. In Table 5.2.1 numerical results for inequality 5.2.1 are presented. In addition, cost of ambiguity and reward of robustness are quantified for increasing ambiguity radii. At the largest calculated neighborhood, the decision maker may be willing to pay 2.02% of the expected final revenue to obtain distributionally robust solution. Besides, in reward of successfully safeguarding against the worst cases she gains 0.4%. This amount is relatively low



$\bar{v}^1$	$8.5 \times 10^7\ m^3$
$\bar{v}^2$	$8.3 \times 10^7\ m^3$
$\bar{v}^3$	$85 \times 10^7\ m^3$

Figure 5.2.4: Reservoirs interconnection w.r.t. ambiguity radii

Table 5.2.1: Numerical results on distributionally robust frontier

	baseline	Ambiguity radius (ϵ)				
		20	80	140	200	400
$\mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathbb{P}), \xi)]$	9,472,018(€)	9,472,018	9,472,018	9,472,018	9,472,018	9,472,018
$\mathbb{E}_{\mathbb{P}}[\bar{H}(x^*(\mathcal{P}_\epsilon), \xi)]$	–	9,444,524	9,385,888	9,354,258	9,328,894	9,279,874
cost of ambiguity	–	27,494	86,130	117,760	143,124	192,144 (2.02%)
$\max_x \min_{\tilde{\mathbb{P}} \in \mathcal{P}_\epsilon} \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x, \xi)]$	–	9,436,328	9,377,692	9,346,062	9,320,698	9,275,367
reward of robustness	–	3,802	21,278	26,095	24,968	39,457 (0.4%)
$\mathbb{E}_{\tilde{\mathbb{P}}}[\bar{H}(x^*(\mathbb{P}), \xi)]$	–	9,432,526	9,356,414	9,319,967	9,295,730	9,235,910

and implies that the baseline optimal solutions are not very much worsened with respect to the worst cases. Figure 5.2.5 pictures this numerical results. We call the dash-dotted line *distributionally robust frontier*, and infer that in this problem, it is much more influenced with respect to the robustness and not as much with respect to the worst case.

Summary

In this chapter we solved a short term hydro electricity production problem under model ambiguity. Ambiguity neighborhoods were constructed around a generated tree consist of a five-variate stochastic parameter at each node. More precisely, the risk neutral problem was solved with respect to five ambiguity radii. At the largest radius, the worst tree consist only of very few scenarios one of which has the largest probability (close to 1). In terms of maximin decision sensitivities, we again observed the increasing of complexity in the decisions which were analyzed according to the state of the system. Finally by considering the objective function as a function of robustness, we drew a distributionally robust frontier and quantified the cost of ambiguity and reward of robustness such that the suboptimal solution at the largest ambiguity neighborhood only accounts for 2.02% decrease of the expected revenue.

In the next chapter we conclude our analysis and findings of this thesis and propose some direction for further research.

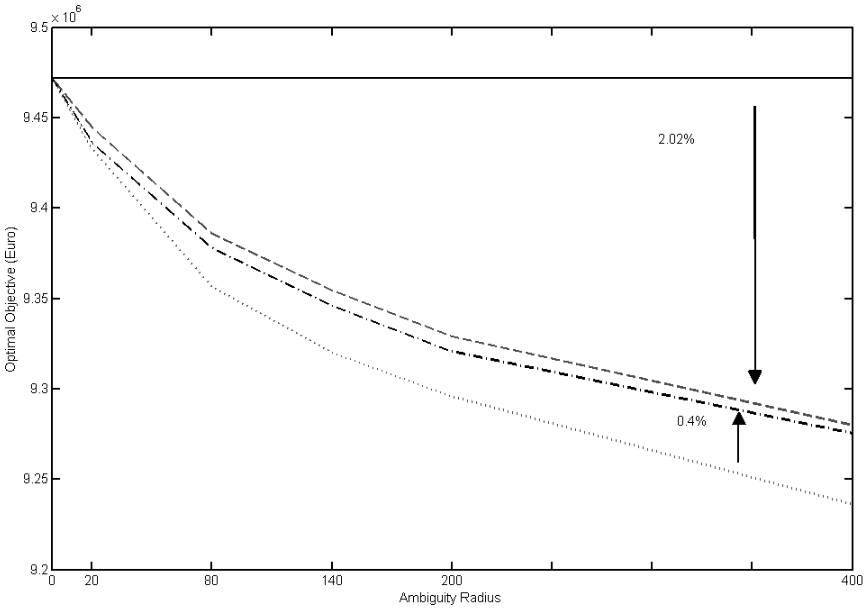


Figure 5.2.5: Cost of ambiguity and reward for robustness - Efficient frontiers with respect to robustness

Chapter 6

Conclusions and Further Extensions

6.1 Concluding Remarks

Incomplete knowledge about the underlying probability model specially in class of real world multistage stochastic optimization problems is inevitable. However, It is very naive to ignore it and at the same time very conservative to adapt to the classical robust approach. In this thesis a new non-parametric theoretical and algorithmic foundation as a quantitative methodology to account for model uncertainty in risk neutral, linear multistage stochastic Optimization is developed.

We have presented a minimax approach for risk neutral linear MSO problems, which accommodate both scenario uncertainty, aleatoric uncertainty, and probability model ambiguity, epistemic uncertainty. The ambiguity set is defined by $\epsilon-$nested neighborhoods around a baseline model and the baseline model is represented by a valuated tree. In addition to providing the theoretical requirements for multistage minimax theorem by introducing convex combination of trees in terms of compounding we present the proof and propose an algorithmic approach for constructing ambiguity neighborhoods and computing a saddle point of DRMSO problems. Specifically, since construction of nested ambiguity neighborhoods appears to be a non-convex problem we provide an approximation scheme in a stage wise manner. We implement this algorithm and provide a tool for quantifying the cost of ambiguity and reward of robustness around a distributionally robust frontier. We solve DR counterpart two MSO problem instances for different ϵ radii, and calculate the maximal corresponding cost and reward. It is observed that the larger the ambiguity radius, the simpler the worst tree becomes and even at the largest ϵ, only very few scenarios remain. On the contrary, the corresponding DR decision scenarios appear to be more and more complex. We argue that the complexity is the result of optimal decision making with respect to several and not only one baseline model, since bang-bang or boundary decisions are avoided. We also observe that the quality of modeling the uncertainties and consequently the approximation of the baseline is crucial and highly influential on the quality of solutions such that the optimal solution under the baseline is not extremely worsened under distortions.

We expect that in presence of high performance computing and more efficient algorithm implementations, ambiguity problems in multistage stochastic optimization settings are more easily solvable for larger trees.

6.2 Furtherance

Until now, we have studied the structural properties and numerical results of risk neutral DRMSO problems based on ambiguity set $\mathcal{P}_\epsilon$ defined in (2.3.7). While in this thesis we restricted our analysis based on small class of ambiguity sets, we believe that in near future this approach could be successfully extended and looked at from additional perspectives too. For the purpose of further research:

- It is desirable to incorporate risk functional in the objective function and study *Mean-Risk* extension of DRMSO. It is well established that when optimizing over a family of random variables, by taking only the expectation of individual members, in fact we are neglecting their variability. One possibility to address risk aversion, is to look for members of the optimization problem whose risk is acceptable. Under Mean-Risk class of models, the objective function is composed of the weighted sum of the expectation and some risk measure $\mathcal{R}$, i.e.,
$$\gamma \mathbb{E}_{\mathbb{P}}[H(x,\xi)] + (1-\gamma)\mathcal{R} \quad \text{for } \gamma \in [0,1]. \tag{6.2.1}$$
Depending on the the choice of risk functional,$\mathcal{R}$, this add-on in the objective function, would certainly affect saddle point computing algorithm. Since the proposed algorithm is designed for a linear criterion in the objective function this extension requires special consideration.
- It may be interesting to allow for a larger class of ambiguity set in which, in addition to probabilities, the filtration structure is subject to changes. In their paper Kovacevic and Pichler [29] proposed an algorithm such that a given

large tree is approximated and improved to a small tree (which is always more desirable in solving MSO efficiently) where the concentration is on improving the nested distance between large and small tree. In this regard, the structure of DRMSO when stage wise construction of worst trees allow for filtration changes by using the approach in [29], is another imaginable direction.

References

[1] B. Analui and R. M. Kovacevic. Medium term hydroelectric production planning - a multistage stochastic optimization model. *Civil Engineering and Infrustructure Journal*, 47(1):139–152, t.b.a June 2014.

[2] B. Analui and G. Ch. Pflug. On distributionally robust multiperiod stochastic optimization. *CMS Journal*, Accepted for Publication April 2014.

[3] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in Linear and Non-Linear Programming*. Stanford University Press, 1958.

[4] B. R. Barmish and P. Shcherbakov. Distributionally robust least squares. In *Proceedings of the International Symposium on Adaptive Systems, St. Petersburg, Russia.*, 1999.

[5] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 1997.

[6] J. Bisschop. *AIMMS Optimization Modelling*. Paragon Decision Technology, 2012.

[7] G. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal of Control Optimization*, 18(3):853–877, 2007.

[8] X. Chen, M. Sim, P. Sun, and J. Zhang. A linear-decision based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357, 2008.

[9] Z. Chen and L. Epstein. Ambiguity, risk and asset returns in continuous time. *Econometrics*, 70(4):1403–1443, 2002.

[10] Y. M. Danilin and V. M. Panin. Methods for searching saddle points. *Kibernetika*, 3:119–124, 1974.

[11] E. H. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problem. *Operations Research*, 58: 596–612, 2010.

[12] V. F. Demynov and A. B. Pevnyi. Numerical methods for finding saddle points. *USSR Computational Mathematics and Mathematical Physics*, 12:1099–1127, 1972.

[13] J. Dupačová. On minimax decision rule in stochastic linear programing. *Studies in Mathematical programming*, pages 47–60, 1980.

[14] J. Dupačová. The minimax approach to stochastic programming and an illustrative application. *Stochastics*, 20:73–88, 1987.

[15] J. Dupačová. Stochastic programming: Minimax approach. *Encyclopedia of Optimization (Floudas, Ch. A. and Pardalos, P.M. eds)*, V:327–330, 2001.

[16] J. Dupačová. Uncertainties in minimax stochastic programs. *Optimization*, 1: 191–220, 2010.

[17] D. Ellsberg. Risk, ambiguity and savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669, 1961.

[18] A. Eydeland and K. Wolyniec. *Energy and Power Risk Management.* John Wiley & Sons, Inc., 2003.

[19] K. Fan. Minimax theorems. *Proc. Nat. Acad. Sci.*, 39:42–47, 1953.

[20] H. I. Gassmann, S. W. Wallace, and W. T. Ziemba. *Stochastic Programming Computer Implementation*, chapter 1 in Applications of Stochastic Programming, pages 3–8. SIAM and MPS, 2005.

[21] R. Giacometti, M. T. Vespucci, M. Bertocchi, and G. B. Adesi. Hedging electricity portfolio for a hydro-energy producer via stochastic programming. *Stochastic Optimization Methods in Finance and Energy, Springer*, pages 163–179, 2011.

[22] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58:902–917, 2010.

[23] H. Heitsch, W. Römisch, and C. Strugarek. Stability of multistage stochastic programs. *SIAM Journal of Optimization*, 17:511–525, 2006.

[24] R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27(3), 2008.

[25] R. Jagannathan. Minimax procedure for a class of linear programs under uncertainty. *Operations Research*, 25:173–177, 1977.

[26] L. Kantorovich. On the transportation of masses. *C.R. Academy of Science USSR*, 37:199–201, 1942.

[27] DM. Keles, M. Genoese, D. Möst, and W. Fichtner. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Economics*, 34(4):1012–1032, 2011.

[28] R. M. Kovacevic and F. Parachiv. Medium-term planning for thermal electricity production. *OR Spectrum*, DOI:10.1007/s00291-013-0340-9, 2013.

[29] R. M. Kovacevic and A Pichler. Scenario trees - a process distance approach. *Available onlile http://www.optimization-online.org*, 2012.

[30] J. W. Labadie. Optimal operation of multi reservoir system: State of the art review. *Water Resources Planning and Management*, 130(2):93–111, 2004.

[31] J. Ledolter. A general class of stochastic models for hydrological sequences. *Journal of Hydrology*, 2(36):309–325, 1978.

[32] G. Monge. Mémoire sue la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires des Mathématique et de Physiue pour la même anneé*, pages 666–704, 1781.

[33] D. M. Papamichail and P. E. Georgiou. Seasonal arima inflow models for reservoir sizing. *Journal of American Water Resource Association*, 37(4):877–885, 2001.

[34] M. V. F. Pereira and L. M. V. G. Pinto. Stochastic optimization of a multireservoir hydroelectric system: A decomposition approach. *Water Resource Research*, 21:779–792, 1985.

[35] G. Ch. Pflug. Version-independence and nested distributions in multistage stochastic optimization. *SIAM Journal on Optimization*, 20(3):1406–1420, 2010.

[36] G. Ch. Pflug and A Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012.

[37] G. Ch. Pflug and A. Pichler. *Multistage Stochastic Optimization.* Preprints Springer, 2014.

[38] G. Ch. Pflug and W. Römisch. *Modelling, Measuring and Managing Risk.* Singapore: World Scientific., 1 edition, 2007.

[39] G. Ch. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

[40] L. Qi and W. Sun. An iterative method for the minimax problem. *Minimax and Applications (Kluwer)*, art. Du and Pardalos, 1995.

[41] S. Rachev and W. Römisch. Quantitative stability in stochastic programming: the method of probability metrics. *Mathematics of Operations Research*, 27: 798–818, 2002.

[42] W. Römisch and R. Schultz. Stability analysis for stochastic programs. *Annals of Operations Research*, 30:241–266, 1991.

[43] W. Römisch and R. Wets. Stability of e-approximate solutions to convex stochastic programs. *SIAM Journal of Optimization*, 18(3):961–979, 2007.

[44] W. Robinson and R. Wets. Stability in two stage stochastic programming. *SIAM Journal of Control Optimization*, 25:1409–1416, 1987.

[45] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[46] R. T. Rockafellar and R. Wets. *Variational Analysis.* Springer, 2004.

[47] B. Rustem and M. Howe. *Algorithms for Worst-Case Design and Applications to Risk Management.* Princeton University Press, 2002.

[48] A. Ruszczynski and A. Shapiro. *Stochastic Programming, 1st ed.* ser. Handbooks in Operations Research and Management science Amsterdam, 2003.

[49] H. Sasai. An interior penalty method for minimax for problems with constraints. *SIAM Journal of Control Optimization*, 12:643–649, 1974.

[50] H. Scarf. *Studies in the mathematical Theory of an Inventory Problems.* Stanford Univeristy Press, 1958.

[51] M. J. Schramm. *Introduction to Real Analysis.* Dover Publication, 2008.

[52] A. Shapiro and Sh. Ahmed. On a class of minimax stochastic programs. *SIAM Journal of Optimization*, 14:1237–1249, 2004.

[53] A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17:523–542, 2002.

[54] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8: 171–176, 1958.

[55] L. A. Terry, M. V. F. Pereira, T. A. A. Neto, L. F. C. A. Silva, and P. R. H. Sales. Coordinating the energy generation of the brazilian national hydrothermal electrical generating system,. *Interfaces*, 16(1):16–38, 1986.

[56] A. Thiele. Robust stochastic programming with uncertain probabilities. *IMA Journal of Management Mathematics*, 19:289–321, 2008.

[57] C. Villani. *Topics in Optimal Transportation.* The Americam Mathematical Society, 2003.

[58] J. von Neumann. Zur theorie der gesellschaftsspiele. *Math. Ann.*, 100:295–320, 1928.

[59] J. Žáčková a.k.a. Dupačová. On minimax solutions of stochastic linear programming problems. *Casopis pro Pestovani Mathematiky*, 91:423–430, 1966.

[60] D. Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research (online First)*, 2010.

[61] L. Xin, D. A. Goldberg, and A. Shapiro. Distributionally robust multistage inventory models with moment constraints. *Submitted to eprints*, April 2013.

Abstract

Multistage Stochastic Optimization (MSO) is a well-established framework where uncertainty is involved and decisions have to be taken in a sequential manner only based on the available information at the time of decision making. These two characteristics are enough to tie multistage stochastic optimization into almost all decision problems in the real life. However, in addition to parameters' uncertainty, in the class of real world problems, the true probability model, which describes these parameters, is itself subject to uncertainty that should not be ignored. Acknowledging the incomplete information about the underlying probability model in multistage stochastic optimization problems, leads to the following critical questions:

- How can we account for model uncertainty when solving a multistage stochastic program?
- What are the associated theoretical and algorithmic complexities?

This thesis develops a new theoretical foundation and a non-parametric approach which provide answers to these questions and can be used in a wide range of applications. In this regard, we start with a risk neutral MSO formulation and study the structure of the problem in presence of model uncertainty which we refer to as "ambiguity problem" too. The decision problem under consideration is defined on a filtered probability space. This filtered probability space is called "baseline" or "best guess" model. For the ambiguity problem, we consider "ambiguity neighborhoods" around the baseline model. The ambiguity neighborhood includes alternative models which are close to the baseline model. Closeness is defined in terms of a distance for nested distributions, called the "nested distance". This distance is appropriate for scenario models of multistage stochastic optimization problems. With this setting, the ambiguity problem is formulated in a *minimax* form. In minimax formulation, the optimal decision is to be found, which minimizes the maximal objective function, within the ambiguity neighborhood. Hence the ambiguity problem, extends the original MSO problem to account for "distributionally robust" solutions. We give a setup for studying saddle point properties of the multistage stochastic minimax problems. Moreover, we present an algorithmic approach for finding the minimax decisions at least asymptotically. In addition, by considering the objective as a function of robustness, we draw the distributionally robust frontier and quantify the costs and rewards of robustness around this frontier. Finally, we present a short term hydro electricity production problem with weekly ordering under weather and market risk. We find the worst models within the corresponding ambiguity neighborhood and determine a solution which is robust with respect to the model uncertainty.

Zusammenfassung

Mehrstufige Stochastische Optimierung (MSO) ist ein weitverbreiteter Zugang zur Optimierung unter Unsicherheit. Dabei müssen Entscheidungen aufeinander aufbauend, nur unter Nutzung der zum Entscheidungszeitpunkt verfügbaren Information, getroffen werden. Wegen diesen beiden Eigenschaften kann mehrstufige stochastische Optimierung in fast allen Entscheidungsproblemen im realen Leben verwendet werden. Neben der Unsicherheit der Parameter ist in Anwendungsproblemen allerdings das eigentliche Wahrscheinlichkeitsmodell selbst mit Unsicherheit behaftet, die nicht ignoriert werden sollte. Das führt zu zwei zentralen Fragestellungen:

- Wie kann Modellunsicherheit bei mehrstufiger stochastischer Optimierung berücksichtigt werden?
- Welche theoretischen und algorithmischen Komplexitäten entstehen dabei?

Diese Arbeit entwickelt eine neue theoretische Grundlage und einen nichtparametrischen Ansatz um Antworten auf diese Fragen zu geben und kann in einer Vielzahl von Anwendungen eingesetzt werden.

Zuerst wird eine risikoneutrale Formulierung der MSO eingeführt und die Struktur des Problems unter Modellunsicherheit untersucht. Das dabei entstehende Problem wird als "ambiguity problem" bezeichnen. Dieses Entscheidungsproblem ist auf einem filtrierten Wahrscheinlichkeitsraum, dem sogenannten "baseline model", definiert. Für das ambiguity problem wird eine Umgebung, die sogenannte "ambiguity neighborhood", um das baseline model konstruiert. Diese Umgebung beinhaltet alternative Modelle die im Sinne der "nested distance" nahe an dem baseline model sind. Das ambiguity problem wird bei diesem Zugang als ein minimax Problem formuliert. Dabei ist jene Entscheidung des MSO optimal welche das Maximum

der Zielfunktion über alle Wahrscheinlichkeitsmodelle in der gesamten Umgebung um das baseline model minimiert. Damit erweitert das ambiguity problem das ursprünglichen MSO, um Lösungen zu gewährleisten die robust bezüglich Wahrscheinlichkeitsmodellen sind.

Weiters wird ein Zugang für das systematische Studium von Sattelpunkteigenschaften von mehrstufigen stochastischen minimax Problemen eingeführt und ein algorithmischer Ansatz für die asymtotische Berechnung von minimax Entscheidungen vorgestellt.

Als numerische Illustration dieses Modellierungszugangs und des Lösungsalgorithmus behandeln wir ein Produktionsproblem für Wasserkraftwerke unter Wetter- und Marktrisiken. Kosten und Nutzen von robusten Entscheidungen werden durch den Vergleich der Qualität von Lösungen in Abhängigkeit von der Größe der ambiguity neighborhood analysiert.

Appendix A

Data, Parameters and Models

A.1 Wasserstein/Kantorovich - Distance of Probability Measures

The essence of transportation distance is the minimal (total) effort or cost required to move a particle from an origin to a destination. This topic was initiated by Monge [32] and was followed later by Kantorovich [26]. Intuitive approach toward distance between probability measures also stems from *Monge Transportation* problem. The cumulative effort is the sum of all distances arising from transporting a mass from ω to $\tilde{\omega}$ over the distance $\mathsf{d}(\omega, \tilde{\omega})$. The optimal value to is called *Wassestein* distance of order r $(r > 1)$ and is denoted by $\mathbf{d}_{WA_r}(P, \tilde{P})$.

$$\begin{aligned} \mathbf{d}_{WA_r}(P, \tilde{P}) = \quad & \inf_{\pi} \left(\int_{\Omega \times \tilde{\Omega}} \mathsf{d}(\omega, \tilde{\omega})^r \; \pi[\mathrm{d}\omega, \mathrm{d}\tilde{\omega}] \right)^{\frac{1}{r}} \\ \text{s.t.} \quad & \pi[A \times \tilde{\Omega}] = P(A) \\ & \pi[\Omega \times B] = \tilde{P}(B) \end{aligned} \tag{A.1.1}$$

and the infimum is with respect to all transportation plans (bivariate probability measures) π with the following marginals, for all measurable sets $A \in \mathcal{F}$ *and* $B \in \tilde{\mathcal{F}}$. For $r = 1$ the distance is called *Kantorovich* distance and is denoted by $\mathbf{d}_{KA}(P, \tilde{P})$.

$$\begin{aligned} \mathbf{d}_{KA}(P, \tilde{P}) = \quad & \inf_{\pi} \int_{\Omega \times \tilde{\Omega}} \mathsf{d}(\omega, \tilde{\omega}) \; \pi[\mathrm{d}\omega, \mathrm{d}\tilde{\omega}] \\ \text{s.t.} \quad & \pi[A \times \tilde{\Omega}] = P(A) \\ & \pi[\Omega \times B] = \tilde{P}(B) \end{aligned} \tag{A.1.2}$$

Solutions to the (linear) optimization problems (A.1.1) and (A.1.2) are very well studied. See Villani [57] for a comprehensive study and overview on this topic.

A.2 Direct Coordinate-wise Algorithm: A Counter Example

This example shows that direct coordinate-wise algorithm results in oscillation and refuses to converge in minimax optimization even in the case of smooth convex-concave functions. As it can be seen from figure, $f(x,y) = x^2 - y^2 + 2xy$ for $X = [-2,2]$ and $Y = [-2,2]$ is convex in $x-$component and concave in $y-$component. Therefore, the minimax optimization problem is denoted by:

$$\min_{x\in X} \max_{y\in Y} f(x,y) \tag{A.2.1}$$

The pair (x^*, y^*) is a saddle point of problem (A.2.1) if and only if $x^* = \arg\min f(\cdot, y)$ and $y^* = \arg\max f(x, \cdot)$. The following iterative algorithm shows oscillation in the $3rd$ iterations.

$\kappa = 0$ and $(x^0, y^0) = (1,1)$

$$\begin{cases} x^{\kappa+1} = \arg\min_{x\in X}\ f(x, y^\kappa) = \arg\min_{x\in X}\ (x^2 - (y^\kappa)^2 + 2xy^\kappa) \\ y^{\kappa+1} = \arg\max_{y\in Y}\ f(x^{\kappa+1}, y) = \arg\max_{y\in Y}\ ((x^{\kappa+1})^2 - y^2 + 2x^{\kappa+1}y) \end{cases}$$

$(x^1, y^1) = (-1,-1)$

$(x^2, y^2) = (1,1)$

i.e., the algorithm oscillates, however the saddle point $(x^*, y^*) = (0,0)$

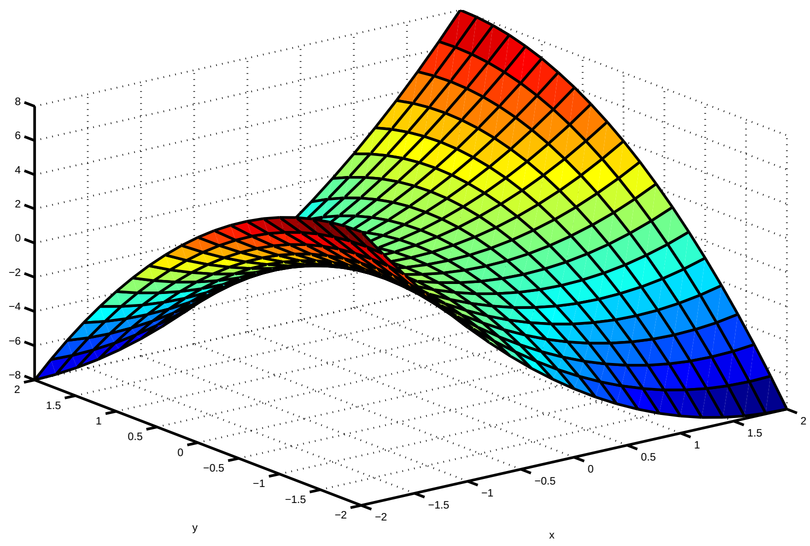


Figure A.2.1: Strictly convex-concave function $f(x, y) = x^2 - y^2 + 2xy$

A.3 Hydroelectricity Generation Model (4.1.2)

A.3.1 Incidence Matrix

Below the incidence matrix of the implemented topology (cf. Figure 4.1.1) is shown. Rows represent the water flows and columns the reservoirs.

$$A = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

A.3.2 $SARIMA$ Models and Parameters Estimation: *Hyndman-Khandkar* Algorithm [24]

Seasonal $ARIMA$ model is formed by including additional seasonal terms in the $ARIMA$ (Autoregressive Integrated Moving Average) models and is written as follows:

$$ARIMA \underbrace{(p, d, q)}_{non-seasonal} \underbrace{(P, D, Q)_m}_{seasonal}$$

where m is the number of periods per season. The seasonal part of the model consists of terms that are very similar to the non-seasonal components, but they involve Backshifts of the seasonal period. A common obstacle in using $SARIMA$ for forecasting is that selecting the orders of the parameters is not an easy task and it is very subjective. One built in algorithm known as *Hyndman-Khandkar* in package `forcast` in statistical software R is designed to select the order of $ARIMA(p, d, q)(P, D, Q)_m$ via a step wise procedure. The novelty of this approach is that it efficiently traverses the model space at each step and finally arrives to the appropriate model with the lowest Akaike's Information Criterion (AIC). Under known[1] d and D, AIC is calculated as:

$$AIC = -2\log(L) + 2(p + q + P + Q)$$

L is the maximized likelihood of the model fitted to the differenced data $(1-B)^d(1-B^m)^D y_t$ .The *Hyndman-Khandkar* algorithm can be summarized in the following two steps:

- For given d and D, start with the possible variations where p and q can take values between 0 to 2 and P and Q between 0 and 1. The one with the smallest AIC is selected and named as "current" model;
- Up to thirteen variations of the current model is considered:
  - only one of p, q, P and Q is allowed to vary by ± 1 from the current model,
  - p and q both are allowed to vary by ± 1 from the current model,
  - P and Q both are allowed to vary by ± 1 from the current model.

As soon as a model with lower AIC is substituted with the "current" model, the procedure is repeated until there is no improvement in AIC.

[1] For technical details please refer to [24].

Table A.3.1: Parameter Estimates

Parameter	Estimate	Standard Error
ϕ_1	0.8521	0.0386
Φ_1	0.1701	0.1824
Φ_2	0.8296	0.1827
θ_1	-0.2893	0.0561
θ_2	-0.2478	0.0432
Θ_1	-0.1608	0.1693
Θ_2	-0.8074	-0.8074

Using *Hyndman-Khandkar* algorithm, the corresponding $SARIMA(1,0,2)(2,0,2)_{52}$, which is without the constant terms, is selected and written in the following closed form.

$$(1-\phi_1 B)(1-\Phi_1 B^m - \Phi_2 B^{2m})y_t = (1+\theta_1 B + \theta_2 B^2)(1+\Theta_1 B^m + \Theta_2 B^{2m})e_t$$

The parameters' estimates are shown in Table A.3.1. These parameters are used to generate inflow scenarios as input for Spot Price-Inflow tree construction in section 4.3.

A.4 Regime Switching Parameters Decoding

According to [27], regime switching approaches should be employed in modeling electricity prices. Since prices may jump (up/down) to another price level, "spike regime", and afterward jumps back to "base regime" . In [28], the authors assume upward and downward spike dynamics are exponentially distributed. The model dynamics for spot prices is represented in the following hourly forms and spot prices fluctuate around Hourly Price Forward Curve[2] F_h:

$$SP_h := \begin{cases} F_h^L - Spike_h^- & u_h < p_h^- \\ F_h \cdot e^{r_h} & p_h^- \leq u_h \leq 1 - p_h^+ \\ F_h^U + Spike_h^+ & u_h > p_h^+ \end{cases} \tag{A.4.1}$$

[2]Especial thanks and credit goes to Dr. F. Paraschiv who has generously provided me with the hourly price forward curve, generated for EEX -Phelix, based on the available information in the market at 01.01.2011.

Table A.4.1: Regime Switching Parameters

Parameter	
p_h^+	probability for upward spike
p_h^-	probability for downward spike
σ_h^2	annual volatility of Gaussian dynamics
λ_h^+	parameter of Exponential distributions for upward spike $Spike_h^+ \sim Exp(\lambda_h^+)$
λ_h^-	parameter of Exponential distributions for downward spike $Spike_h^- \sim Exp(\lambda_h^-)$
α_h	delimiting the upper and lower bounds for the base regime
$r_h \sim N(0, \sigma_h^2)$	deviations from observed hourly price forward F_h
$F_h^L = F_h * e^{-\alpha_h}$	lower limit of the Gaussian dynamics
$F_h^U = F_h * e^{\alpha_h}$	upper limit of the Gaussian dynamics
$u_h \sim U[0, 1]$	

The corresponding parameters are explained in Table A.4.1. These hourly prices are the input for simulating weekly spot and pumping prices and further on together with generated weekly inflows used to construct the Price-Inflow tree. To give an overview of hourly spike behaviors in a week, in Figure A.4.1, the expected magnitude of upward and downward spikes are pictured. It is likely that large upward spikes happen during the evening peak hours of weekdays whereas downward spikes are more expected during day time in the weekends.

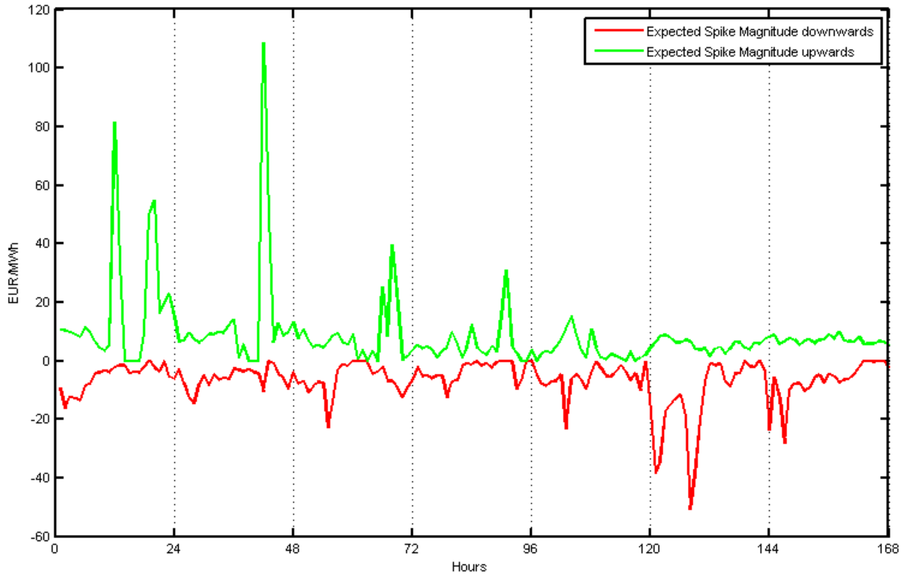


Figure A.4.1: Expected magnitude of hourly spikes (Data from [28], Table 7.)

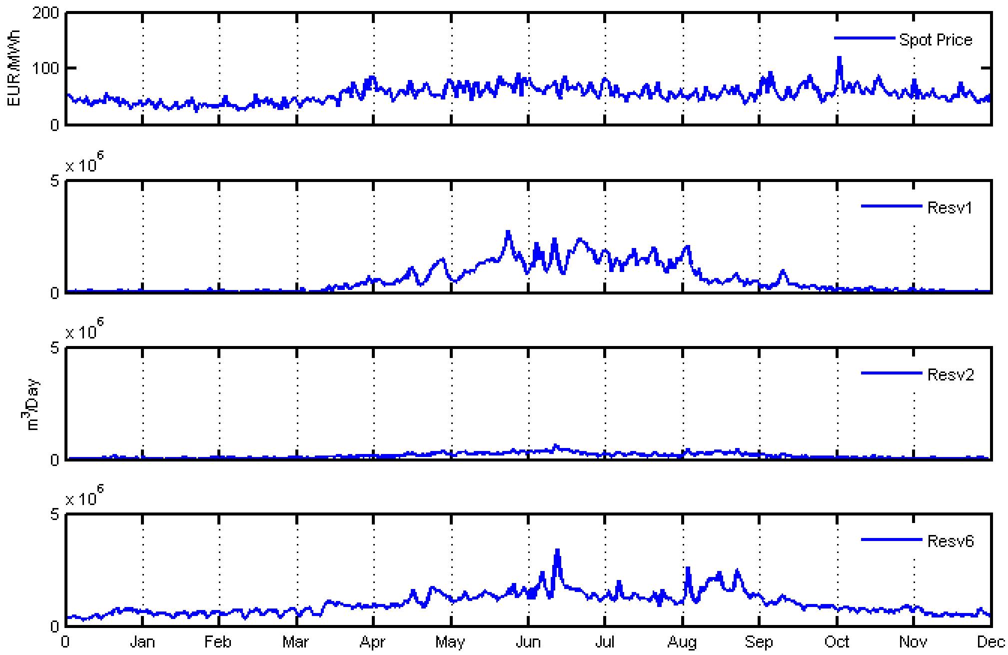


Figure A.4.2: Scaling Requirement for Price-Inflow Tree Generation

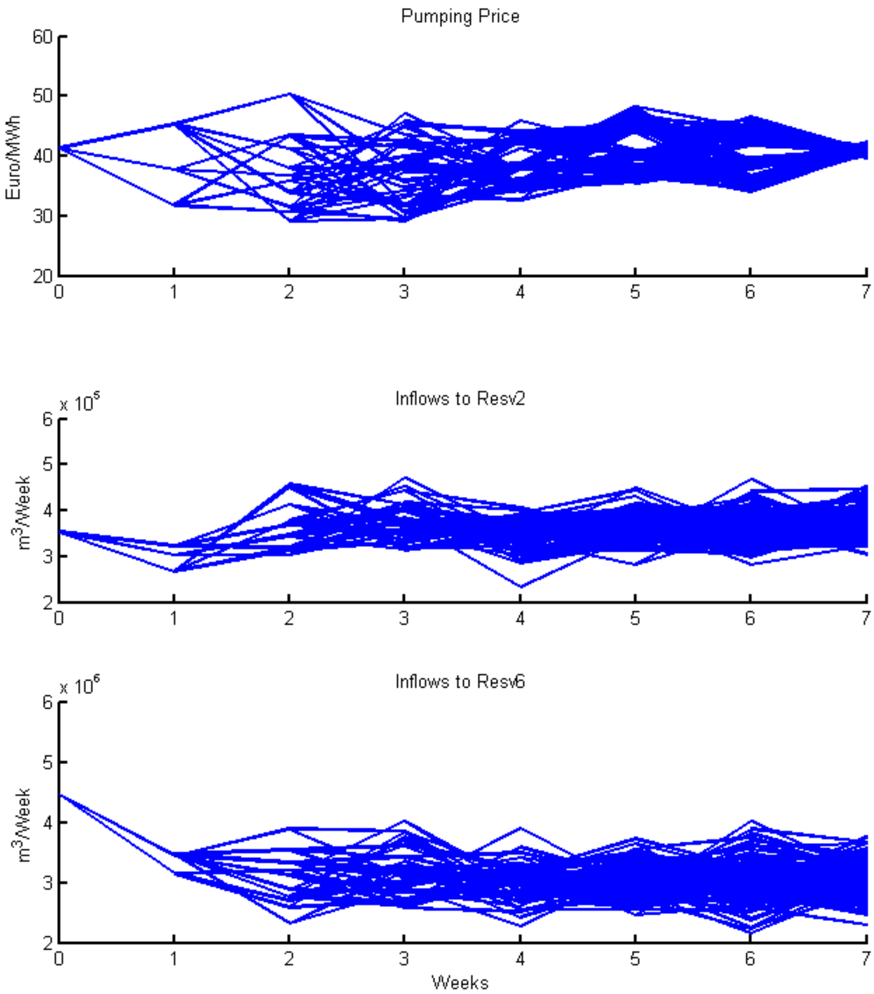


Figure A.4.3: **Top**: Pumping Price component **Middle**: Reservoir 2 Inflows component **Bottom**: Reservoir 6 Inflows component

Bita Analui

Curriculum Vitae

Oskar-Morgenstern-Platz 1
1090 Vienna Austria
✉ *bita.analui@univie.ac.at*

Personal Information

Nationality Iran

Place of birth Tehran

Education

Expected June 2014 **PhD candidate in Statistics and Operations Research**, *University of Vienna*, Vienna, Austria.
Theme: Multistage Stochastic Optimization of Energy Portfolios under Model Ambiguity, Approved January 2012, Supervisor : Prof. Georg Ch. Pflug.

Oct.2008– Dec.2009 **MSc Statistics**, *The University of Sheffield*, Sheffield, United Kingdom.
Thesis with **Merit**: Relationship of Univariate and Multivariate Methods for Selecting Variables in High Dimensional Data. Supervisor: Dr. Nick J. Fieller.

Feb.2005– Jun.2007 **MSc Applied Mathematics**, *Iran University of Science and Technology(IUST)*, Tehran, Iran.
Thesis with **Distinction**: Developing Evaluation Models in Organizations using Data Envelopment Analysis. Supervisor: Dr. Mohammad R. Alirezaee.

Oct.1996– Feb.2000 **BSc. Mathematics**, *Alzahra University*, Tehran, Iran.

Experience

Jun.2012– Present **College Assistant 'IK-Computational Optimization'**, *Institute of Statistics and OR*, University of Vienna.

May2011– May2012 **Project Assistant**, *Institute of Statistics and OR*, University of Vienna.
WWTF Funded project 'Energy Policies in21st Century'

Jul.2010– Feb.2012 **Project Assistant and Programmer**, *Institute of Statistics and OR*, University of Vienna.
Modeling and implementing in the OPL 6.3 in the field of 'Stochastic optimization in power systems' for the project in Siemens Österreich.

Sep.2008– Sep.2009 **MSc Representative**, Department of Probability and Statistics.
University of Sheffield

Jan.2001– Aug.2008 **Statistics Analyst and IT Liaison** , Tehran South West Power Distribution Co..

2005–2008 **Mathematical Modeling and Strategic Management Assistant**, Behin-Cara International Institute of Operations Research.

Publications and Preprints

1- **Bita Analui** and Georg Ch. Pflug, 'On Distributionally Robust Multiperiod Stochastic Optimization' Computational Management Science Journal, April 2014 accepted for publication.
2- **Bita Analui** and Raimund M. Kovacevic, 'Medium Term Hydroelectric Production Planning - a Multistage Stochastic Optimization Model'.Civil Engineering and Infrastructure Journal, 47(1) :139-152, June 2014
3- **Bita Analui**, Mohammad R. Alirezaee and Mohsen Afshariyan, 'Evaluating TFP growth in Iran Electric Power Industry with DEA'. Journal of Economical Research, University of Tehran No. 78: 177-206, June 2007.

Talks and Presentations

1- Medium Term Hydroelectric Production Planning: A Multistage Stochastic Optimization Model. IFIP TC7/2013, Sep. 2013, Klagenfurt, Austria. Joint work with R. Kovacevic.
2- On Robust Multiperiod Stochastic Optimization. ICSP Jul. 2013, Bergamo, Italy.
3- Sensitivity Analysis in Multistage Stochastic Optimization Problems Under Model Ambiguity. VOCAL Dec. 2012, Veszprém, Hungary.
4- Multistage Stochastic Optimization Problems Under Model Ambiguity. 21^{st} ISMP, Aug. 2012, Berlin, Germany.
5- Multistage Stochastic Optimization in Hydroelectric Operation. APMOD Mar. 2012, Paderborn, Germany.Joint work with R. Kovacevic.
6- Stochastic Optimization in Water Resource Management. Center for Hydrometeorology and Remote Sensing, Mar. 2012, University of California Irvine, CA, the USA.
7- Regression Quantiles with errors in variables. Workshop "Modeling in Finance and Energy" Sep. 2010, University of Vienna, Vienna, Austria.
8- Developing Evaluation Models using Data Envelopment Analysis (DEA). APMOD May 2008, Bratislava, Slovakia.
9-Frontier Relationship between Performance Metrics. Euro XXII Jul. 2007, Prague, Czech Rep. Joint work with M. R. Alirezaee.

Grants and Scholarships

- PhD scholarship. IK Computational Optimization, University of Vienna, June 2012-May2014.
- Summer School Grant. 'Numerical analysis Summer school 2012'. 26.6.2012-6.7.2012, Cadarache-France.