DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Robust pricing in insurance and energy markets“

verfasst von / submitted by

Lic. Debora Daniela Escobar Luna Barrera

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

Doctor of Philosophy (PhD)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 794 370 136
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Statistics and Operations Research
Betreut von / Supervisor:	o. Univ.-Prof. Mag. Dr. Georg Pflug

To the students who remind me of the true purpose for academia.

In mathematics,
the art of proposing a question
must be held of higher value than solving it.
Georg Cantor

Acknowledgements

I would like to express my gratitude to my advisor Georg Pflug for giving me the great opportunity to be part of his research group and his warm welcome inviting me to the Opera. These years I met wonderful people and learned so much from all the conferences, winter schools and seminars in which I participated.

From all the people I met in these different academic trips, I would like to express my gratitude to Francesca Maggioni for welcoming me in her research community. To Florentina Paraschiv and Michael Schürle, for all their time spent discussing our project and for making me feel like home in my visits to Trondheim.

In the second half of my journey, I had the pleasure to meet Christa Cuchiero, Thorsten Schmidt and Josef Teichmann in the Freiburg-Wien-Zurich seminar. I am thankful to all of them for having me in this fabulous atmosphere they created to discuss mathematics. I express my gratitude to Irene Klein for inviting me in the first place, for her caring and kind support in different stages of my studies. I would also like to show my gratitude to Josef Teichmann for his support and concern.

Most of my academic trips were possible thanks to the financial support of the University of Vienna. I also would like to thank the non-academic staff of the University, mostly to Lisa Carli for handling all my paperwork at the end of my studies and for being so thoughtful.

To my students, for being so active in class, for your kind words and wishes. You were, are and will be a great inspiration for my job.

Thanks to my colleagues Alexander Ristig and Teresa Scarinci, from whom I learned so much, and for the friendship we have. My dear friend Teresa, for all the weekends we spent working and the coffee breaks. Thank you for being there when I had to celebrate, and when I needed encouragement.

Finally, to Vivian and Alfonso, my beloved parents, for all the lessons learned and empowering me to stay studying all day. For all I missed when I was there and while I am here. Thank you.

Contents

Preface

Most of the activities of financial corporations and other complex industries such as the energy sector are subject to different uncertainties. Certain outcomes of exogenous variables have a financial impact on them, causing great losses if they are not prevented. Although not all the uncertainties can be measured, some of them, such as asset prices, commodity prices or weather forecast among others, are exhaustively studied and modelled. With these models, one could be interested in, for example, maximizing (minimizing) expected profits (losses), on the basis of those random variables. Nevertheless, different statistical methods and samples of past observations result in different models for the same random variable. Each of these models would assign different probabilities to the outcomes of the random variables of interest, resulting in different optimal profits. Although modelling is itself a large part of risk management, different business sectors increased their interest in robust assessment, especially after the financial crisis. This observation leads to the necessity of considering, not one, but several models for our data. In other words, we incorporate model ambiguity for evaluating a concrete problem. Robust prices are optimal prices, whose values offer an achievable profit (loss) under possible severe scenarios. The primary focus of my work is the assessment of complex contracts or even complex real options under model ambiguity. Moreover, I am interested in the estimation of parameters that explain the pricing mechanism from observed prices. To analyse these aspects, I present three different studies of robust pricing with applications in insurance and energy markets.

Different pricing rules are used for pricing losses. In the simple case of the assessment of univariate random variables, a pricing rule is a functional that assigns a positive value to a random variable of losses. A trivial example is the mathematical expectation. Although this functional would summarize the average of losses, it does not provide information about extreme events. More sophisticated pricing rules are to be considered according to the context of the problem. For instance, insurance companies are concerned about pricing losses by measuring their risk. Insurance contracts are priced by following specific pricing methods and principles (see Young 2014 [111] for a general summary). Classic insurance premium principles are the certainty equivalence principle (CEQ) (Von Neumann and Morgenstern 1947 [97], Borch 1961 [13]) and the Wang principle (Wang 2000 [101]). In the first case, the premium charged depends on a utility function, while in the latter the premium depends on a distortion function. Since these contracts are linked to risky losses, risk measures also serve for pricing purposes. In the literature of risk measures, we can find coherent risk measures (Artzner et al. 1999 [4]), spectral risk measures (Acerbi 2002 [2] and Kusuoka 2001 [60]), and deviation measures (Rockafellar et al. 2006 [82]) among others. Each of them is characterized by specific properties. As risk measures, some insurance premium principles are also characterized by properties. In fact, some classic insurance principles are risk measures. In this thesis, I pay particular attention to distortion risk measures, or what is the same, the distortion premium principle. The distortion premium principle (Denneberg 1990 [18]) has been first characterized by a dual theory of risk (Yaari 1987 [109]) and later by axioms (Wang et al. 1997 [102]). From the risk measure theory point of view, distortion risk measures are coherent risk measures (Wirch and Hardy 2001 [106]) and spectral risk measures under some assumptions. Insurance principles and risk measures serve for similar purposes. Their calculation is applied in many aspects of pricing, from insurance contracts, portfolio optimization, dynamic problems and others.

Considering the above, this thesis examines the incorporation of model ambiguity for robust pricing in three different aspects: pricing in the insurance sector, estimation of the pricing mechanism in the market of electricity trading, and internal pricing for the valuation of power plants. Each of them corresponds to a separate chapter.

Chapter 1 contains the results of my first paper. This paper studies the theoretical aspects of risk measures when ambiguity is considered. An overview of different insurance premium principles and risk measures is presented. The main focus of this paper is twofold: the identification of distortion functions and the incorporation of model ambiguity in the distortion premium. Chapter 2 includes a second manuscript. This paper studies the pricing of futures in electricity markets. It focuses on understanding the market in terms of distortion functions and model ambiguity. The inverse problem presented in the first paper is extended here for a modification of the distortion premium, where ambiguity is also a parameter to recover. Chapter 3 is the last paper of this compilation of manuscripts. This paper focuses mainly on the incorporation of ambiguity in the pricing of a real option. The difference with the previous two papers is that here a dynamic system with more than one risk factor is studied. The applications of this study are illustrated for the valuation of power plants.

An overview of the three respective papers is written below.

Chapter 1. The distortion principle for insurance pricing: properties, identification and robustness

This paper is a joint work with Georg Pflug. Published in 2018 (see Escobar and Pflug 2018 [25]).

This paper is organized into four main parts. The first part gives an overview of different insurance pricing rules, such as the certainty equivalence principle (CEQ), the ambiguity principle and the distortion premium. Combinations of them are also reviewed here. Some examples of the distortion premium are presented, such as the

Average Value-at-Risk, the power distortion and the Wang distortion.

The second part is concentrated on theoretical properties of the distortion premium. Possible generalizations of this premium are also discussed here. This premium is obtained by distorting the original probability of the loss random variable and calculated as an expectation w.r.t. the distorted probability. The question we asked ourselves is whether the prices of this premium are close if we give two initial probabilities, which are close to each other. The purpose of a continuity result of such a type is to study robustness against model misspecification. As a distance between probability measures, we choose the Wasserstein distance. We consider the premium defined on a space of distribution functions that assigns a value to each distribution with fixed distortion. We show that this functional is continuous w.r.t. the Wasserstein distance. We also provide a bound for the distance between distorted probabilities in terms of the original probabilities.

The third part is focused on the distortion functions that are applied to the loss distributions. The question we ask here is whether two different distortion functions can give the same distortion premium for a loss distribution. This is a key property if we aim to identify the distortion density. We prove the identifiability of distortions (almost surely). We also propose a methodology for the identification of distortion functions for simple examples: distortion of the Average Value-at-risk and the power distortion. Although the former is a well-known risk measure, it does not use all the information of the loss distribution (Balbás et al. 2009 [5]), hence the importance of including strictly monotone distortion functions.

Finally, the last part of the paper studies robustness of the distortion premium. For a given baseline model, we consider a set of models around the given one. This set is called ambiguity set and we will define it in terms of the Wasserstein distance. Robustness means to be safe in the worst case, therefore the robust distortion premium is obtained as the maximum distortion premium calculated over the ambiguity set. Under some assumptions of the models, the distortion functions and the order of the Wasserstein distance, we prove optimal solutions for this problem. In some cases, we find the distribution that gives the maximum price in the optimization, called the worst case model. For other cases we can approximate it, extending known results (Pichler 2010 [79]). Additionally, we also characterize unboundedness of the calculations of the robust distortion premium.

Chapter 2. Recovering distortion functions in power markets under model ambiguity

This preprint is a joint work with Florentina Paraschiv and Michael Schürle.

The goal of this paper is to fill the gap between electricity futures and hourly electricity prices (spot prices). With this purpose, we first explain the pricing problem of electricity futures. Secondly, we propose and study a pricing rule in terms of distortion functions. We aim at recovering distortion functions among other parameters. To this end we present a new regime-switching model for spot prices. With the parameters estimated from our pricing rule we aim to explain main characteristics of futures prices and the risk premia in this market.

Participants in electricity markets face risks of extreme prices in different commodities, such as coal, oil and gas prices. Although large volumes of electricity are traded in the day-ahead market, producers and consumers trade in the future market for hedging purposes. Producers want to protect themselves against low spot prices, while consumers are concerned about high spot prices. Since electricity is non-storable, the relationship between futures and spot prices has been studied extensively. Many authors examine futures prices by studying the risk premia of these prices (e.g., Benth and Sgarra 2012 [8], Veraart and Veraart 2013 [94]). As an extension to the existent literature, we propose to explain the relationship between futures and spot prices in terms of three different premia: a distortion premium, an information premium and an ambiguity premium.

The distortion premium quantifies the risk aversion of traders in terms of distortion functions. The information premium, which can be seen as a correction factor, is related to the enlargement of information studied in Benth and Meyer Brandis 2009 [7] and in our model it depends on a shifting parameter. The risk premium will be explained by these two premia. In our estimations we also include an ambiguity premium, which models the level of uncertainty. Although insurance tools were already used in this context (as the CEQ premium in Benth et al. 2008 [9]),this is the first study that approaches futures pricing with distortion functions, with a correction factor and model ambiguity. In addition, our goal is to recover these different premia from observed prices and we use non-parametric distortion densities for this. Another highlight is the model for spot prices. This is given in the spirit of Paraschiv et al. 2015 [72]. Different from the latter model, here we specify a stochastic level for the base regime and change the distribution of the spike regimes. This model captures seasonality, it has mean reversion and model spikes.

Although the identification of distortion functions was studied in Escobar and Pflug 2018 [25], in this paper we also identify shifting parameters and ambiguity radii (measured with the Wasserstein distance). We relate our results with traders risk aversion, production costs and model uncertainty. In addition, we find dependence on time-to-delivery and seasonality. The results are linked to other findings, e.g., Bessembinder and Lemmon 2002 [10], Benth et al. 2008 [9] and Lucia and Torró 2011 [65]. Results are shown for base and peak monthly futures traded in the European Energy Exchange (EEX).

Chapter 3. Distributionally robust optimization with multiple time scales: valuation of a thermal power plant

This paper is a joint work with Wim Van Ackooij, Martin Glanzer and Georg Pflug. Submitted in Computational Management Science (CMSC) on April, 2019 [90].

In this paper, we consider ambiguity in a dynamical system (in discrete time), describing the operation of a thermal power plant. The main goal of this paper is the valuation of this real option in discrete time under model ambiguity. The principal characteristics of this problem are the presence of a multiscale model in the optimization and a tractable distance for multistage problems. We apply this general procedure in a specific problem, which is the valuation of a power plant over several weeks.

In general, we have a system where decisions take place in equally spaced moments in time up to a finite horizon. Although decisions are taken, say weekly, uncertainties still take values between stages, affecting the weekly profit and hence the entire valuation. Therefore, we also need to consider a finer scale between stages, where we can simulate paths of the uncertainties. These problems are called multiscale stochastic optimization problems. The valuation of this power plant is a concrete example of this problem, whose setting was provided by the company Électricité de France (EDF). The owner of the power plant decides on the production at the beginning of every week. Within weeks, prices of electricity, fuel, and CO2 allowances influence the overall weekly profit. Each decision made leads to a new state of the power plant, resulting in a dynamic problem. This problem can be solved backwards in time, by optimizing the profit-to-go function stage-wise.

The paper is organized as follows: the first part describes this particular setting for a specific thermal power plant. The second part describes the modelling of uncertainties and its discretization together with an interpolation process called the bridge process. The third part of the paper studies the incorporation of ambiguity. Subsequently, the last part shows the results applied to specific profit functions and production choices.

As for the first part, we describe in detail all the cost functions affecting the weekly profit. We also describe precisely the state variables of the power plant and the decision variables (production profiles). The second part models the three uncertainties following a two-factor model (short and long term). The resulting model follows a three-dimensional log-normal distribution. We quantize this process obtaining a lattice process (see Pflug and Pichler 2014 [76] for discretization in scenario trees). The lattice obtained can be seen as a Markovian process with given transition matrices between stages. We also describe the bridge process, used to link the discretized realizations in consecutive stages by simulating outcomes of the uncertainties on a finer scale. In this case, we extend cases such as Kaut et al. 2014 [57], Werner et al. 2013 [103], among others.

The third part defines the dynamic optimization problem over this Markov process and discusses the incorporation of ambiguity when solving the backward equations. We define a distance for Markov processes - a variant of the Wasserstein distance - called uniform Wasserstein distance. This distance is consistent with the backward equations and its computation reduces to a stage-wise procedure. Besides, the Wasserstein distance is defined with a state-dependent metric, penalizing realizations according to the current state of the system. The last part of this paper puts together all the previous results to solve the valuation of this specific power plant under model ambiguity. For each ambiguity radius, the solution of this problem gives a probabilistic tree of production profiles (among the pre-fixed production choices). This is associated with a probabilistic tree of profits. We compare the stage-wise probabilities of the optimal production profiles when ambiguity radii increase over time. In general, we show the larger the ambiguity radius, the more conservative the production and the less the overall valuations are.

Individual contribution

Chapter 1

Chapter 1 includes my first joint paper with Georg Pflug. This paper has been published in Annals of Operations Research (see Escobar and Pflug 2018).

In the beginning of my PhD I attended a seminar my advisor taught on risk measures and optimal transportation. I started to research different risk measures and found special interest in distortion risk measures. I continued investigating theoretical properties of this risk measure and proved the continuity of the distortion premium w.r.t. the Wasserstein distance. These results where the basis for solving the robust distortion premium under ambiguity sets measured with the Wasserstein distance. Later, I generalized these results and presented them to my advisor. Georg Pflug supervised my work and we decided to write our first paper on theoretical aspects of distortion risk measures. I also studied the continuity of this premium w.r.t. the distortion function. To enclose this part, we decided to add empirical results for the identification of distortion functions. We discussed the different options for the methodology and I am also responsible for the computational part of it. Regarding the writing of the paper, I prepared the initial draft and the consecutive versions of it. My advisor helped me with some examples and the Introduction. The final version is the product of more results we discussed after the comments of the reviewers, to whom we are very grateful.

Chapter 2

The subsequent chapter is a joint work with Florentina Paraschiv and Michael Schürle, with current status as preprint.

This paper follows the previous chapter as an extended application in a real market. The topic is inspired by a proposal my advisor made at the beginning of my PhD studies. He posed the question of explaining the dynamics of the risk premia in the electricity market. Given my interest in the distortion premium, I decided to study the changes in the risk premia in terms of this principle. After writing the first paper I had the pleasure to meet Florentina Paraschiv and Michael Schürle in Norway. I decided to resume this work together with them. We organized several meetings and exchanged our thoughts frequently on the paper. The individual contributions of this paper can be separated into two parts. Firstly, the theory on the pricing properties together with the computational results of the identification. Secondly, the theoretical construction of the model for spot prices and its calibration. I am responsible for the first part and to put together all our results in the different versions of this paper. As my co-authors have a strong background in energy markets, I attribute to them the construction and computational results for the calibration of the model, who also provided the data. Regardless of this separation, we all reviewed every step of the paper and discussed the interpretations of the results.

Chapter 3

Chapter 3 is a joint work with Wim Van Ackooij, Martin Glanzer and Georg Pflug. This paper was submitted in Computational Management Science on April, 2019.

My advisor and I applied to participate in the Programme Gaspard Monge (PGMO), which supports collaboration between academia and industry for solving specific problems related to the energy sector. This chapter is a compilation of results of the project we did together with my colleague Martin Glanzer and Wim Van Ackooij from Électricité de France (EDF). Wim Van Ackooij provided us with real data and the benchmark of the problem we solve here. From our first meeting in Paris, we all participated in vivid discussions on the project. Wim Van Ackooij helped us with his practitioner's insight and together with Georg Pflug's expertise, we defined a distance that accounts for specific details in our problem. The paper can be separated into two parts. On the one hand, it concerns the construction of the stochastic model for the uncertainties, the quantization and an interpolation process. On the other hand, the paper defines different aspects of a power plant and focuses on solving a dynamic optimization problem under model ambiguity for this real option. The main contributor to the first part is Martin Glanzer while I am the main contributor to the second part. The computational codes, results and writing of the second part were my primary tasks. I received frequent feedback from all of my co-authors for my part. Despite the separate tasks, we all discussed the paper. As this paper is another example of robust pricing we decided to include it in my thesis.

CHAPTER 1

The distortion principle for insurance pricing: properties, identification and robustness

Joint work with Georg Ch. Pflug

Abstract

Distortion (Denneberg 1990) is a well known premium calculation principle for insurance contracts. In this paper, we study sensitivity properties of distortion functionals w.r.t. the assumptions for risk aversion as well as robustness w.r.t. ambiguity of the loss distribution. Ambiguity is measured by the Wasserstein distance. We study variances of distances for probability models and identify some worst case distributions. In addition to the direct problem we also investigate the inverse problem, that is how to identify the distortion density on the basis of observations of insurance premia.

Keywords— Ambiguity · Distortion premium · Dual representation · Premium principles · Risk measures · Wasserstein distance.

1.1 Introduction

The function of the insurance business is to carry the risk of a loss of the customer for a fixed amount, called the premium. The premium has to be larger than the expected loss, otherwise the insurance company faces ruin with probability one. The difference between the premium and the expectation is called the *risk premium*. There are several principles, from which an insurance premium is calculated on the basis of the loss distribution.

Let X be a (non-negative) random loss variable. Traditionally, an insurance premium is a functional, $\pi : \{X \geq 0 \text{ defined on } (\Omega, \mathcal{F}, P)\} \to \mathbb{R}_{\geq 0}$. We will work with functionals that depend only on the distribution of the loss random variable (sometimes called law-invariance or version-independence property, Young 2014 [111]). If X has distribution function F we use the notation $\pi(F)$ for the pertaining insurance premium, and $\mathbb{E}(F)$ for the expectation of F. We use alternatively the notation $\pi(F)$ or $\pi(X)$, resp. $\mathbb{E}(F)$ or $\mathbb{E}(X)$ whenever it is more convenient. To the extent of the paper, a more specific notation is used for particular cases of the premium.

We consider the following basic pricing principles:

- The distortion principle (Denneberg 1990 [18]).
- The certainty equivalence principle (v Neumann and Morgenstern 1947 [97]).
- The ambiguity principle (Gilboa and Schmeidler 1989 [44]).
- Combinations of the previous (for instance Luan 2001 [63]).

1.1.1 The distortion principle

The distortion principle is related to the idea of stress testing. The original distribution function F is modified (distorted) and the premium is the expectation of the modified distribution. If $g : [0, 1] \to \mathbb{R}$ is a concave monotonically increasing function with the property $g(0) = 0$, $g(1) = 1$, then the distorted distribution F^g is given by

$$F^g(x) = 1 - g(1 - F(x)).$$

The function g is called the *distortion function* and

$$h(v) = g'(1 - v),$$

with g' being the derivative of g, is the *distortion density*.[1] Notice that h is a density in $[0, 1]$. We denote by $H(u) = \int_0^u h(v)\, dv$ the *distortion distribution*. Since the assumptions imply that $g(x) \geq x$ for $0 \leq x \leq 1$, $F^g \leq F$, i.e. F^g is first order

[1] The derivative of a concave function is a.e. defined, even if it is not differentiable everywhere.

stochastically larger than F.[2] The distortion premium is the expectation of F^g

$$\pi_h(F) = \int_0^\infty g(1 - F(x))\, dx \geq \int_0^\infty (1 - F(x))\, dx = \mathbb{E}(X).$$

By a simple integral transform, one may easily see that the premium can equivalently be written as

$$\pi_h(F) = \int_0^1 F^{-1}(v)\, h(v)\, dv = \int_0^1 \mathrm{V@R}_v(F)\, h(v)\, dv, \tag{1.1}$$

where $\mathrm{V@R}_v(F) = F^{-1}(v)$, the quantile function. Note that a functional of this form is called an L-estimates (Huber 2011 [53]). If the random variable X takes as well negative values, we could generally define the premium as a *Choquet integral*

$$\pi_h(F) = \int_{-\infty}^0 g(1 - F(x)) - 1\, dx + \int_0^\infty g(1 - F(x))\, dx. \tag{1.2}$$

In principle, any distortion function which is monotonic and satisfies $g(u) \geq u$ is a valid basis for a distortion function. However, the concavity of g guarantees that the pertaining distortion density h is increasing, which - in insurance application - reflects the fact that putting aside risk capital gets more expensive for higher quantiles of the risk distribution. Nondecreasing distortion functions lead to non-negative distortion densities with the consequence that

$$\pi_h(F_1) \leq \pi_h(F_2) \qquad \text{whenver } F_2 \text{ is stochastically larger than } F_1.$$

Relaxing the monotonicity assumption for g would violate in general the monotonicity w.r.t. first stochastic order.

1.1.2 Examples of distortion functions

Widely used distortion functions g resp. the pertaining distortion densities h are

- the power distortion with exponent s. If $0 < s < 1$,

  $$g^{(s)}(v) = v^s, \qquad h^{(s)}(v) = s(1 - v)^{s-1}. \tag{1.3}$$

  The premium is known as the Proportional Hazard transform (Wang 1995 [99]) and calculated as

  $$\pi_{h^{(s)}}(F) = \int_0^\infty 1 - F(x)^s\, dx = s \int_0^1 F^{-1}(v)(1 - v)^{s-1}\, dv. \tag{1.4}$$

[2] F_1 is first order stochastically larger than F_2 if $F_1(x) \leq F_2(x)$ for all x.

If $s \geq 1$, then we take

$$g^{(s)}(v) = 1 - (1-v)^s, \qquad h^{(s)}(v) = sv^{s-1}. \tag{1.5}$$

The premium is

$$\pi_{h^{(s)}}(F) = \int_0^\infty 1 - (1 - F(x))^s \, dx = s \int_0^1 F^{-1}(v) v^{s-1} \, dv. \tag{1.6}$$

If we consider integer exponent, the premium has a special representation.

Proposition 1.1. *Let $X^{(i)}$, $i = 1, \ldots, n$ be independent copies of the random variable X, then the power distortion premium with integer power s has the representation*

$$\pi_{h^{(s)}}(X) = \mathbb{E}\left(\max\{X^{(1)}, \ldots, X^{(s)}\}\right).$$

Proof. Let F be the distribution of X. The power distortion premium for integer power s is computed with $g^{(s)}$ in (1.5) and by definition

$$\pi_{h^{(s)}}(F) = \int_0^\infty g^{(s)}(1 - F(x)) = \int_0^\infty 1 - F(x)^s \, dx.$$

The assertion follows from the fact that the distribution function of the random variable $\max\{X^{(1)}, \ldots, X^{(s)}\}$ is $F(x)^s$. $\square$

Finally, notice that the distortion density is bounded for $s \geq 1$, but unbounded for $0 < s < 1$.

- the Wang distortion or Wang transform (Wang 2000 [101])

$$g(v) = \Phi\left(\Phi^{-1}(v) + \lambda\right), \qquad h(v) = \frac{\varphi(\Phi^{-1}(1-v) + \lambda)}{\varphi\left(\Phi^{-1}(1-v)\right)}, \quad \lambda > 0,$$

where Φ is the standard normal distribution and φ its density.

- the AV@R (average value-at-risk) distortion function and density are

$$g_\alpha(v) = \min\left\{\frac{v}{1-\alpha}, 1\right\}, \qquad h_\alpha(v) = \frac{1}{1-\alpha} \mathbb{1}_{v \geq \alpha}, \tag{1.7}$$

where $0 \leq \alpha < 1$. The pertaining premium has different names, such as conditional tail expectation (CTE), CV@R (conditional value at risk) or ES (expected shortfall) (Embrechts et al. 1997 [22]). The premium is

$$\pi_{h_\alpha}(F) = \int_0^\infty \min\left\{\frac{1 - F(x)}{1-\alpha}, 1\right\} dx = \frac{1}{1-\alpha} \int_\alpha^1 F^{-1}(v) \, dv. \tag{1.8}$$

- piecewise constant distortion densities. The insurance industry uses also piecewise constant increasing distortion functions. For example, the following distortion function is used by a large reinsurer.

v	$h(v)$	v	$h(v)$
[0,0.85]	0.8443	[0.988,0.992)	3.6462
[0.85,0.947)	1.1731	[0.992,0.993)	4.0572
[0.947,0.965)	1.4121	[0.993,0.996)	6.5378
[0.965,0.975)	1.7335	[0.996,0.996)	12.7020
[0.975,0.988)	2.4806	[0.996,1]	14.9436.

For more examples on different choices of h and also for different families of distributions, see Wang 1996 [100] and Furman and Zitikis 2008 [38].

1.1.3 Certainty equivalence principle

Let V be a convex, strictly monotonic disutility function.[3] The certainty equivalence premium is the solution of

$$V(\pi) = \mathbb{E}(V(X)),$$

i.e. it is obtained by equating the disutility of the premium and the expected disutility of the loss. The premium is written as follows

$$\pi^V(F) = V^{-1}\left(\mathbb{E}(V(X))\right) = V^{-1}\left(\int_0^1 V\left(F^{-1}(v)\right)\, dv\right).$$

By Jensen's inequality $\pi^V(F) \geq \mathbb{E}(F)$. Examples for disutilities V are the power utility $V(x) = x^s$ for $s \geq 1$ or the exponential utility $V(x) = \exp(x)$.

Related to this premium, one could consider just the expected value and compute the expected disutility (Borch 1961 [13]) obtaining

$$\pi(F) = \mathbb{E}(V(X)). \tag{1.9}$$

For generalizations of the CEQ premium see Vinel and Krokhmal 2017 [96].

1.1.4 The ambiguity premium

Let $\mathfrak{F}$ be a family of distributions, which contains the "most probable" loss distribution F. The ambiguity insurance premium is

$$\pi^{\mathfrak{F}}(F) = \sup\{\mathbb{E}(G) : G \in \mathfrak{F}\}.$$

$\mathfrak{F}$ is called the *ambiguity set*. In an alternative, but equivalent notation, the ambiguity premium is given by

$$\pi^{\mathcal{Q}}(X) = \max\left\{\mathbb{E}_Q(X) \,:\, Q \in \mathcal{Q}\right\}, \tag{1.10}$$

[3]The original notion of a utility function introduced by Neumann/Morgenstern was a concave monotonic U, such that the decision maker maximizes the expectation $\mathbb{E}(U(Y))$ of a profit variable Y. A disutility function can be defined out of a utility function by setting $V(u) = -U(-u)$.

where $\mathcal{Q}$ is a family of probability models containing the baseline model P. The functional inside the maximization needs not to be the expectation, but can be general, see e.g. Wozabal 2012 [107], Wozabal 2014 [108], Gilboa and Schmeidler 1989 [44] and our Section 1.6.

Remark 1.1. *In their seminal paper from 1989, Gilboa and Schmeidler [44] give an axiomatic approach to extended utility functionals of the form*

$$\min \left\{ \mathbb{E}_Q(U(Y)) \,:\, Q \in \mathcal{Q} \right\},$$

where U is a utility function and Y is a profit variable. For the insurance case, U should be replaced by a disutility function V and Y should be replaced by a loss variable X leading to an equivalent expression

$$\max \left\{ \mathbb{E}_Q(V(X)) \,:\, Q \in \mathcal{Q} \right\}.$$

The link to (1.10) is obvious and it can be seen as a combination of expected disutility (1.9) and ambiguity.

Remark 1.2. *Recall the fundamental pricing formula of derivatives in financial markets states that the price can be obtained by taking the maximum of the discounted expected payoffs, where the maximum is taken over all probability measures, which make the discounted price of the underlying a martingale. This can be seen as an ambiguity price.*

The ambiguity premium is characterized by the choice of the ambiguity set $\mathfrak{F}$. In principle, this set can be arbitrary given as long as it contains F. Convex premium functionals have a dual representation, which are also in the form of an ambiguity functional. For distortion functionals, this will be illustrated in the next section. Other important examples for ambiguity premium prices can be defined through distances for probability distributions. Let D be such a distance, then an ambiguity set is given by

$$\mathfrak{F} = \{G : \, D(F,G) \le \epsilon\},$$

with ambiguity premium

$$\pi_D^\epsilon(F) = \max\{\mathbb{E}(G) : D(F,G) \le \epsilon\}.$$

We call ϵ the *ambiguity radius.* This radius quantifies not only the risk premium, but also the model uncertainty, since the real distribution is typically not exactly known and all we have is a baseline model F. In our Section 1.6 we base ambiguity models on the Wasserstein distance WD.

1.1.5 Combined models

Luan 2001 [63] introduced a combination of distortion and certainty equivalence premium prices by defining a variable W distributed according to F^g and setting

$$\pi_h^V(F) = V^{-1}(\mathbb{E}[V(W)]) = V^{-1}\left(\int_0^1 V\left(F^{-1}(v)\right) h(v)\, dv\right).$$

Notice that $(F^g)^{-1}(v) = F^{-1}(1 - g^{-1}(1 - v))$.

More generally, one may also add ambiguity respect to the model and set

$$\pi_h^{V,\epsilon}(F) = \sup\left\{ V^{-1}\left(\int_0^1 V\left(G^{-1}(v)\right)\, h(v)\, dv\right) : D(F,G) \le \epsilon \right\}. \tag{1.11}$$

Notice that (1.11) contains all previous definitions by making some of the following parameter settings

$$h(v) = 1, V(v) = v, \epsilon = 0.$$

If all parameters are set like that, we recover the expectation.

We could also consider the expected disutility premium (1.9) and combine it with the distortion premium,

$$\int_0^1 V(F^{-1}(v))\, h(v)\, dv = \mathbb{E}[V(W)].$$

Section 1.6 will be dedicated to study the combination of distortion and ambiguity premium prices.

As to notation, we denote by $\mathcal{L}^p$ the space of all random variables with finite p-norm for all $p \ge 1$

$$\|X\|_p = [\mathbb{E}(|X|^p)]^{1/p},$$

resp. $\|X\|_\infty = \operatorname{ess\,sup}(|X|)$, the essential supremum. The same notation is used for any real valued function on $[0, 1]$ and p and q are conjugates if $1/p + 1/q = 1$.

1.2 The distortion premium and generalizations

The characterization and represestations of the distortion premium were studied exhaustively. Among some of the most classic contributions we mention the dual theory of Yaari 1987 [109]; and the characterization by axioms of this premium developed in Wang, et al. 1997 [102], where the power distortion for $0 < s < 1$ is also characterized in a unique manner. A summary of other known representations and new generalization of this premium will be presented below. Recall that any mapping $X \mapsto \pi(X)$ which is monotone, convex and fulfils translation equivariance[4] is a risk measure. Furthermore, if π is also positively homogeneous, monotonic w.r.t. the first stochastic order and subadditive[5], then it is a *coherent* risk measure (Artzner et al. 1999 [4]). The distortion premium fulfils all these properties, therefore by the Fenchel-Moreau-Rockefellar Theorem, it has a dual representation.

Theorem 1.1. *(see Pflug 2006 [73]) The dual representation of the distortion pre-*

[4] π has translation equivariance property, if $\pi(X + c) = \pi(X) + c$, for $c \in \mathbb{R}$.

[5] A premium π is called subadditive, if $\pi(X + Y) \le \pi(X) + \pi(Y)$. Subadditivity and positive homogeneity imply convexity.

mium with distortion density h is given by

$$\pi_h(X) = \sup\{\mathbb{E}(X \cdot Z) : Z = h(U), \text{ where } U \text{ is uniformly distributed on } [0,1]\}.$$

Note that all admissible Z's in Theorem 1.1 are densities on [0,1], since $h \geq 0$ and $\mathbb{E}(h(U)) = 1$. To put it differently, given X defined on $(\Omega, \mathcal{F}, P)$ and let $\mathcal{Q}$ be the set of all probability measures on $(\Omega, \mathcal{F})$ such that the density $\frac{dQ}{dP}$ has distribution function H, the distortion distribution, then

$$\pi_h(X) = \sup\{\mathbb{E}_Q(X) : Q \in \mathcal{Q}\}.$$

Therefore, every distortion premium can be seen as well as an ambiguity premium with $\mathcal{Q}$ as the ambiguity set.

Let us look into more detail to the special case of the AV@R premium. In this case, the dual representation specializes to

$$\pi_{h_\alpha}(X) = \sup\{\mathbb{E}(X \cdot Z) : 0 \leq Z \leq \frac{1}{1-\alpha}; \mathbb{E}(Z) = 1\}.$$

From the previous representation, we can see that the AV@R-distortion densities h_α are the extremes of the convex set of all distortion densities. This fact implies that any distortion premium can be represented as mixtures of AV@R's, such representations are called Kusuoka representations (Kusuoka 2001 [60], Jouini et al. 2006 [56]). Coherent risks have a Kusuoka representation of the form

$$\pi(F) = \sup_{K \in \mathcal{K}} \int_0^1 \text{AV@R}_\alpha(F)\, dK(\alpha),$$

where $\mathcal{K}$ is a collection of probability measures in [0,1]. In particular, for the distortion premium we have the following result (Pflug/Römisch 2007 [77]).

Theorem 1.2. *Any distortion premium can be written as*

$$\pi_h(F) = \int_0^1 AV@R_\alpha(X)\, dK(\alpha).$$

The mixture distribution K is given by the way how h is represented as a mixture of the AV@R-distortion densities, i.e.

$$h(v) = \int_0^v \frac{1}{1-\alpha}\, dK(\alpha).$$

The pure AV@R$_\beta$ is contained in this class by setting $K(\alpha) = \delta_\beta$, the Dirac measure at β. Moreover, the integral of the AV@R's is obtained for $K(\alpha) = \alpha$ and is defined as

$$\int_0^1 \text{AV@R}_\alpha(F)\, d\alpha = \int_0^1 F^{-1}(v)\, [-\log(1-v)]\, dv,$$

if the integral exists.

Remark 1.3. *Some other generalizations of the distortion premium were studied in Greselin and Zitikis 2018 [49], where they consider a class of functionals*

$$\int_0^1 \nu(AV@R_\alpha(X), AV@R_0(X))\, d\alpha,$$

with $\nu(\cdot,\cdot)$ an integrable function and show the Gini-index and Bonferroni-index belong to this class. These generalizations lead to inequality measures instead of risk measures.

As a related generalization of the distortion premium one may consider

$$R(X) = \int_0^1 \nu(\mathrm{AV@R}_\alpha(X))\, k(\alpha)\, d\alpha, \tag{1.12}$$

for some convex and monotonic Lipschitz function ν and some non-negative function k on [0,1]. Clearly, $R(X)$ is convex and monotonic, but in general is neither positively homogeneous nor translation equivariant unless ν is the identity (see Appendix A for a proof). To our knowledge, functionals of the form (1.12) are not used in the insurance sector. For this and some other generalizations see the papers of Goovaerts et al. 2004 [47] and Furman and Zitikis 2008 [38].

1.3 Continuity of the premium w.r.t. the Wasserstein distance

In this section we study sensitivity properties of the distortion premium respect to the underlying distribution. Some results in this section are related to those in Pichler 2013 [80], Pflug and Pichler 2014 [76] and Kiesel et al. 2016 [59]. Similar results of continuity for variability measures are studied in Furman et al. 2017 [39]. To start, we recall the notion of the Wasserstein distance.

Definition 1.1. *Let (Ω, d) be a metric space and P, $\tilde{P}$ be two Borel probability measures on it. Then the Wasserstein distance of order $r \geq 1$ is defined as*

$$WD_{r,d}(P,\tilde{P}) = \left(\inf_{\substack{X\sim P\\ Y\sim\tilde{P}}} \mathbb{E}\left(d(X,Y)^r\right) \right)^{1/r}.$$

Here the infimum is over all joint distributions of the pair (X, Y), such that the marginal distributions are P resp. $\tilde{P}$, i.e. $X \sim P$, $Y \sim \tilde{P}$.

For two distributions F and G on the real line endowed with metric

$$d_1(x,y) = |x-y|.$$

this definition specializes to (see Vallender 1974 [89])

$$WD_{1,d_1}(F,G) \;=\; \int_{-\infty}^{\infty} |F(x) - G(x)|\, dx = \int_0^1 |F^{-1}(v) - G^{-1}(v)|\, dv.$$

Therefore, the Wasserstein distance is the (absolute) area between the distribution functions which is also the (absolute) area between the inverse distributions. By a similar argument one may prove that the Wasserstein distance of order $r \geq 1$ with the d_1 metric on the real line is

$$WD^r_{r,d_1}(F,G) = \int_0^1 |F^{-1}(v) - G^{-1}(v)|^r\, dv = \|F^{-1} - G^{-1}\|_r^r. \tag{1.13}$$

We now study continuity properties of the functional $F \mapsto \pi_h(F)$.

Proposition 1.2 (Continuity for bounded distortion densities)**.** *Let F and G be two distributions on the real line and h a distortion density function. If the distributions have both finite first moments and h is bounded, then*

$$|\pi_h(F) - \pi_h(G)| \leq ||h||_\infty \cdot WD_{1,d_1}(F,G).$$

Proof. See Pichler 2010 [79]. □

Remark 1.4. *The boundedness of h is ensured if g has a finite right hand side derivative at 0, and also if g has finite Lipschitz constant L, since $\|h\|_\infty \leq L$.*

Proposition 1.2 can be easily generalized as follows.

Proposition 1.3 (Continuity for distortion densities in $\mathcal{L}^q$ for $q < \infty$)**.** *Let F and G be two distributions on the real line and h a distortion density function. If F, G have finite $p-$moments and $h \in \mathcal{L}^q$, then*

$$|\pi_h(F) - \pi_h(G)| \leq ||h||_q \cdot WD_{p,d_1}(F,G),$$

where p and q are conjugates.

Proof. By Hölder's inequality for p and q we obtain

$$\begin{aligned} |\pi_h(F) - \pi_h(G)| &= \left| \int_0^1 h(v) \cdot \left(F^{-1}(v) - G^{-1}(v)\right)\, dv \right| \\ &\leq \left(\int_0^1 |h(v)|^q\, dv \right)^{1/q} \cdot \left(\int_0^1 \left|F^{-1}(v) - G^{-1}(v)\right|^p\, dv \right)^{1/p} \\ &\leq ||h||_q \cdot WD_{p,d_1}(F,G). \end{aligned}$$

□

Example 1.1. *Let F and G be two distributions with finite first moments.*

- *For the AV@R distortion premium* $||h_\alpha||_\infty = \frac{1}{1-\alpha}$, *and therefore*

$$|\pi_{h_\alpha}(F) - \pi_{h_\alpha}(G)| \leq \frac{1}{1-\alpha} \cdot WD_{1,d_1}(F,G).$$

- *For the power distortion with* $s \geq 1$, $||h^{(s)}||_\infty = s$, *and therefore*

$$|\pi_{h^{(s)}}(F) - \pi_{h^{(s)}}(G)| \leq s \cdot WD_{1,d_1}(F,G).$$

The power distortion with $0 < s < 1$ is not bounded. The next result is dedicated for this particular case.

Proposition 1.4 (Continuity for the the power distortion with $0 < s < 1$)**.** *Let F and G be distribution functions and $h^{(s)}$ the distortion density defined in (1.3). If F and G have finite $p-$moments for $p > \frac{1}{s}$ and $h \in \mathcal{L}^q$, then*

$$|\pi_{h^{(s)}}(F) - \pi_{h^{(s)}}(G)| \leq \frac{s}{\sqrt[q]{1+q\,(s-1)}} \cdot WD_{p,d_1}(F,G),$$

where p and q are conjugates.

Proof. We first note that $p > \frac{1}{s}$ implies $q < \frac{1}{1-s}$ and let $t = 1 + q\,(s-1) > 0$.

$$\begin{aligned}
\left(\int_0^1 h^{(s)}(v)^q \, dv\right)^{1/q} &= \left(\int_0^1 s^q \cdot (1-v)^{q\cdot(s-1)} \, dv\right)^{1/q} \\
&= \left(\int_0^1 s^q \cdot (1-v)^{t-1} \, dv\right)^{1/q} \\
&= \frac{s}{\sqrt[q]{t}} \cdot \left(\int_0^1 t\,(1-v)^{t-1} \, dv\right)^{1/q} = \frac{s}{\sqrt[q]{t}}.
\end{aligned}$$

Proposition 1.3 proves the statement. □

The next result is a direct consequence of Proposition 1.4.

Corollary 1.1 (Continuity for distortion densities dominated by power distortion densities with $0 < s < 1$)**.** *Let F and G be distribution functions and h a distortion density. If h is such that $h(v) \leq c \cdot h^{(s)}(v)$, for all $v \in [0,1]$, $c > 0$ and $0 < s < 1$, F and G have finite $p-$moments for $p > \frac{1}{s}$, then $h \in \mathcal{L}^q$ and*

$$|\pi_h(F) - \pi_h(G)| \leq \frac{c \cdot s}{\sqrt[q]{1+q\,(s-1)}} \cdot WD_{p,d_1}(F,G),$$

where p and q are conjugates.

Corollary 1.2 (Convergence)**.** *If F, F_n for all $n \geq 1$ have finite uniformly bounded $p-$moments, $h \in \mathcal{L}^q$ and $WD_{p,d_1}(F_n, F) \to 0$ as $n \to \infty$, then*

$$|\pi_h(F) - \pi_h(F_n)| \xrightarrow[n\to\infty]{} 0,$$

where p and q are conjugates.

Remark 1.5. *Corollary 1.2 holds when the sequence of distributions are the empirical distributions $\widehat{F}_n$ defined on an i.i.d. sample of size n, $(x_1, \ldots, x_n)$ from $X \sim F$. If F has finite $p-$moments, then $WD_{p,d_1}(\widehat{F}_n, F) \xrightarrow[n\to\infty]{} 0$, hence $\left|\pi_h(\widehat{F}_n) - \pi_h(F)\right| \xrightarrow[n\to\infty]{} 0$. This result follows by applying Lemma 4.1 in Pflug and Pichler [76].*

Finally notice that, for continuity, the order of the Wasserstein distance r coincides with the number of finite moments of F.

1.3.1 Partial coverage

Many insurance contracts do not guarantee complete indemnity, but their payoff is just a part of the full damage. Such contracts include proportional insurance, deductibles and capped insurance. In general, there is a (monotonic) payoff function T such that the payoff is $T(X)$, if the total loss is X. A quite flexible form is for instance the excess-of-loss insurance (XL-insurance), which has a payoff function

$$T(x) = \begin{cases} 0 & \text{if } x \leq a \\ x - a & \text{if } a \leq x \leq e \\ e - a & \text{if } x \geq e. \end{cases} \tag{1.14}$$

Denote by F^T the distribution of $T(X)$, if F is the distribution of X. The distortion premium for partial coverage is $\pi_h(F^T)$. We study the relationship between F^T and G^T as well as between $\pi_h(F^T)$ and $\pi_h(G^T)$ in a slightly more general setup, namely for Hölder continuous T. Recall that T is Hölder continuous with constant H_β, if $|T(x) - T(y)| \leq H_\beta \cdot |x - y|^\beta$, for some $\beta \leq 1$.

Theorem 1.3 (Distance between the original and image probabilities by T)**.** *Let P and Q be two probability measures and consider their image probabilities under T denoted by P^T and Q^T, respectively. If T is a $\beta-$Hölder continuous mapping, then*

$$WD_{r_\beta,d_1}(P^T, Q^T) \leq H_\beta \cdot WD^\beta_{r,d_1}(P, Q),$$

for $r_\beta = \frac{r}{\beta} \geq 1$ and $r \geq 1$, where H_β is the $\beta-$Hölder constant.

Proof. Let the joint distribution of X and Y such that

$$WD_{r,d_1}(X, Y) = \mathbb{E}^{1/r}(|X - Y|^r),$$

then

$$\begin{aligned} WD^{r_\beta}_{r_\beta,d_1}(P^T, Q^T) &\leq \mathbb{E}(|T(X) - T(Y)|^{r_\beta}) \\ &\leq H_\beta^{r_\beta} \cdot \mathbb{E}(|X - Y|^r) = H_\beta^{r_\beta} \cdot WD^r_{r,d_1}(P, Q). \end{aligned}$$

Taking the r_β root on both sides finished the proof. ☐

For the XL-insurance, the Hölder-constant is a Lipschitz constant ($\beta = 1$) and has the value 1.

From the previous Theorem we can conclude that, if two probabilities are close, then the image probabilities by a mapping T with the characteristics of Theorem 1.3, are close in Wasserstein distance as well. Theorem 1.3 isolates the argument also used in Theorem 3.31 in [76]. Note that the underlying distances for the Wasserstein distances are the metrics of the respective spaces.

Corollary 1.3. *Let F, G be two distributions defined by the probabilities P and Q, respectively, and F^T, G^T be their image distributions by T, respectively. If T is a $\beta-$Hölder continuous mapping with constant H_β, $h \in \mathcal{L}^q$, the distributions F^T, G^T with finite $p-$moments, then for all $r = p \cdot \beta$ ($r \geq 1$), the distortion premium with payment function T satisfies*

$$|\pi_h(F^T) - \pi_h(G^T)| \leq ||h||_q \cdot WD_{p,d_1}(P^T, Q^T) \leq ||h||_q \cdot H_\beta \cdot WD^\beta_{r,d_1}(P, Q). \quad (1.15)$$

We proceed now to study sensitivity properties of the distortion premium w.r.t. the distortion density.

1.4 Continuity of the premium w.r.t. the distortion density

Previously, we studied the mapping $F \mapsto \pi_h(F)$ for fixed h. In this section, we consider and present properties of the mapping $h \mapsto \pi_h(F)$ for fixed F. Different sensitivity properties w.r.t. the distortion parameters were studied in Gourieroux and Liu 2006 [48].

Proposition 1.5 (Continuity of the distortion premium w.r.t. the distortion density h)**.** *Let F be a distribution and consider two different distortion densities h_1, h_2. If F has finite $p-$moments and $h_1, h_2 \in \mathcal{L}^q$, then*

$$|\pi_{h_1}(F) - \pi_{h_2}(F)| \leq ||F^{-1}||_p \cdot ||h_1 - h_2||_q,$$

where p and q are conjugates. Here the choices $p = 1$, $q = \infty$ and $p = \infty$, $q = 1$ are included.

Proof. Use Hölder inequality and the result is direct. □

We can conclude that, if h_1 and h_2 are close, then also the premium prices are close. However, h is always identifiable by the following Proposition.

Proposition 1.6. *If $\pi_{h_1}(F) = \pi_{h_2}(F)$ for all distribution functions F (the value ∞ is not excluded), then*

$$h_1(v) = h_2(v) \ \text{a.s.}$$

Proof. Let F_a be the distribution which takes the value 0 with probability a and the value 1 with probability $1-a$, for some $a \in (0,1)$, then its inverse F_a^{-1} is the indicator function of the interval $[a,1]$. Hence,

$$\pi_{h_1}(F_a) = \int \mathbb{1}_{[a,1]}(v)\, h_1(v)\, dv = \int_a^1 h_1(v)\, dv = \pi_{h_2}(F_a) = \int_a^1 h_2(v)\, dv.$$

Thus, the distortion distributions H_1 and H_2 are equal and therefore $h_1 = h_2$ almost surely. □

Remark 1.6. *Note the previous proposition is true if the family of distributions where the premium prices coincide contains all the Bernoulli variables. Compare also Theorem 2 in [102].*

Remark 1.7. *Another family with the property that the premium prices for this family determine the distortion in a unique manner is the family of Power distributions of the form $F_\gamma(u) = u^\gamma$ on $[0,1]$ and more general of the form $F_{\gamma,\beta}(u) = \beta^{-\gamma}u^\gamma$ on $[0,\beta]$. The distortion premium prices for this family are*

$$\int_0^1 \beta\, v^{1/\gamma}\, h(v)\, dv,$$

and the uniqueness of h and β is obtained since

$$\beta = \lim_{\gamma\to\infty} \int_0^1 \beta v^{1/\gamma}\, h(v)\, dv,$$

and the inversion formula for the Mellin transform (see Zwillinger 2002 [114]).

1.5 Estimating the distortion density from observations

The way how insurance companies calculate a premium is typically not revealed to the customer. Notice that risk premia appear not only in the insurance business, see the link of insurance premium prices and asset pricing in Nguyen et al. 2012 [69]. Risk premia appears in other areas such as

- **Power future markets** A future contract fixes the price today for delivery of energy later. There is the risk of price changes between now and the delivery period. Thus, such a contract has the character of an insurance and the pricing principles apply, although the price is found in exchange markets (e.g. electricity future markets).
- **Exotic options** While standard options are priced through a replication strategy argument, this argument does not apply for other types of options and these options have the character of insurance contracts. Pricing of such contracts is often done over the counter, but again the pricing principle is not revealed to the counterparty.

- **Credit derivatives** Also these contracts carry the character of insurance and can be priced according to insurance price principles.

In this section we assume that we know the distortion premium prices of m contracts, which are all priced with the same distortion density h. For each contract j, we also have a sample $x_1^{(j)}, \ldots, x_n^{(j)}$ of size n drawn from the loss distribution of this contract at our disposal. For simplicity we assume that n is the same for all contracts, but this is not crucial.

The goal of this section is to show how the distortion density h can be regained from the observations of the insurance prices, which would help us to shed more light on the price formation of contract counterparties. Notice that our aim is not to estimate the distortion premium prices from empirical data as is done in Gourieroux and Liu 2006 [48] or Tsukahara 2013 [88].

A simulation example As an example we consider m *different* loss distributions, all of Gamma type. From each distribution, we obtain a sample of size n. For each sample, we calculate the AV@R and power distortion premium prices. Based on the prices obtained and our samples, we aim to recover the distortion density h. We denote the ordered sample from the j-th loss distribution by $x_{[1]}^{(j)}, \ldots, x_{[n]}^{(j)}$. The distortion premium, with distortion density h for each sample $j = 1, \ldots, m$, is

$$\pi^{(j)} = \sum_{i=1}^{n} x_{[i]}^{(j)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} h(v)\, dv = \sum_{i=1}^{n} x_{[i]}^{(j)} \left(H\left(\frac{i}{n}\right) - H\left(\frac{i-1}{n}\right)\right). \tag{1.16}$$

On the following, we develop (1.16) for the particular cases of AV@R and power distortion premium prices for each sample $j = 1, \ldots, m$.

AV@R distortion premium The price for h_α defined on (1.7) is

$$\pi^{(j)} = \frac{1}{n\,(1-\alpha)} \cdot \sum_{i=i_a}^{n} x_{[i]}^{(j)}, \tag{1.17}$$

where $1 < i_\alpha < n$ s.t. $\frac{i_\alpha - 1}{n} \leq \alpha < \frac{i_\alpha}{n}$.

Power distortion premium The price given by the power distortion $h^{(s)}$ defined in (1.3) with $0 < s < 1$ is

$$\pi^{(j)} = \sum_{i=1}^{n} x_{[i]}^{(j)} \cdot \left(\left(1 - \frac{i-1}{n}\right)^s - \left(1 - \frac{i}{n}\right)^s \right), \tag{1.18}$$

and the price given by $h^{(s)}$ defined in (1.5) with $s \geq 1$ is

$$\pi^{(j)} = \sum_{i=1}^{n} x_{[i]}^{(j)} \cdot \left(\left(\frac{i}{n} \right)^s - \left(\frac{i-1}{n} \right)^s \right). \tag{1.19}$$

The inverse problem consists on estimating the distortion density h from observed prices. Recall that among the examples we presented of common distortion densities we had step functions and continuous functions, therefore we will use step and spline functions in order to estimate estimate h. We do so for the prices obtained in (1.17), (1.18) and (1.19).

1.5.1 Estimation of the distortion density with a step function

Distortion density as a step function Let $\widehat{h}_l^1$ denote the step function consisting of l equal-size steps, defined as

$$\widehat{h}_l^1(v) = \sum_{k=1}^{l} \lambda_k \cdot I_{\left[L \cdot \frac{k-1}{n}, L \cdot \frac{k}{n}\right)}(v) = \sum_{k=1}^{l} \lambda_k \cdot I_{\left[\frac{k-1}{l}, \frac{k}{l}\right)}(v), \tag{1.20}$$

where $L = n/l$, $\lambda_s \in \mathbb{R}$ for $k = 1, \ldots, l$ and l denotes the dimension of the step function space. We also impose

$$\int_0^1 \widehat{h}_l^1(v)\, dv = \sum_{k=1}^{l} \int_{\frac{k-1}{l}}^{\frac{k}{l}} \lambda_k \, dv = \frac{1}{l} \cdot \sum_{k=1}^{l} \lambda_k = 1, \tag{1.21}$$

with $0 \leq \lambda_1 \leq \cdots \leq \lambda_l$. In this way, $\widehat{h}_l^1$ fulfils the density constraints as well as the non-decreasing constraints.

Prices with the step function For each sample $j = 1, \ldots, m$, the prices with $\widehat{h}_l^1$ are

$$\widehat{\pi}^{(j)} = \sum_{i=1}^{n} x_{[i]}^{(j)} \cdot \int_{\frac{i-1}{n}}^{\frac{i}{n}} \widehat{h}_l^1(v)\, dv = \sum_{k=1}^{l} \sum_{i=(k-1)L+1}^{L \cdot k} x_{[i]}^{(j)} \cdot \int_{\frac{i-1}{n}}^{\frac{i}{n}} \lambda_k \, dv = \sum_{k=1}^{l} \frac{\lambda_k}{n} \cdot \sum_{i=(k-1)L+1}^{L \cdot k} x_{[i]}^{(j)}, \tag{1.22}$$

Estimation In order to estimate $\widehat{h}_l^1$ we will minimize the squares of the differences between the prices obtained by a distortion function h and the premium obtained by $\widehat{h}_l^1$ in (1.22). We will test our results with the given prices $\pi^{(j)}$ calculated in (1.17), (1.18) and (1.19). We solve,

$$\begin{aligned}
\min_{\lambda} \quad & \sum_{j=1}^{m} (\hat{\pi}^{(j)} - \pi^{(j)})^2 \\
\text{s.t.} \quad & \frac{1}{l} \cdot \sum_{i=1}^{l} \lambda_i = 1 \\
& 0 \le \lambda_1 \le \cdots \le \lambda_l \,.
\end{aligned} \tag{P_1}$$

1.5.2 Estimation of the distortion density with a cubic monotone spline

B-splines construction For our purposes we will define the splines on the interval $[0,1]$. Any B-spline is a linear combinations of the B-spline basis functions. The B-spline basis functions have all the same degree, b and we choose to define them at equally spaced knots $t_k = k/L$, for $k = 0, \ldots, L$, hence L subintervals. The functions for this basis are denoted as $B_{k,b}$ and constructed following a recursion formula. The B-spline basis function of degree 0 is denoted and defined as

$$B_{k,0}(v) = \begin{cases} 1 & t_k \le v \le t_{k+1} \\ 0 & \text{otherwise.} \end{cases}$$

The B-spline basis functions of degree b, $B_{k,b}$ are obtained as an interpolation between $B_{k,b-1}$ and $B_{k+1,b-1}$, following the recursion formula

$$B_{k,b}(v) = \frac{v - t_k}{t_{k+b} - t_k} B_{k,b-1}(v) + \frac{t_{k+b+1} - v}{t_{k+b+1} - t_{k+1}} B_{k+1,b-1}(v).$$

In the recursion we need to define fake knots $t_{-k} = 0$ and $t_{L+k} = 1$ for $k = 1, \ldots, b$. In our case, we consider splines of degree $b = 2$. If we divide $[0,1]$ in L equally sized intervals, the basis has $L + 2$ functions

$$\{B_{-2,2}, B_{-1,2}, B_{0,2}, B_{1,2}, \ldots, B_{L-1,2}\}. \tag{1.23}$$

Notice that all the elements of the basis can be obtained by translating the B-spline basis function $B_{0,2}$ defined on the first $b + 2 = 4$ knots. In order to have a base of increasing monotone cubic splines we integrate the functions of (1.23) and obtain a new base

$$\{S_{-2}, S_{-1}, S_0, \ldots, S_{L-1}\}, \tag{1.24}$$

where $S_k(v) = \int_0^v B_{k,2}(w)\, dw$ for all $k = -2, \ldots, L-1$. We scale the functions of (1.23) so the splines in (1.24) are distribution functions. Note that no linear combination of (1.24) gives us a constant function, due to construction of (1.24). Therefore, we need one element to our base, say $S_L(v) = c$ and hence

$$\{S_{-2}, S_{-1}, S_0, \ldots, S_{L-1}, S_L\}, \tag{1.25}$$

is our final base with $l = L + 3$ elements, where l denotes its dimension.

As an example we illustrate the base obtained for $L = 5$. Starting with $B_{0,2}$ defined on $t_0 = 0, t_1 = 1/5, t_2 = 2/5, t_3 = 3/5$, precisely

$$B_{0,2}(v) = \frac{5^3}{2} \cdot \left(v^2 \mathbb{1}_{[t_0,t_1)} + (v(t_2 - v) + (t_3 - v)(v - t_1)) \mathbb{1}_{[t_1,t_2)} + (t_3 - v)^2 \mathbb{1}_{[t_2,t_3)} \right)$$

We denote by S_0 the distribution of $B_{0,2}$ and obtain the rest of the monotone cubic splines by translating S_0. The basis of cubic monotone splines of dimension $l = 8$, illustrated in Figure 1.1, is denoted as

$$\{S_{-2}, S_{-1}, S_0, \ldots, S_4, S_5\}, \tag{1.26}$$

where $S_k(v) = S_0(v - k/5)$ for $k = -2, \ldots, 4$ and $S_5(v) = c$.

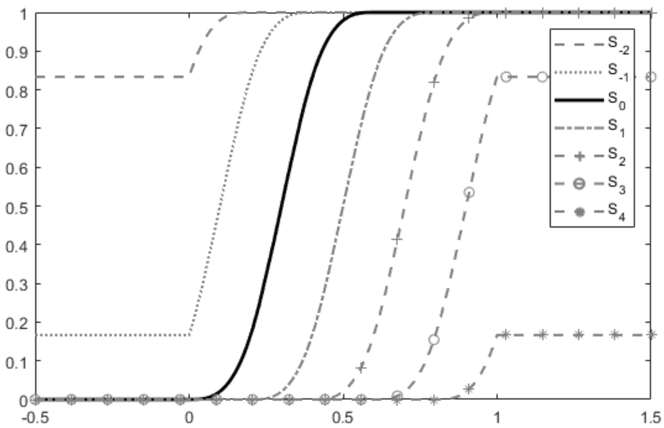


Figure 1.1: Cubic increasing monotonic base functions.

Any linear combination with positive scalars of the splines in (1.26) define a spline which is an increasing and positive function.

Distortion density as a spline Let $\widehat{h}_l^2(v)$ denote an increasing monotone cubic density defined as a linear combination of $l = L + 3$ splines in (1.25)

$$\widehat{h}_l^2(v) = \sum_{k=-2}^{L} \lambda_k \cdot S_k(v), \tag{1.27}$$

where $\lambda_k \geq 0$ for all $k = -2, \ldots, L$. Notice that by setting the scalars to be non-negative, $\widehat{h}_l^2$ is increasing. However, $\widehat{h}_l^2$ must integrate to 1 on $[0, 1]$, hence

$$\int_0^1 \widehat{h}_l^2(v)\, dv = \sum_{k=-2}^{L} \lambda_k \cdot \int_0^1 S_k(v)\, dv = \sum_{k=-2}^{L} \lambda_k \cdot \left(\sum_{i=1}^{n} A_{ik} \right) = \sum_{k=-2}^{L} \lambda_k \cdot a_k = 1,$$

where

$$A_{ik} = \int_{\frac{i-1}{n}}^{\frac{i}{n}} S_k(v)\, dv, \quad a_k = \sum_{i=1}^{n} A_{ik}. \tag{1.28}$$

Prices with the spline For each sample $j = 1, \ldots, m$, the prices with $\widehat{h}_l^2$ are

$$\widehat{\pi}^{(j)} = \sum_{i=1}^{n} x_{[i]}^{(j)} \cdot \int_{\frac{i-1}{n}}^{\frac{i}{n}} \widehat{h}_l^2(v)\, dv = \sum_{i=1}^{n} x_{[i]}^{(j)} \cdot \left(\sum_{k=-2}^{L} \lambda_k\, A_{ik} \right). \tag{1.29}$$

Estimation. Given prices $\pi^{(j)}$ calculated as in (1.17), (1.18) or (1.19) and the prices calculated in (1.29) for every sample $j = 1, \ldots, m$, we solve

$$\begin{aligned} \min_{\lambda} \quad & \sum_{j=1}^{m} (\hat{\pi}^{(j)} - \pi^{(j)})^2 \\ \text{s.t.} \quad & \sum_{k=-2}^{L} \lambda_k \cdot a_k = 1 \\ & \lambda_k \geq 0,\ k = -2, \ldots, L, \end{aligned} \tag{P_2}$$

where a_k is defined in (1.28).
The estimations obtained by solving (P_1) and (P_2) are presented below.

AV@R distortion premium We consider particular cases of h_α for $\alpha = 0.9, 0.95$. We estimate the distortion density for each of the cases, with two different step functions, corresponding to $l = 8, 10$ steps, and two different spline basis functions of dimensions $l = 8, 13$, respectively.
Step function. The estimated step distortions $\widehat{h}_l$ for $l = 8, 10$ are obtained by solving (P_1) and illustrated below (Fig. 1.2).
Splines. The estimated spline distortions $\widehat{h}_l^2$ for $l = 8, 13$ are obtained by solving (P_2) and illustrated below (Fig. 1.3).

Power distortion premium For this case we consider$h^{(s)}$ for $s = 0.8, 3$. We solve (P_1) and (P_2) with the same number of steps and number of spline basis functions as before.
Step function. The estimated step distortions $\widehat{h}_l^1$ for $l = 8, 10$ are obtained by solving (P_1) and illustrated below (Fig. 1.4).
Splines. The estimated spline distortions $\widehat{h}_l^1$ for $l = 8, 13$ are obtained by solving (P_2) and illustrated below (Fig. 1.5).

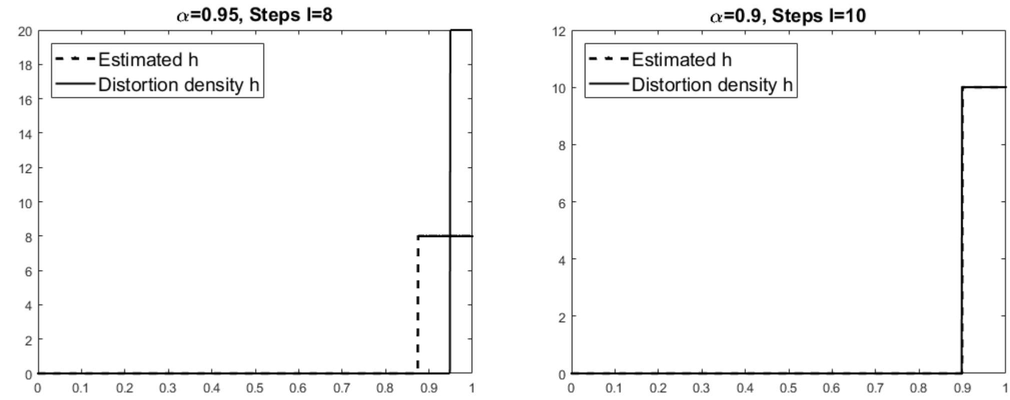


Figure 1.2: The true distortion density h_α for $\alpha = 0.9, 0.95$ and their respective step functions estimators for $l = 8$ steps, and $l = 10$ steps.

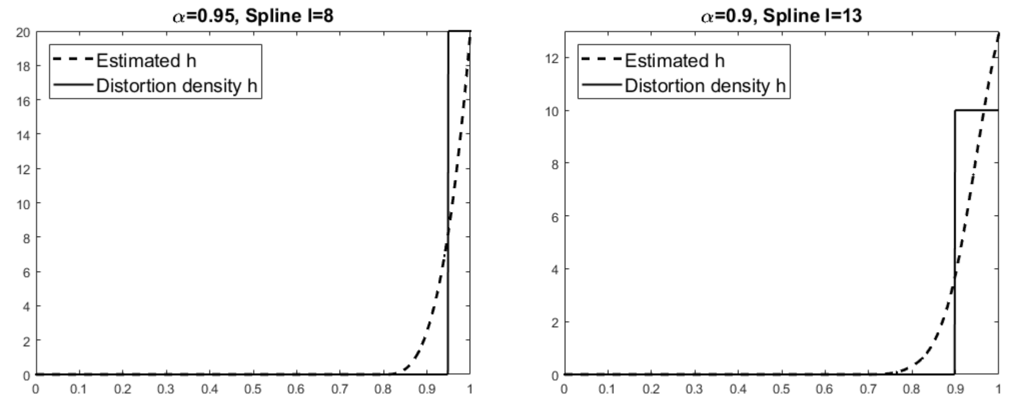


Figure 1.3: The true distortion density h_α for $\alpha = 0.9, 0.95$ and their respective spline estimators for $l = 8$ and $l = 13$ spline base dimension.

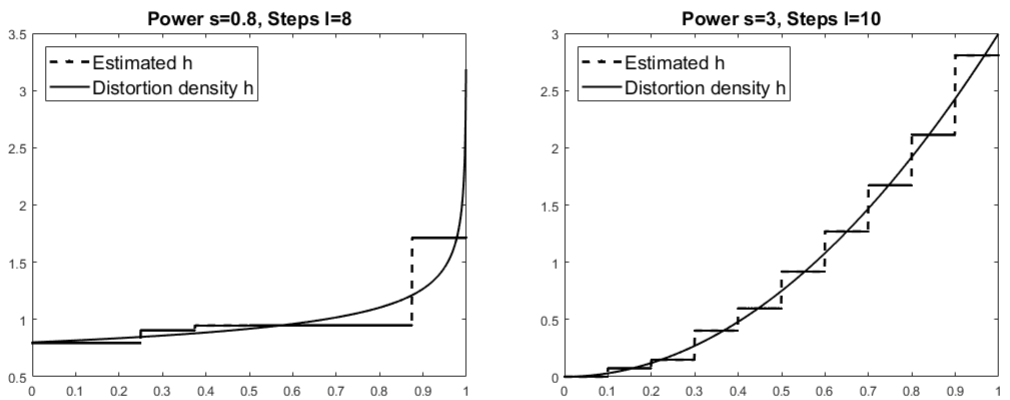


Figure 1.4: The true distortion density $h^{(s)}$ for $s = 0.8, 3$ and their respective estimated step distortions with $l = 8, 10$ steps.

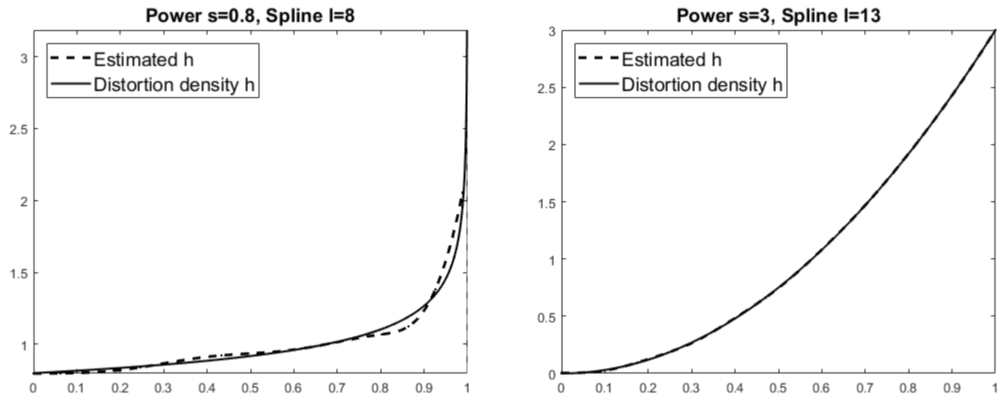


Figure 1.5: The true distortion density $h^{(s)}$ for $s = 0.8, 3$ and their respective estimated spline distortions with $l = 8, 13$ spline base dimension.

The optimal values of the optimization problems for all the cases can be seen in the following Table 1.1.

AV@R	$\alpha = 0.9$	$\alpha = 0.95$	Power	$s = 0.8$	$s = 3$
Step $l = 8$	7.3248	107.1562	Step $l = 8$	0.0012	1.1466e-04
Step $l = 10$	0	58.4835	Step $l = 10$	0	5.1772e-05
Spline $l = 8$	0.0322	13.0785	Spline $l = 8$	3.6976e-04	0
Spline $l = 13$	0.0154	0.0502	Spline $l = 13$	1.3251e-04	0

Table 1.1: Optimal values of the problems (P_1) and (P_2) for the AV@R$-$distortion and the power distortion.

1.6 Ambiguity

In this section we combine the distortion premium with the ambiguity principle. Such an approach allows us to incorporate model uncertainty into the premium. Recall that, by setting the distortion density to $h = 1$, we would price just with the ambiguity principle. As was mentioned in Section 1.1, distances can be used to define ambiguity sets. Here, closed Wasserstein balls will serve as ambiguity sets. These sets will be centred at F, an initial distribution, that we refer to as our baseline model.

Definition 1.2 (Robust distortion premium under Wasserstein balls with d_1)**.** *Let F be the baseline loss distribution, h a distortion density. The robust distorted price of order $r \geq 1$ is*

$$\pi^{\epsilon}_{h,r,d_1}(F) = \sup\{\pi_h(G) : G \in \mathcal{B}_{r,d_1}(F,\epsilon)\}, \tag{P-r}$$

where $\mathcal{B}_{r,d_1}(F,\epsilon) = \{G : WD_{r,d_1}(G,F) \leq \epsilon\}$. We call the worst case distribution and denote it by F^* if $F^* \in \mathcal{B}_{r,d_1}(F,\epsilon)$ and is such that

$$\pi^{\epsilon}_{h,r,d_1}(F) = \pi_h(F^*).$$

Remark 1.8. *Notice that for $r_1 \le r_2$*

$$WD_{r_1,d_1} \le WD_{r_2,d_1}, \tag{1.30}$$

thus $\mathcal{B}_{r_1,d_1} \supseteq \mathcal{B}_{r_2,d_1}$.

We can say more about the value and solution of (P-r) if we choose $r = p$. We start with bounded distortion densities, i.e. for $p = 1$ and $q = \infty$.

Proposition 1.7 (Characterization of the worst case distribution for $\mathbf{r} \ge \mathbf{p} = \mathbf{1}$). *Let the baseline distribution F have its first moment finite.*

(i) If h is unbounded, then (P-r) for $r = 1$ is unbounded.

(ii) If h is bounded with $\sup_v h(v) = \|h\|_\infty$, then (P-r) is bounded for all $r \ge 1$. If $r = 1$, the optimal value of (P-r) is

$$\pi^{\epsilon}_{h,1,d_1}(F) = \pi_h(F) + \epsilon \cdot \|h\|_\infty.$$

We interpret the additional term $\epsilon \cdot \|h\|_\infty$ as the ambiguity premium. *For the worst case distribution,*

- *if $h(v) = \|h\|_\infty$ for $v \ge 1 - \eta$ and $0 < \eta \le 1$, then the supremum is attained at*

$$F^*_\eta(x) = \begin{cases} F(x) & x < F^{-1}(1-\eta), \\ 1-\eta & F^{-1}(1-\eta) \le x < F^{-1}(1-\eta) + \epsilon/\eta, \\ F\left(x - \epsilon/\eta\right) & x \ge F^{-1}(1-\eta) + \epsilon/\eta. \end{cases}$$

- *Otherwise, the supremum is not attained, but can be approximated by the sequence $F^*_{1/n}(x)$, $\forall n \in \mathbb{N}$.*

Proof. (i) Given that h is increasing and unbounded, the increasing sequence $K_n = h\left(1 - 1/n\right)$, is such that $\lim_{n\to\infty} K_n = \infty$. For all $n \in \mathbb{N}$ we define a distribution G_n such that

$$G_n^{-1}(v) = F^{-1}(v) + \epsilon \cdot n\, \mathbb{1}_{[1-1/n,1]}.$$

G_n is on the boundary of $\mathcal{B}_{1,d_1}(F,\epsilon)$ and

$$\pi_h(G_n) = \pi_h(F) + \epsilon \cdot n \int_{1-1/n}^{1} h(v)\, dv \ge \pi_h(F) + \epsilon\, K_n.$$

Hence, (P-r) is unbounded for $r = 1$. (ii) It is sufficient to prove (P-r) is bounded for $r = 1$ since $\mathcal{B}_{1,d_1} \supseteq \mathcal{B}_{r,d_1}$ for all $r \ge 1$ (see Remark 1.8). Any admissible G for $r = 1$ can be written as $G^{-1}(v) = F^{-1}(v) + G_1^{-1}(v)$, where G_1 is such that $\int_0^1 G_1^{-1}(v)\, dv \le \epsilon$. Since F has its first moment finite, the following upper bound is finite:

$$\pi_h(G) = \pi_h(F) + \int_0^1 G_1^{-1}(v)\, h(v)\, dv \le \pi_h(F) + \epsilon \cdot \|h\|_\infty. \tag{1.31}$$

The distribution $F^*_\eta(x)$ given in the Proposition has inverse

$$(F^*_\eta)^{-1}(v) = F^{-1}(v) + \frac{\epsilon}{\eta}\mathbb{1}_{[1-\eta,1]}.$$

Therefore, F^*_η is on the boundary of $\mathcal{B}_{1,d_1}(F,\epsilon)$ and

$$\pi_h(F^*_\eta) = \int_0^1 \left(F^{-1}(v) + \frac{\epsilon}{\eta}\mathbb{1}_{[1-\eta,1]}\right) h(v)\,dv = \pi_h(F) + \frac{\epsilon}{\eta}\int_{1-\eta}^1 h(v)\,dv.$$

If $h(v) = \|h\|_\infty$ for $v \geq 1-\eta$, then F^*_η attains the upper bound in (1.31). Otherwise, $F^*_{1/n}$ approaches the maximum from below, since

$$(F^*_{1/n})^{-1}(v) = F^{-1}(v) + \epsilon \cdot n\mathbb{1}_{[1-1/n,1]},$$

and

$$\pi_h(F^*_{1/n}) = \pi_h(F) + \epsilon \cdot n \int_{1-1/n}^1 h(v)\,dv \uparrow \pi_h(F) + \epsilon \cdot \|h\|_\infty.$$

□

Remark 1.9. *The solution F^*_η in Proposition 1.7 is not unique. Any distribution $\tilde{F}_\eta$ such that $\tilde{F}_\eta^{-1}(v) = F^{-1}(v) + \frac{\epsilon}{\eta}\cdot k(v)\mathbb{1}_{[1-\eta,1]}$, with $\frac{1}{\eta}\cdot k(v)\mathbb{1}_{[1-\eta,1]}$ a density on $[0,1]$, attains the supremum.*

As an example, we illustrate the worst case distribution for the AV@R premium.

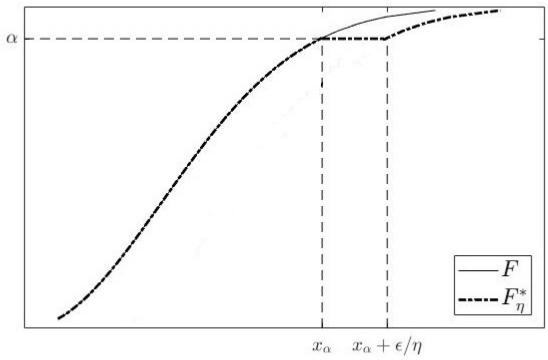


Figure 1.6: The worst case distribution F^*_η for h_α with $\alpha = 0.9$ is obtained by shifting F from x_α, a length ϵ/η, where $x_\alpha = F^{-1}(\alpha)$ and $\eta = 1-\alpha$.

If h is unbounded we can characterize the solution of (P-r) as follows.

Proposition 1.8 (Characterization of the worst case distribution for $\mathbf{r \geq p > 1}$**).** *Let the baseline distribution F have finite $p-$moments. If $h \in \mathcal{L}^q$, then (P-r) is bounded for $r \geq p$. If $r = p$, the optimal value of (P-r) is*

$$\pi^\epsilon_{h,p,d_1}(F) = \pi_h(F) + \epsilon \cdot \|h\|_q^q.$$

Also in this case, the term $\epsilon \cdot \|h\|_q^q$ is interpreted as ambiguity premium.

Furthermore, the worst case distribution F^* of (P-r) for $r = p$ is such that

$$F^{*-1}(v) = F^{-1}(v) + \epsilon \cdot \left(\frac{h(v)}{||h||_q}\right)^{q/p}.$$

Proof. We prove (P-r) is bounded for $r = p$ and by Remark 1.8 we have boundness for all $r \geq p$. Notice that, for all admissible G, if $r = p$, we have

$$\begin{aligned}
\int_0^1 G^{-1}(v)\, h(v)\, dv &\leq \int_0^1 F^{-1}(v)\, h(v)\, dv + \int_0^1 |G^{-1} - F^{-1}|\, h(v)\, dv \\
&\leq \pi_h(F) + \left(\int_0^1 |G^{-1} - F^{-1}|^p\, dv\right)^{1/p} ||h||_q \\
&\leq \pi_h(F) + \epsilon \cdot ||h||_q.
\end{aligned}$$

F^* is admissible since it is on the boundary of $\mathcal{B}_{p,d_1}(F, \epsilon)$

$$WD_{p,d_1}(F, F^*) = \left(\int_0^1 \epsilon^p \cdot \left(\frac{h(v)}{||h||_q}\right)^q dv\right)^{1/p} = \epsilon,$$

and F^* attains the upper bound

$$\pi_h(F^*) - \pi_h(F) = \int_0^1 \epsilon \cdot \left(\frac{h(v)}{||h||_q}\right)^{q/p} h(v)\, dv = \epsilon \cdot \int_0^1 \frac{h(v)^q}{||h||_q^{q-1}}\, dv = \epsilon \cdot ||h||_q.$$

□

Under some conditions on h we can also prove unboundness of (P-r) for $r > p > 1$ in the case where h is not in L^q, where q is the conjugate of p, the finite moments of F.

Proposition 1.9 (Unboundness for $\mathbf{r > p > 1}$)**.** *Let the baseline distribution F have finite $p-$moments and let $h \notin \mathcal{L}^q$, for p, q conjugates and r, s conjugates with $r > 1$. If there exists $s_1 < s$ such that $\int_0^1 h(v)^{s_1}\, dv = \infty$ and $h \in \mathcal{L}^t$, for all $t < s_1$, then (P-r) is unbounded for all $r > p$.*

Proof. Define $\psi_\eta(v) = h(v)^{s_1 - 1}\mathbb{1}_{[1-\eta,1]}$. Since $\psi_\eta \in \mathcal{L}^r$ for $r > 1$ (note that $r(s_1 - 1) < s_1$), there exists an $0 < \eta < 1$ such that

$$\int_0^1 \psi_\eta(v)^r\, dv = \int_{1-\eta}^1 h(v)^{r(s_1-1)}\, dv < \epsilon.$$

Thus, the distribution G_η such that $G_\eta^{-1}(v) = F^{-1} + \psi_\eta(v)$ is in $\mathcal{B}_{r,d_1}(F, \epsilon)$. And its premium is unbounded

$$\pi_h(G_\eta) = \pi_h(F) + \int_0^1 \psi_\eta(v)\, h(v)\, dv = \pi_h(F) + \int_{1-\eta}^1 h(v)^{s_1} > \infty.$$

□

Remark 1.10. *If instead of the metric d_1 we consider $d_p(x,y) = |x^p - y^p|$ as underlying metric for the Wasserstein distance, we could define the ambiguity principle*

$$\pi^{\epsilon}_{h,1,d_p}(F) = \sup\{\pi_h(G) : G \in \mathcal{B}_{1,d_p}(F,\epsilon)\}, \tag{P-dp}$$

where $\mathcal{B}_{r,d_p}(F,\epsilon) = \{G : WD_{r,d_p}(G,F) \le \epsilon\}$. It is easy to see that, if F has $p-$moments the constraint of the balls make all of admissible distributions to have also $p-$moments, therefore for Proposition 1.3, if $h \in \mathcal{L}^q$, then (P-dp) is bounded. Furthermore, continuity respect to this Wasserstein distance implies our continuity results in Section 1.3.

1.7 Conclusions

After some introduction about general premium principles we propose generalizations of the distortion premium. In addition, we have studied in detail three functional relationships for the distortion premium

- the premium function $F \mapsto \pi_h(F)$, i.e. the properties of π_h as a premium principle,
- the direct function $h \mapsto \pi_h(F)$, i.e. the dependency on the distortion density,
- the inverse functions $\pi_h(F) \mapsto h$.

The smoothness properties are important for robustness aspects, however it is well known that a quite smooth direct function makes the inverse problem difficult. We showed however that the inverse problem is identifiable and we gave a simple quadratic optimization problem to estimate it from empirical data. We successfully illustrated this in a simulation study, the application on real data is left for further research. We also identified the ambiguity premium for Wasserstein balls as ambiguity sets offering, in some cases, a specific formulation of the worst case distribution. It turned out that the extra premium for ambiguity depends on the distortion function h and in a multiplicative way on the ambiguity radius ϵ, but does not on the loss distribution F itself. Thus it is the same for all contracts and can be calculated in a separate manner. Finally, by using different distances as underlying metrics for the Wasserstein ball, and hence, for the ambiguity set, we could find bounds for the robust premium is always bounded.

Acknowledgements Open access funding provided by University of Vienna.

CHAPTER 2

Recovering distortion functions in power markets under model ambiguity

Joint work with Florentina Paraschiv and Michael Schürle

Abstract

This paper addresses the relationship between spot and futures prices in power markets. It is well-known that the classical arbitrage-based methods are not applicable for pricing electricity derivatives in the same way as for other commodities due to the non-storability of electricity. Consequently, futures valuation do not follow a general rule. As an extension of the existing literature, we propose the valuation of futures in terms of distortion functions, which quantify the risk preferences of consumers and producers. Since our goal is the identification of a pricing rule for futures from spot prices, a model for the evolution of spot prices is needed that is estimated (only) with information known from the day-ahead market. For this purpose, we propose a regime-switching model that reflects characteristic features of prices on the spot market for electricity. While the distortion principle is widely used for pricing insurance contracts, we make it applicable for pricing electricity futures, by including negative risk premia. It has also been established that futures prices contain more information than the one provided by the spot market, therefore we include a correction parameter for its quantification. Additionally, we incorporate a premium for ambiguity in terms of the Wasserstein distance. The novelty of this paper lies in the identification of these three components: distortion functions, correction parameters and ambiguity premia. Overall, this explains not only the prices of futures as well as the magnitude and sign of risk premia in power markets, but the distinction between these components allows a better understanding of the pricing mechanism.
Keywords— Ambiguity · Distortion premium · Electricity futures · Energy markets · Regime-switching models · Risk premia · Wasserstein distance.

2.1 Introduction and motivation

Power markets are constantly changing over the years and have particular characteristics that distinguish them from classical financial markets and other commodity markets. Electricity prices are not only affected by weather conditions but depend also on prices of different fuels used for power generation, e.g., coal, gas or oil, as well as the production costs of the various types of plants (thermal, hydro, photovoltaic, wind). In the medium and long term, political decisions also have an impact on electricity prices since policy makers promote the transition towards low-carbon power generation. Most of the physical electricity is traded on a day-ahead market in an auction mechanism, where individual prices are calculated for each hour of the following day by matching hourly demand and supply. Although participants in the market trade large volumes of electricity in the day-ahead market, they also can trade in the future market.

Futures can be traded ahead in time to deliver a certain amount of electricity over a longer period, e.g., a week, month, quarter or year. Thus, futures contracts effectively allow locking in the price for each hour during the delivery period. These contracts are available in two different profiles: base and peak. The former refers to the arithmetic mean of spot prices over the delivery period, while the latter refers to only peak hours of the delivery period (from 8:00 a.m. to 8:00 p.m. excluding weekends). Understanding the price mechanism of these contracts will be the main topic of this paper. For our empirical tests, we work with the German/Austrian market, where spot prices are obtained from the European Power Exchange (EPEX SPOT) and prices of Phelix futures with monthly delivery periods from the European Energy Exchange (EEX). Both spot and futures are quoted in EUR/MWh.

Let X^m denote the average spot prices over a month $m = [T_1, T_2]$. This average can be calculated for both profiles, base and peak as

$$X^{m,B} = \frac{1}{|hour(m)|} \sum_{t \in hour(m)} spot_t\,, \qquad X^{m,P} = \frac{1}{|peak(m)|} \sum_{t \in peak(m)} spot_t\,, \tag{2.1}$$

where $hour(m)$ discretizes m in hours and $peak(m)$ selects the peak hours of $hours(m)$. Monthly base (peak) futures refer to the base (peak) average spot prices. For simplicity, we denote the monthly averages by X^m, and denote by $\mathbb{F}(\tau, m)$ the futures price traded at time $\tau < T_1$ with maturity m and underlying X^m (unknown at time τ).

Futures that refer to storable commodities are priced following the theory of storage, inventory and convenience yield (Fama and French 1987 [32]). Prices of these contracts are given by the actual market price of the commodity including interest rates, value of holding the commodity and time-to-delivery (see also Gibson and Schwartz 1990 [43], Eyedeland and Wolyniec 2003 [31] and Geman 2009 [41]). Holding the commodity is key for this pricing rule. Since electricity cannot be stored, the convenience yield and cost of carry are hardly identifiable. A naive approach to explain the link between spot prices and futures is given by the expectations of the average spot prices during the delivery period (see Stoft 2002 [87]), e.g.,

$$\mathbb{F}(\tau, m) = \mathbb{E}_P(X^m | \mathcal{F}_\tau)\,, \tag{2.2}$$

where $\mathcal{F}_\tau$ is the $\sigma-$algebra of spot prices until time τ and P the physical measure of X^m. Another possibility, also proposed by Fama and French 1987 [32], is to add a risk premium to the previous expectation, obtaining

$$\mathbb{F}(\tau, m) = \mathbb{E}_P(X^m | \mathcal{F}_\tau) + \text{ risk premium}\,.$$

The risk premium is widely studied to explain the formation of futures prices in terms of traders' risk aversion. Producers want to protect themselves from lower spot prices, while consumers against large spot prices. Therefore, producers are willing to accept a discount and secure for not selling below that price, while buyers are willing to pay an extra load to secure not to pay above that price. Several authors study whether concrete markets have significant risk premia and its sign relating them to risk aversion of traders. If consumers are very risk-averse, their participation overcomes producers' and prices become larger. In the case where producers have more risk aversion, futures are traded at lower prices. A classic and well-known result is the equilibrium model proposed in Bessembinder and Lemmon 2002 [10], where the risk premium is explained in terms of moments (variance and skewness) of spot prices. This work is the benchmark for several authors, see Shawky 2003 [85], Longstaff and Wang 2004 [62], Diko et al. 2006 [19], Lucia and Torró 2011 [65], among others. Another example, where equilibrium arguments are used to explain futures pricing, is Benth et al. 2008 [9]. The authors use parametric utility functions to model the risk aversion of producers and consumers and explain the size and sign of risk premia. Different probability choices for the expectation were also proposed, for example, Benth and Sgarra 2012 [8] use an Esscher transform to explain the change of sign of the risk premia. In the spirit of the latter paper, Veraart and Veraart 2013 [94] propose modelling the spot price with Lévy semi-stationary processes and derive an analytical expression for the risk premia. A different and relevant aspect for pricing electricity futures concerns the $\sigma-$algebra $\mathcal{F}_\tau$. For non-storable commodities, $\mathcal{F}_\tau$ does not contain forward-looking information, such as planned future maintenance of power plants or reservoir levels of hydropower. Benth and Meyer-Brandis 2009 [7] emphasize the need to enlarge the filtration of spot prices to take into account future market information. In this case, they use the information approach to explore the relation between spot and futures prices instead of an equilibrium approach.

Taking all the previous facts into account, the goal of this paper is to explain the relationship between spot prices and futures and derive conclusions on the magnitude and sign of the risk premium. We propose to price futures following the *distortion premium principle* with some additional parameters that will be explained in the sequel. First of all, we consider the distortion premium, which is an expectation of X^m with respect to P_h given by

$$\mathbb{F}(\tau, m) = \mathbb{E}_{P_h}(X^m | \mathcal{F}_\tau)\,,$$

where P is the physical measure of X^m, h is a *distortion density* on [0,1] and P_h is the distorted probability. The choice of h will reflect different risk preferences and include a risk premium when h is different from the constant function 1. For the calculation of the premium, a distribution for the monthly averages X^m is required. To this end, we propose a regime-switching model for the spot prices that is inspired by the model described in Paraschiv et. al 2015 [72]. By modelling spot prices, we will be able to simulate hourly prices over a specific month, say m. Each simulated path will be averaged and give a sample point for X^m, the random variable describing monthly averages over m. Our model for spot prices reflects characteristic features of observed spot prices, like seasonal patterns (for different hours of a day, different weekdays and times of the year) and their spiking behaviour, i.e., prices may jump from one hour to the next to extremely high or low levels, remain there possibly for some hours, and then return to their previous level (Bunn and Karakatsani 2003 [15], Weron 2007 [104]). However, while the original model in [72] uses information from futures prices to obtain the spot price level, we model the price level as an additional latent factor that is identified in the estimation solely from spot market information by a Kalman filter.

In order to take into account additional information at trading time, which is not included in the filtration of spot prices $\mathcal{F}_t$, we introduce a correction term. This term is defined by a *shifting parameter* $\theta < 1$, which shifts the distortion premium towards smaller or larger prices. This way, we obtain the *shifted distortion premium*

$$\mathbb{F}(\tau, m) = (1-\theta)\, \mathbb{E}_{P_h}(X^m|\mathcal{F}_\tau)\,.$$

Since the majority of valuation methods proposed in the literature are model dependant, we incorporate in our estimations model ambiguity and account for model misspecification. Hence, our pricing approach is a robust version of the shifted distortion premium given by

$$\mathbb{F}(\tau, m) = (1-\theta)\, \mathbb{E}_{P_h}(X^m|\mathcal{F}_\tau) + \text{ambiguity premium}\,. \tag{2.3}$$

The parameter θ, the distortion density h and the ambiguity premium may be different depending on the trading time and delivery month, but we remove the dependence for legible purposes. In our case, we define the ambiguity premium in terms of the Wasserstein distance. The risk premium will also be explained in terms of the previous parameters. Notice that the risk premium can be calculated before and after delivery. Let $\tau < T_1$, the *ex-ante risk premium* is defined as

$$r_{ante}(\tau, m) = \mathbb{F}(\tau, m) - \mathbb{E}_P(X^m|\mathcal{F}_\tau)\,. \tag{2.4}$$

The *ex-post risk premium* at time τ is

$$r_{post}(\tau, m) = \mathbb{F}(\tau, m) - X^m\,. \tag{2.5}$$

The relationship between ex-ante risk premium and ex-post risk premium is given by

$$r_{post}(\tau, m) = r_{ante}(\tau, m) + \mathbb{E}(X^m|\mathcal{F}_\tau) - X^m\,, \tag{2.6}$$

where $\mathbb{E}(X^m|\mathcal{F}_\tau) - X^m$ is the forecast error and X^m the realized monthly average (known average). With our pricing rule (2.3), we can obtain closed-forms of the risk premia in (2.4) and (2.5) in terms of h, θ and the Wasserstein distance. In other words, futures prices and their risk premia will be explained by how the shifting parameter θ, the function h and the ambiguity, altogether distort the initial distribution of X^m to match the futures prices in the market. With the identification of these parameters, we study the presence of seasonality in futures as stated in Lucia and Schwartz 2002 [64]. We will also examine the dependence of the size and sign of the risk premia on the time-to-maturity. Some insurance principles were used in this context (see Benth et al. 2008 [9], Benth and Sgarra 2012 [8]) and we also address the model of Bessembinder and Lemmon can be seen as an insurance premium. Different than the previous studies, our approach recovers the risk preferences from observed prices with non-parametric distortions and performs estimations under model ambiguity, i.e., we solve an inverse problem. The estimations of the different parameters in (2.3) will be done in groups of contracts extending an idea presented in Escobar and Pflug 2018 [25].

To the best of our knowledge, this paper is the first that uses the distortion premium for the explanation of risk preferences in power markets. Besides, we include a shifting parameter θ to correct for information not available in the spot market and the incorporation of model ambiguity. With our estimations, we derive conclusions of the main features of futures prices and the risk premia in the EEX.

The details of insurance premiums and specifically the distortion premium with the previous considerations will be presented in Section 2.2. Section 2.3 specifies the spot price model and the data of futures studied here. Section 2.4 extends the identification procedure in Escobar and Pflug 2018 [25]. The results of the identification of different parameters are given in Section 2.5 and Section 2.6 concludes.

2.2 The shifted distortion premium and robustness

In this section, we introduce different insurance premia and relate them to present literature on pricing electricity futures. Since consumers are willing to pay an extra premium and producers to accept lower prices against even smaller spot prices, we argue that futures in energy markets can be seen as insurance contracts. We start with an overview of insurance principles and then we present the details of the distortion premium with a correction factor and robustness.

A trivial example of an insurance premium is the expected value. However, the valuation of futures with only an expectation does not fully cover the gap between spot prices and electricity futures. In the same manner, insurance premia include an extra load to the expected losses for pricing insurance contracts. Thus, more sophisticated pricing rules are required. Formally, for a non-negative loss random variable X with cumulative distribution function (cdf) F, an insurance premium is a functional, $\pi : \{X \geq 0 \text{ defined on } (\Omega, \mathcal{F}, P)\} \to \mathbb{R}_{\geq 0}$, whose output is at least the expected value of the loss. A positive extra load can be added to the expectation by simply calculating the premium as

$$\pi(X) = (1+\theta)\,\mathbb{E}(X)\,, \tag{2.7}$$

where $\theta > 0$. The extra load (obtained as the difference of any insurance premium and the expected loss) is defined as the risk premium. For insurance premia, the risk premia are always positive. More examples that include a positive risk premium can be obtained by adding moments to the expectation. For instance, the classic *variance premium principle*

$$\pi(X) = \mathbb{E}(X) + \alpha\, Var(X), \quad \alpha > 0\,. \tag{2.8}$$

Although the variance premium involves more information of the random variable in the pricing, it does not fulfill desirable properties for insurance premia. Examples of these properties are translation invariance ($\pi(X+a) = \pi(X) + a$, for any $a \in \mathbb{R}$) and positive homogeneity ($\pi(aX) = a\pi(X)$, for any $a > 0$). The latter property is closely related to non-arbitrage, which is extensively discussed in Venter 1991 [93] and Albrecht 1992 [3]. These properties are also common requirements for risk measures. Alternatives to variance premium principle for insurance pricing were proposed, e.g., Denneberg 1990 [18], or coherent risk measures Artzner 1999 [4] for risk measures. Despite this, one can extend the variance premium by incorporating more moments. We notice this is exactly the idea in the equilibrium model proposed by Bessembinder and Lemmon 2002 [10] for electricity futures in the Pennsylvania-new Jersey-Maryland (PJM) market. Although the implications of this model are verified in different electricity markets, the model itself may not be applicable in other markets for recent data, for example in the Nordic market after 2002 (see Lucia and Torró 2011 [65]).

A different way to add a risk premium in the pricing is given by transforming the losses. A classic example of this is the *certainty equivalence premium* (CEQ). This premium is defined as the solution of the equation

$$V(\pi) = \mathbb{E}(V(X))\,, \tag{2.9}$$

where V is a convex, strictly monotone disutility function (see Von Neumann and Morgenstern 2007 [98]). The choice of a disutility (or utility) function to describe participants' preferences, is widely used in economics. We also find them in the framework of power markets in Benth et al. 2008 [9]. We can also link the CEQ premium and the equilibrium model proposed by Bessembinder and Lemmon 2002 [10], see Appendix B.2.

Notice that the previous expectations are calculated respect to the physical measure P (or cdf F) of the random loss. Instead of transforming the losses X as in the CEQ premium, we may apply a transformation to its probability (or to its cdf F). This is exactly the idea of the distortion premium (Denneberg 1990 [18], Wang 1995 [99]). The distortion premium is defined by an expectation of the random variable with respect to a distortion of P. In practice, we distort the cdf of random losses, say F. For this reason, our notation will apply the premium to the cdf F and the same notation is used for the expectation. The definition we present here allows the distortion premium to price with a negative risk premium.

Definition 2.1. *Let X be a positive random variable with cdf F. Let $g: [0,1] \to \mathbb{R}$ be a non-decreasing function, have fixed points 0 and 1 and have second derivative. Denote its symmetric derivative by $h(v) = g'(1-v)$. The distorted distribution is given by*

$$F^h(x) = 1 - g(1 - F(x)) .$$

The distortion premium is the expectation respect to F^h

$$\pi_h(F) = \int_0^\infty 1 - F^h(x)\, dx . \tag{2.10}$$

The premium is above the mean for g concave and below the mean for g convex. The function g is called distortion function and h, which is a density on $[0,1]$ is called distortion density.

Remark 2.1. *If g is concave, $g(v) \geq v$, hence $F^h \leq F$, then the premium prices are above the mean*

$$\pi_h(F) = \int_0^\infty g(1 - F(x))\, dx \geq \int_0^\infty (1 - F(x))\, dx = \mathbb{E}(F) .$$

If g is convex, $\pi_h(F) \leq \mathbb{E}(F)$ follows.

Remark 2.2. *The distortion function g is concave (convex) if and only if h is increasing (decreasing). Therefore, an increasing distortion density h describes a risk-averse preference while a decreasing h describes a risk-seeking preference.*

Remark 2.3. *The distortion premium defined with g concave results in a coherent risk measure (see Artzner et al. 1999 [4]). Moreover, this premium is characterized in Yaari 1987 [109] and in Wang et al. 1997 [102]. For g convex, the distortion premium can be seen as an acceptability functional (see Pflug and Römisch 2007 [77]).*

An equivalent calculation of the premium in terms of h and the inverse of F is given by

$$\pi_h(F) = \int_0^1 F^{-1}(v)\, h(v)\, dv , \tag{2.11}$$

since the inverse of the distorted distribution F^h is given by $(F^h)^{-1}(v) = F^{-1}(v)\, h(v)$. Note that the function h is a weighting function that changes the original probabilities. In fact, it is the Radon-Nikodym derivative that represents the change from the original probability to the distorted probability.

The distortion premium will be applied to the empirical distributions of the pertaining monthly averages (given solely with information from the spot market). In order to capture forward-looking information not included in the day-ahead market, we incorporate a correction premium. Exogenous information can come from different sources, e.g., inventory, available capacities, production costs, temperature forecasts, or others. Another example of additional information at trading time is related to government policies, which target future dates. For this reason, we introduce a correction premium defined in terms of a shifting parameter. This parameter corrects our initial models for the monthly averages by adjusting their moments (mean and variance). We accomplish this with a shifting parameter θ, likewise in

(2.7). The distortion premium including a shifting parameter $\theta < 1$ will be called shifted distortion premium and it is be given by

$$\pi_{h;\theta}(F) = (1-\theta)\,\pi_h(F) = \int_0^1 (1-\theta)F^{-1}(v)\,h(v)\,dv\,, \tag{2.12}$$

where the correction premium is $\pm\theta\cdot\pi_h(F)$. Additionally, the previous formulation can also be written as the distortion premium applied to the cdf $F_\theta(x) = F(x/(1-\theta))$ with same distortion density h. We call F_θ the *shifted distribution* of F and we note its inverse is given by $F_\theta^{-1}(v) = (1-\theta)\,F^{-1}(v)$. Hence, the shifted distortion premium can be written as

$$\pi_{h;\theta}(F) = \pi_h(F_\theta)\,. \tag{2.13}$$

Thus, the shifted distortion premium gives a price by calculating an expectation after transforming F with θ and h. Below, we present examples of how the monotonicity of h influences in the distortion premium, as well as how the sign of θ changes the moments of F.

Average value-at-risk (AV@R). Also called conditional tail expectation (CTE), conditional value at risk (CVaR) and known in the insurance sector as expectred shortfall (ES) (Embrechts et al. 1997 [22]). The AV@R is a particular case of a distortion premium. To illustrate also the case for decreasing distortion densities we differentiate between lower value-at-risk (decreasing h) and upper value-at-risk (increasing h). The lower average value-at-risk for $\alpha > 0$ is

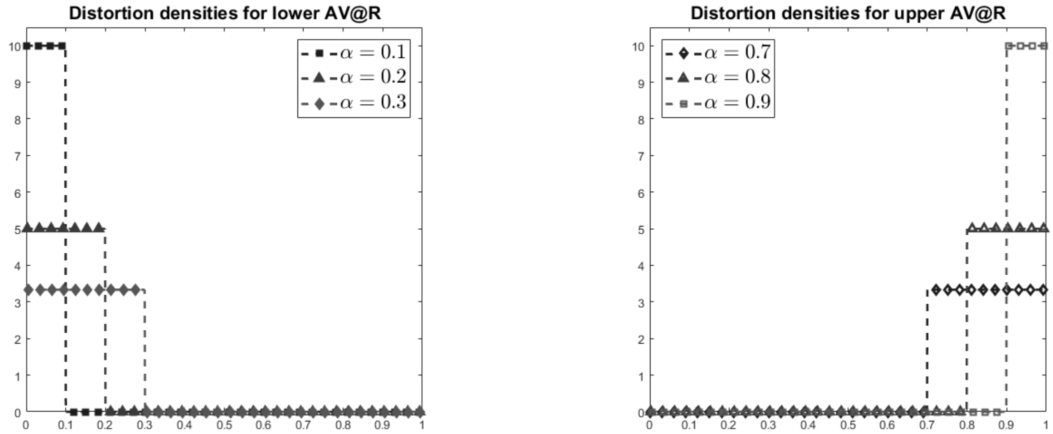
$$\pi_{h_{\underline{\alpha}}}(F) = \int_0^1 F^{-1}(v)\,h_{\underline{\alpha}}(v)\,dv \quad \text{with} \quad h_{\underline{\alpha}}(v) = \frac{1}{\alpha}\mathbb{1}_{[0,\alpha]}(v)\,,$$

and the upper average value-at-risk for $\alpha < 1$ is

$$\pi_{h_{\overline{\alpha}}}(F) = \int_0^1 F^{-1}(v)\,h_{\overline{\alpha}}(v)\,dv \quad \text{with} \quad h_{\overline{\alpha}}(v) = \frac{1}{1-\alpha}\mathbb{1}_{[\alpha,1]}(v)\,.$$

We illustrate some examples of both distortion densities for different quantiles (see Figure 2.1) and their respective prices (see Figure 2.2).

Shifted distribution. For the illustration of this example, assume F is the cdf of a continuous random variable and let $f(x) = F'(x)$ be its density function. Therefore the shifted density is given by $f_\theta(x) = f(x/(1-\theta))/(1-\theta)$, where $f(x) = F'(x)$ is the given density. Thus, for $0 < \theta < 1$, the distribution is shifted to the left, the mean and variance are smaller than in the baseline model, θ can be seen as a wholesale discount factor. Whereas, for $\theta < 0$, the distribution is shifted to the right obtaining larger mean and variance. Figure 2.3 shows the changes in the given density f for a positive and a negative value of θ. In general, the random variable with cdf F_θ is $Y = (1-\theta)\cdot X$. The moment generating function of Y is $M_Y(t) = M_X((1-\theta)t)$, where $M_X(t) = \mathbb{E}(e^{tX})$. We easily see that $\mathbb{E}(Y) = (1-\theta)\mathbb{E}(X)$ and $Var(Y) = (1-\theta)^2 Var(X)$. Then if $0 < \theta < 1$, $\mathbb{E}(Y) < \mathbb{E}(X)$ and $Var(Y) < Var(X)$. Otherwise, if $\theta < 0$, the contrary follows. The skewness and kurtosis are the same as they are compensated by the variance.



(a) (b)

Figure 2.1: Examples of (a) distortion densities $h_{\underline{\alpha}}$ for $\alpha \in \{0.1, 0.2, 0.3\}$, and (b) distortion densities $h_{\overline{\alpha}}$ for $\alpha \in \{0.7, 0.8, 0.9\}$.

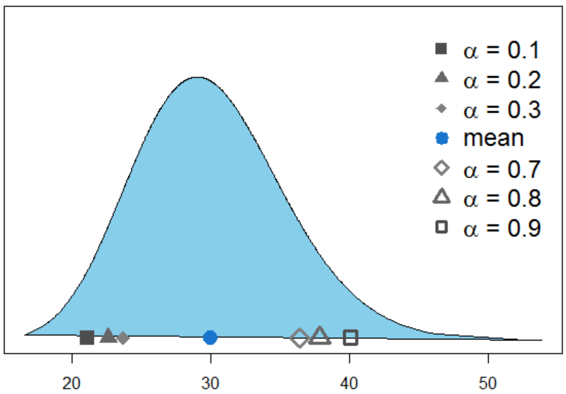


Figure 2.2: Example of lower and upper average value-at-risk prices for a Gamma distribution. The colored area is the Gamma density. The lower average-value-at-risk is calculated for $\alpha \in \{0.1, 0.2, 0.3\}$, and the upper value-at-risk is calculated for $\alpha \in \{0.7, 0.8, 0.9\}$.

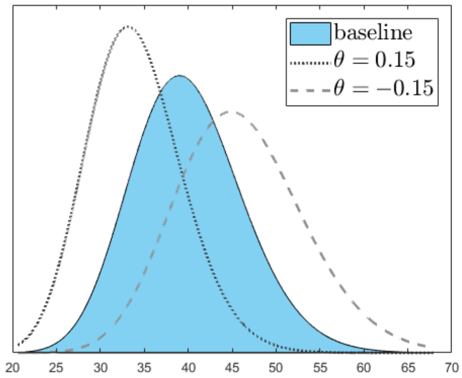


Figure 2.3: Comparison of a given density f (coloured area) and the densities f_θ (dashed lines) corresponding to positive and negative shifting parameters.

The combination of the distortion and the shifting parameter transform the initial cdf for the calculation of the shifted distortion premium in (2.12). We illustrated how the sign of the shifting parameter is closely related to the mean and variance, which are involved in the equilibrium model in Bessembinder and Lemmon 2002 [10]. We also illustrated how the monotonicity of h and its steepness influence in the magnitude and sign of the risk premia. In our example, the distortion density is a step function, but in the literature, continuous functions are also well-known (e.g., the Wang distortion Wang 2000 [101]). Our goal is the identification of θ and h behind electricity futures prices under model ambiguity. Below, we add ambiguity to this premium.

2.2.1 Incorporation of model ambiguity

Most of the literature studying the relationship between futures and spot, as well as the risk premia in energy markets, is model dependent. Furthermore, even if we are close to the true model, since the filtration matters, we are including a shifting parameter. However, this parameter is a general quantity and may not be able to capture the different sources of information. To account for model misspecification we incorporate model ambiguity for our estimations. The main purpose to add ambiguity is the calculation of robust prices. Instead of considering only one model (baseline model), we calculate the premium taking into account many models for our random variables. The set of all these models is called *ambiguity set*. In this paper, we build ambiguity sets using the Wasserstein distance (Dobrushin 1970 [20], Villani 2008 [95]). This distance is itself an optimization problem but has a tractable representation under some assumptions proved by Vallender 1974 [89]. Below we present this representation, which will be used for the rest of our calculations.

Definition 2.2. *Let F and G be two cdfs of positive random variables, the Wasserstein distance of order 1 can be written as*

$$\mathrm{WD}(F,G) = \int_0^1 |F^{-1}(v) - G^{-1}(v)|\, dv = \int_0^\infty |F(x) - G(x)|\, dx. \tag{2.14}$$

Although the previous representation is valid for random variables that take values in the real line, we restrict to positive random variables. We recall the solution of the robust distortion premium. Given (2.13), the solution for the robust shifted distortion premium will follow.

The robust distortion premium with increasing distortion density h and baseline model F is defined as

$$\pi_h^\varepsilon(F) = \sup\{\pi_h(G) \;:\; \mathrm{WD}(F, G) \leq \varepsilon\}\,, \tag{2.15}$$

where $\varepsilon > 0$ is the *ambiguity radius.* The ambiguity sets are Wasserstein balls centred at F of radius ε. Under some assumptions,[1] the solution of (2.15) with h increasing is

$$\pi_h^\varepsilon(F) = \pi_h(F) + \varepsilon\,||h||_\infty\,. \tag{2.16}$$

The second term is the ambiguity premium. If h is decreasing, the robust distortion premium is defined as in (2.15) but with an infimum. Its solution is also written as in (2.16), but with negative ambiguity premium. In general, we obtain

$$\pi_h^\varepsilon(F) = \pi_h(F) \pm \varepsilon\,||h||_\infty\,, \tag{2.17}$$

where the ambiguity premium is $\pm\varepsilon\,||h||_\infty$. If we include the shifting parameter, the robust shifted distortion premium with $\theta < 1$ has the following closed-form (under the same assumptions as before)[2]

$$\pi_h^\varepsilon(F_\theta) = \pi_h(F_\theta) \pm \varepsilon\,||h||_\infty\,. \tag{2.18}$$

The relation between (2.17) and (2.18) is given in the Appendix B.1

We see that the price to pay for ambiguity only depends on the radius of the ambiguity set and the distortion. On the following, we formulate the risk premium of pricing with the shifted distortion premium. Notice that we shall not mix it with the ambiguity premium. Then the risk premia can be calculated as follows

$$\begin{aligned}
\pi_h(F_\theta) - \mathbb{E}(F) &= \int_0^1 F_\theta^{-1}(v)\,h(v)\,dv - \int_0^1 F^{-1}(v)\,dv \\
&= \int_0^1 (F_\theta^h)^{-1}(v) - F^{-1}(v)\,dv \\
&= \pm\mathrm{WD}(F, F_\theta^h)\,.
\end{aligned}$$

Then we can write

$$\pi_h(F_\theta) = \mathbb{E}(F) \pm \mathrm{WD}(F, F_\theta^h)\,. \tag{2.19}$$

[1]For this result we need h to be bounded and X to have a finite expectation. For generalizations see Escobar and Pflug 2018 [25]

[2]With h bounded, we require F_θ to have finite mean, hence θ.

If we incorporate the ambiguity premium in the previous we have

$$\pi_h^{\varepsilon}(F_\theta) = \mathbb{E}(F) \pm \mathrm{WD}(F, F_\theta^h) \pm \varepsilon\,||h||_\infty\,. \tag{2.20}$$

The robust shifted distortion premium can be written the expected loss plus a risk premium, whose magnitude is the area between F and F_θ^h, and an ambiguity premium.

Remark 2.4. *We can also decompose the risk premium* $\pm\mathrm{WD}(F, F_\theta^h)$ *in two terms: one that only depends on* h *and the other as a percentage of a distortion premium.*

$$\begin{aligned}\pi_h^{\varepsilon}(F_\theta) &= (1-\theta)\,\pi_h(F) \pm \varepsilon\,||h||_\infty \\ &= \pi_h(F) - \theta\,\pi_h(F) \pm \varepsilon\,||h||_\infty \\ &= \mathbb{E}(F) \pm \mathrm{WD}(F, F^h) - \theta\,\pi_h(F) \pm \varepsilon\,||h||_\infty\,.\end{aligned}$$

Hence, the risk premium $\pm\mathrm{WD}(F, F_\theta^h) = \pm\mathrm{WD}(F, F^h) - \theta\,\pi_h(F)$*. Where the first term is positive (negative) for increasing (decreasing)* h*. The second terms is positive for* $\theta < 0$*, and negative for* $0 < \theta < 1$*.*

In this paper, we choose to price electricity futures following the pricing rule of the robust shifted distortion premium, as foretold in (2.3). We prove the risk premium and the ambiguity premium have closed forms. Note that the risk premium we calculate here is an ex-ante risk premium (expectations are involved) and we separate it as a sum of a distortion risk premium ($\pm\mathrm{WD}(F, F^h)$) and an information premium ($-\theta\,\pi_h(F)$) for interpretation purposes. While θ adjusts the cdf F, the incorporation of ambiguity involves many models and its closed form of depends only on h and the radius ε. Since futures have time and space components (trading time and delivery period), we will estimate different shape parameters, distortion functions and ambiguity radius for different futures. Also note that in (2.3), the shifted distortion premium is applied to the average spot prices X^m, given the information of spot prices up to trading time. In this section, we omitted the calculation of this premium with conditional information not to overload the notation. Before the identification of our pricing rule's elements, we present a model for the spot prices in the upcoming Section.

2.3 Spot prices model and futures prices

Electricity prices show certain characteristics like seasonality, mean reversion or volatility clustering. A specific feature is the "spiking behaviour": prices on the spot market may jump from one hour to the next to extremely high or low levels, possibly remain there for some hours, and then move back to the original level. This occurs in situations when it is difficult to balance load and generation in the power grid, e.g., due to large demand or a high percentage of infeed from renewable energies (wind, photovoltaic). Then already relatively small changes in the load or generation can cause extreme price changes between consecutive hours. The fact that electricity cannot be stored efficiently prevents an inter-temporal smoothing of the load.

In particular, even negative prices (i.e., paying for generation) may occur mainly at times when a large infeed from renewable generation meets a low demand like in the night or on weekends. Since renewable energies have priority, conventional generation must be shut down and later ramped up again, which can be associated with significant costs in case of inflexible coal-fired plants. Therefore, some units rather accept negative prices to avoid switching off. On the other hand, renewable generation is largely not affected by negative market prices since it is compensated by a fixed premium in the German market design. To have a closer look into electricity prices and production costs, Figure 2.4 shows the evolution of monthly spot price averages for base and peak load with delivery in Germany/Austria at the EPEX.

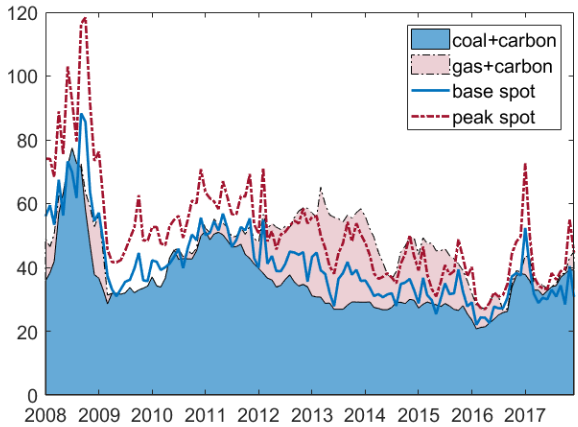


Figure 2.4: Base and peak monthly averages compared to marginal prices of coal and gas. Coal and gas values include carbon emissions.

Spot prices increased up to September 2008 and then dropped by nearly two thirds in less than one year. This decline goes along the collapse of energy and commodity prices in the onset of the global recession following the financial crisis. To illustrate the relationship between primary energy and electricity prices, we compare them with the costs of coal- and gas-based generation (monthly averages). Coal and gas are the most common fossil fuels used for electricity generation in Germany with shares of more than 40% and 10%, respectively, during the observation period (AG Energiebilanzen, 2018 [23]). The comparison of generation costs, expressed in EUR/MWh, instead of market prices of the fuels reveals the impact on electricity prices more clearly since production costs determine the market clearing price. See details in the Appendix B.3.

It is noticeable that base-load prices move largely parallel to the costs of coal-based production. This is consistent with findings in the literature according to which the impact of carbon and coal prices on electricity wholesale prices has (almost) been twice as high as the effect of subsidies for renewable energy (wind, photovoltaic), see Everts et al. 2016 [29] and Bublitz et. al 2017 [14]. Hirth 2018 [52] argues that the decreasing effect of the expansion of renewable energies was compensated by Germany's nuclear phase-out so that the net effect of the energy transition on power prices has been negligible. Furthermore, Paraschiv et al. 2014 [71] investigate also the effect of the changes in the input mix in the context of the energy transition on electricity prices for different delivery hours. This study concludes that in general, the sensitivity concerning coal prices remained high, whereas the sensitivity to gas prices decreased after 2011 since gas-fired plants with higher marginal costs are replaced by renewable energies, particularly around noon due to the high photovoltaic infeed. This corresponds to Figure 2.4, where costs of gas-based production exceed also the average peak-load price after 2011 and the spread between base- and peak-load prices decreased significantly over time.

2.3.1 Spot-forward model

Several authors, e.g., Huisman and De Jong 2003 [54] or Weron et al. 2004 [105], propose to reflect the spiking behaviour by regime-switching models, where the price may change from a "base" to a "jump" regime. These changes are governed by an unobservable stochastic process, and the probabilities of remaining in the current regime or switching to another one are described by a transition matrix. In particular, Keles et al. 2012 [58] find that for an electricity price simulation a regime-switching approach leads to a significant improvement compared to classic time series models.

In this paper, we propose to use a Markov regime-switching model to generate simulations of future spot prices, where the price distributions in the individual regimes may differ and depend on the season (summer/winter), day of the week (workday vs. weekend) or time of the day. This flexible modelling framework reflects the above-described characteristic features of spot prices realistically. It also allows to test if the observable large price fluctuations are reflected in the distortion premiums and if their premia are season-dependent, i.e., vary for different months of the year. While only for the pricing of base futures a simple model for the base price dynamics (average of the 24 hourly prices of each day) might have been sufficient, we have chosen here an approach where the electricity price dynamics is modelled in hourly steps. This allows applying the approach also to products without continuous provision of electricity during the delivery periods like peak futures. This framework was proposed in Paraschiv et al. 2015 [72], abbreviated by PFS in the sequel. Below we describe the PFS model and the modifications we propose for our model.

PFS specifications. In order to use different parameter sets for the price distributions in different seasons, weekdays or times of the day, we introduce the functions $\beta := \beta(t)$ and $\gamma := \gamma(t)$ that map time t to a set of indices, but the dependency on t will be omitted in the sequel for simplicity. We denote by ℓ_t^U and ℓ_t^L those limit values at time t that separate the base regime from an upper and a lower spike regime, which relate to extremely large or low (possibly negative) prices. ξ_t^U and ξ_t^L are deviations from these limits in the corresponding spike regimes, which are exponentially distributed. Prices in the base regime are log-normally distributed and vary around a time-dependent level L_t that reflects the typical seasonal pattern of power prices. The price level L_t defines also the location of the regime limits by the relations $\ell_t^U = L_t \cdot \exp(\alpha_\beta^U)$ and $\ell_t^L = L_t / \exp(\alpha_\beta^L)$, respectively, where $\alpha_\beta^U > 0$ and $\alpha_\beta^L > 0$ are season-dependent model parameters. Overall, the formulation of the regime-switching model for the spot price at time t reads as follows:

$$spot_t = \begin{cases} \ell_t^L - \xi_t^L, & \text{if the system is in the lower spike regime,} \\ L_t \cdot \exp(r_t), & \text{if the system is in the base regime or} \\ \ell_t^U + \xi_t^U, & \text{if the system is in the upper spike regime,} \end{cases} \tag{2.21}$$

and a transition matrix specifies the probabilities of remaining in the current regime or switching to another one:

$$\Pi_\gamma = \begin{pmatrix} 1 - \pi_\gamma^{LB} & \pi_\gamma^{LB} & 0 \\ \pi_\gamma^{BL} & 1 - \pi_\gamma^{BL} - \pi_\gamma^{BU} & \pi_\gamma^{BU} \\ 0 & \pi_\gamma^{UB} & 1 - \pi_\gamma^{UB} \end{pmatrix} \tag{2.22}$$

For example, π_γ^{BU} (π_γ^{BL}) denotes the probability that in the next hour an upward (downward) spike occurs, given that the system is now in the base regime, and π_γ^{UB} (π_γ^{LB}) is the probability of a transition from the upper (lower) spike back to the base regime. As the index $\gamma := \gamma(t)$ indicates, the probabilities in (2.22) may depend on season, weekday and time of the day.

The dynamics of the normally distributed random variable r_t in (2.21) is modeled by an autoregressive process with several lags up to 24 hours. This is motivated by "spillover" effects of not anticipated differences between supply and demand for electricity, e.g., unavailable production capacity or temperature fluctuations, that cause price deviations from the expected level L_t and persist for several hours. Furthermore, the price level L_t is a deterministic forecast obtained from an hourly price-forward curve, which consists of two components:

$$L_t = \tilde{L}_t \cdot s_t. \tag{2.23}$$

s_t is the relative weight of the price at time t compared to the annual average and varies around the mean 1.0. It reflects the seasonal pattern characteristics, i.e., prices are typically larger in winter than in summer, higher on workdays than on weekends or holidays, and vary also during the day. More details and the formal specification of s_t are given in Appendix B.5. The price level after correcting for seasonal effects $\tilde{L}_t$ in (2.23) is chosen in PFS to match the prices of base futures observed on the market.

Our model specifications. Aligning the level of spot prices to observed futures prices is not useful for our application as here the premium of futures products is estimated, i.e., futures prices should be an output of the model rather than an input. Therefore, we extend the above outlined model and assume that $\tilde{L}_t$ follows itself the stochastic process

$$d\tilde{L}_t = \mu \tilde{L}_t \, dt + \eta \tilde{L}_t \, dW_t^X \tag{2.24}$$

with $\mu \in \mathbb{R}$, $\eta > 0$ and W_t^X is a Wiener process. The intuition behind this assumption is the empirical evidence that the level of power prices in Germany is mainly determined by prices of coal and CO2 emission allowances since the production is largely coal-based, see Figure 2.4 and description. The choice of a Geometric Brownian motion in (2.24) is a reasonable modelling assumption for prices of fossil fuels. However, here the level $\tilde{L}_t$ is not linked to observable data but derived from the estimation of the spot price process. For convenience, we define $X_t := \ln \tilde{L}_t$ and obtain with Ito's Lemma a process without stochastic volatility term:

$$dX_t = \left(\mu - 0.5 \cdot \eta^2\right) \, dt + \eta \, dW_t^X. \tag{2.25}$$

In order to facilitate the estimation procedure, we define further Y_t as logarithm of the spot price after correcting for the seasonal effects s_t introduced in (2.23):

$$Y_t := \ln(spot_t / s_t). \tag{2.26}$$

The PFS model (2.21) considers short-term (i.e., some hours) deviations of the spot price from the deterministic (absolute) price level L_t by the quantity r_t. Since in our modified approach L_t is not observable, we model instead the adjusted spot price Y_t defined in (2.26) and assume it follows an Ornstein-Uhlenbeck process around the stochastic mean X_t:

$$dY_t = \kappa_\beta (X_t - Y_t) \, dt + \sigma_\beta \, dW_t^Y \tag{2.27}$$

with season-dependent parameters $\kappa_\beta, \sigma_\beta > 0$ and W_t^Y is Wiener process with $dW_t^X \cdot dW_t^Y = 0$. Including the logarithm of the price level X_t in the process for the log prices Y_t allows us to identify the unobservable level from the observed prices during the model estimation. A Kalman filter is used for this purpose, where the state equation is obtained from (2.25) and the measurement equation from (2.27). Further details are described in Appendix B.6.

Finally, we make here the assumption that in the spike regimes price deviations ξ_t^U and ξ_t^L from the regime limits ℓ_t^U and ℓ_t^L are Weibull distributed, which offers a greater flexibility to fit the model to observed data. With these modifications, the complete extended model reads as

$$spot_t = \begin{cases} \ell_t^L - \xi_t^L, & \text{if the system is in the lower spike regime,} \\ s_t \cdot \exp(Y_t), & \text{if the system is in the base regime or} \\ \ell_t^U + \xi_t^U, & \text{if the system is in the upper spike regime,} \end{cases} \tag{2.28}$$

where the process for Y_t is defined by (2.27), and

$$\xi_t^U \sim \text{Wei}(\lambda_\gamma^+, k_\gamma^+), \qquad \xi_t^L \sim \text{Wei}(\lambda_\gamma^-, k_\gamma^-) \tag{2.29}$$

with time-dependent parameters $\lambda_\gamma^+, k_\gamma^+, \lambda_\gamma^-, k_\gamma^- > 0$. The locations of the regime limits are constructed around the stochastic level:

$$\ell_t^U = s_t \cdot \tilde{L}_t \cdot \exp(\alpha_\beta^U), \qquad \ell_t^L = s_t \cdot \tilde{L}_t / \exp(\alpha_\beta^L) \tag{2.30}$$

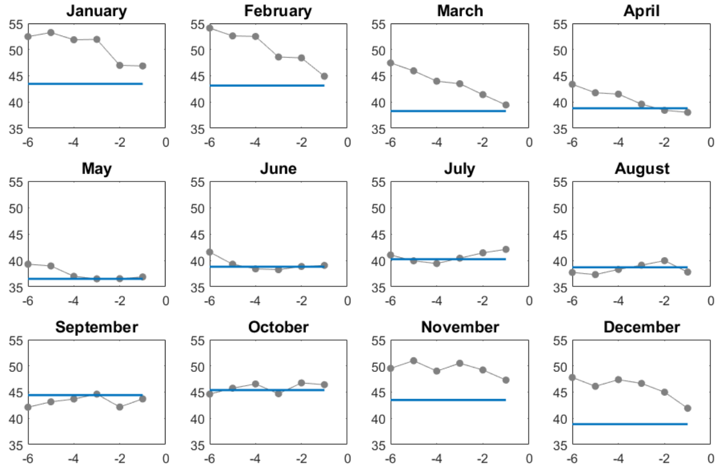
with season-dependent $\alpha_\beta^U, \alpha_\beta^L > 0$. The transition matrix remains as in (2.22). The calibration of the model is outlined in Appendix B.5 and B.6. In particular, it is only based on historical data that is known at the time when the premium is derived and does not exploit any information from the futures market. Preliminary model tests as well as the results in Paraschiv et al. 2015 [72] indicate that two different parameter sets for summer (April to September) and winter (October to March) are sufficient in case of the price process (2.27) for the base regime and the factors $\alpha_\beta^U, \alpha_\beta^L$ that define the location of the regime limits, i.e., $\beta \in \{1, 2\}$. In case of the parameters for the spike regimes and the transition matrices, it should be distinguished between workdays vs. weekends/holidays, day vs. night hours and summer vs. winter, i.e., $\gamma \in \{1, \ldots, 8\}$. Note that the parameters μ, η of the process for the unobservable adjusted level $\tilde{L}_t$ in (2.24) are not season-dependent.

2.3.2 Electricity futures

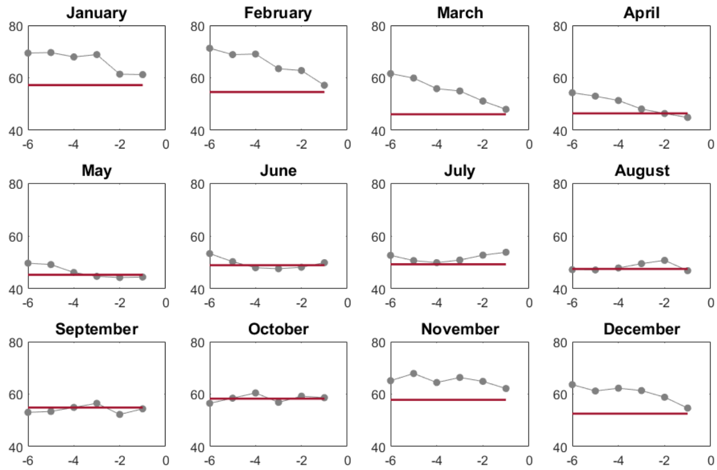
The volumes traded in the day-ahead market in Germany increased significantly over time. In particular, the volume traded by the transmission system operators (TSO) was reported to be of 81 TWh in 2010 versus 20 TWh in 2009 (see Erni 2012 [24]). Although the day-ahead market in Germany is very liquid, producers and consumers have both incentives to trade in the future market to hedge their market exposures. To be precise, to monthly futures we study in this paper are Phelix futures traded in the EEX with monthly deliveries and underlying the average spot prices in (2.1).

Phelix monthly futures can be traded up to six months ahead, and they can also be traded during delivery (in which case, there is no physical delivery), but in this paper, we study futures traded before delivery. Precisely, for a fixed delivery month m, we choose the last settlement price at each of the six previous months. Let $\tau(k, m)$ be the corresponding trading times, each at $k-$th month(s) before m. Then, for a fixed delivery month m we select six different futures: $\mathbb{F}(\tau(k, m), m)$, for $k = 1, \ldots, 6$. Figure 2.5 shows the average evolution of base and peak futures towards the realized monthly averages over 2008-2017. In general, it is observed that for delivery in winter (October to March), futures tend to be above the mean, while in summer (April to September) they oscillate around the mean. Additionally, futures are more separated from their pertaining realized monthly averages in winter than in summer. It is also clear that peak futures are greater than base futures. In average, base futures prices move in a range from 34 €/MWh to 47 €/MWh, while peak futures take values between 41 €/MWh and 61 €/MWh. We also illustrate the dynamics of futures prices traded one month ahead delivery versus futures traded 6 months ahead delivery. Figure 2.6 shows these prices for base and peak profiles. Prices are illustrated above their delivery month from 2008-2017. As we observed before an average pattern of futures, in Figure 2.6, we can observe a seasonal pattern, with higher prices in winter than in summer.

In this paper, we assume base and peak futures are priced following the robust shifted distortion principle, with closed-form solution in (2.18). We specify how this premium applies to our futures prices and their risk premia. For a fixed delivery month m, and fixed time to maturity, say k months before m starts, we assume



(a) Base



(b) Peak

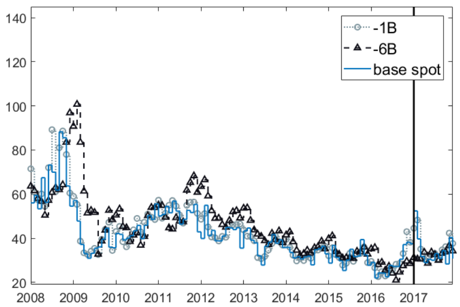
Figure 2.5: For monthly futures that deliver in 2008-2017, we show their average evolution as the trading time approaches to the delivery, i.e., we show average evolution of $\mathbb{F}(\tau(k,m),m)$ (dotted line) per months, for $k=6,\ldots,1$.. They are compared to the mean of realized monthly averages (straight line).

there is a correction factor on the model of X^m, a level of traders' risk aversion given by h, and ambiguity on the knowledge of X^m with information up to trading time $\tau(k,m)$. We write the last settlement price k months before m as

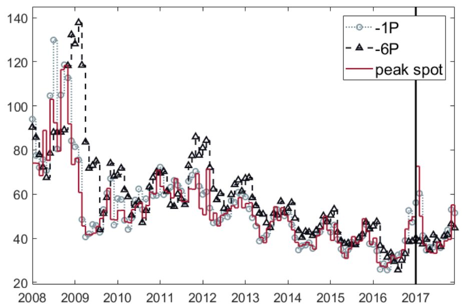
$$\mathbb{F}\left(\tau(k,m),m\right) = \pi_{h;\theta}\left(X_{\tau(k,m)}\right) \pm \varepsilon\,||h||_\infty\,, \tag{2.31}$$

where $X_{\tau(k,m)}$ denotes the random variable X^m adapted to $\mathcal{F}_{\tau(k,m)}$ (the information of spot prices up to trading time $\tau(k,m)$). Recall the value of θ, the shape of h, and the value of ε are not the same for all the futures, but we avoid the time-space dependence not to overload the notation.

In the previous section, we specified our model for spot prices. To simulate values of



(a) Base



(b) Peak

Figure 2.6: Comparison of base (B) and peak (P) futures traded one month ahead and 6 months ahead against the realized monthly averages. The months in the x-axis are delivery months.

$X_{\tau(k,m)}$ (monthly average with conditional information), we first calibrate our model for spot prices using only information up to trading time $\tau(k,m)$. Once the model is calibrated we generate different paths of spot prices over the delivery month m, and average each simulated path to obtain a finite sample of $X_{\tau(k,m)}$. Thus, for fixed m and k, we have an empirical cdf of $X_{\tau(k,m)}$, which we denote by $F_{X_{\tau(k,m)}}$. These samples serve for the calculation of the shifted distortion premium. In the next section, we discuss a procedure to recover θ, h and ε from these futures. With our estimations, we intend to explain the price formation of futures and their risk premia. With our empirical distributions, the ex-ante risk premia are

$$r_{ante}\left(\tau(k,m),m\right) = \pm \mathrm{WD}\left(F_{X_{\tau(k,m)}}, (F_{X_{\tau(k,m)}})^h\right) - \theta\, \pi_h(F_{X_{\tau(k,m)}})\,. \tag{2.32}$$

The ex-post risk premia are

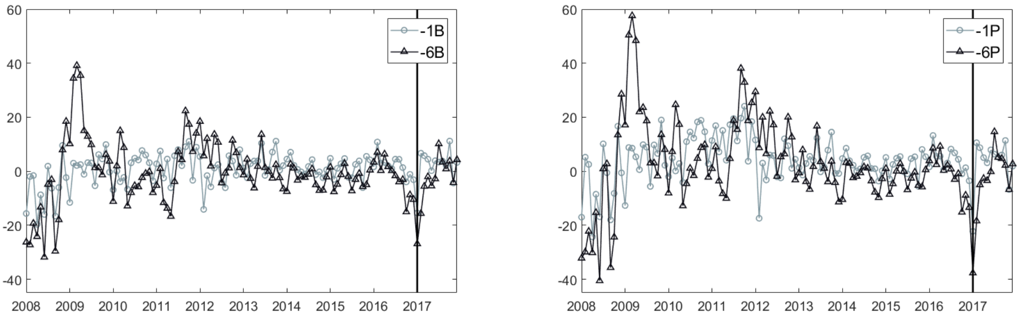
$$r_{post}\left(\tau(k,m),m\right) = \pi^{\varepsilon}_{h;\theta}(F_{X_{\tau(k,m)}}) - X^m\,, \tag{2.33}$$

where X^m is the realized monthly average for each delivery month. The relationship between the two premia is

$$r_{post}\left(\tau(k,m),m\right) = r_{ante}\left(\tau(k,m),m\right) + \text{ ambiguity premium } + \text{ forecast error}\,. \tag{2.34}$$

Figure 2.7 shows the forecast error using the model presented in 2.3.1. We observe the forecast error is higher in 2008-2009 compared to the rest years. We justify this by the impact of the financial crisis and its effect on the drop in production costs (see Figure 2.4). Later, in 2011 we also observe large forecast errors. Our model overestimates expected monthly averages, possibly given by the increasing gas generation prices, which even were larger than peak futures.

In the next section, we discuss a procedure to estimate θ, h and ε by minimizing the gap between the left and right side in (2.31).



(a) Base (b) Peak

Figure 2.7: Forecast error for (a) base monthly averages and (b) peak monthly averages. We compare the forecast error when the model parameters are estimated with information available 6 months before delivery versus 1 month before delivery.

2.4 Recovering procedure for groups of futures

A non-parametric approach to estimate distortion densities for simulated prices was presented in Escobar and Pflug 2018 [25]. The goal of this paper is not only to identify h but also θ and ε. Differently from the previous paper, we identify these parameters for prices in a real market. Additionally,we allow for h to be decreasing.

The main idea for the identification is to minimize the difference between observed futures and empirical prices calculated with our samples, given by the robust shifted distortion premium , i.e., we minimize the right and left-hand side of (2.31). The procedure presented in Escobar and Pflug 2018 [25] is an identification procedure for groups of contracts, with the primary assumption that each contract in the same group, is priced with the same distortion. Therefore, we will not be able to recover θ, h and ε for each futures $\mathbb{F}(\tau(k,m),m)$, but for a group of futures. In other words, we separate futures in different groups, with the main assumption that every contract in the same group is priced with the same value θ, distortion density h and will have the same ambiguity radius ε. Before we specify the different groups of futures, we describe the identification procedure for a general group of contracts. Let us start recalling the recovering of only the distortion density (for h increasing).

Assume C different contracts are priced following the distortion premium with fixed distortion density, say h. Let $\pi^{(j)}$ be the corresponding prices of these contracts given by

$$\pi^{(j)} = \pi_h(X^{(j)})\,, \tag{2.35}$$

where $X^{(j)}$ is the respective underlying of each contract $j = 1, \ldots, C$, and fixed h. To identify h, we introduce a non-parametric function $\hat{h}$ fulfilling appropriate constraints (increasing and integrates 1 in $[0,1]$). Let $(x_{[1]}^{(j)}, \ldots, x_{[n]}^{(j)})$, be an iid ordered sample of $X^{(j)}$, the distortion premium for each sample is

$$\widehat{\pi}^{(j)} = \sum_{i=1}^{n} \left(x_{[i]}^{(j)} \cdot \int_{\frac{i-1}{n}}^{\frac{i}{n}} \widehat{h}(v)\, dv \right) . \tag{2.36}$$

In order to capture different shapes of distortion densities, we define $\widehat{h}$ as either a step function, or as a combination of splines functions. As a step function, $\widehat{h}$ is given by

$$\widehat{h}(v) = \sum_{k=1}^{l} \lambda_k \cdot \mathbb{1}_{[\frac{k-1}{l}, \frac{k}{l})}(v)\,, \tag{2.37}$$

where $\lambda_k > 0$, for $k = 1, \ldots, l$, are the steps defined on l equal-sized subintervals in $[0, 1]$. Optionally, we can also write $\widehat{h}$ as a combination of a base of splines $\{S_{-2}, S_{-1}, S_0, \ldots, S_L\}$

$$\widehat{h}(v) = \sum_{k=-2}^{L} \lambda_k \cdot S_k(v)\,, \tag{2.38}$$

where S_k are translations of S_0, for $k = -2, \ldots, L-1$, and S_L is a constant function with all $\lambda_k \geq 0$. The details on the construction of this base are in Escobar and Pflug 2018 [25].

For the estimation of $\widehat{h}$, we minimize the distance between market prices $\pi^{(j)}$ and empirical prices given in (2.36). We solve

$$\begin{aligned} \min \quad & \sum_{j=1}^{C} \left(\hat{\pi}^{(j)} - \pi^{(j)}\right)^2 \\ \text{s.t.} \quad & (\widehat{h} \text{ increasing and density})\,. \end{aligned} \tag{P1}$$

The average value-at-risk is the most popular premium defined with a step distortion density. Nevertheless, since step functions are constant over an interval, and even constant zero in the case of the AV@R, they do not capture all the information of the distribution (see Balbas et al. [5]). For this reason, we also use continuous distortion densities, which are also common in insurance pricing (e.g., Wang premium, Wang 2000 [101]).

In this paper we estimate not only h, but also θ and ε. For a group of contracts we match

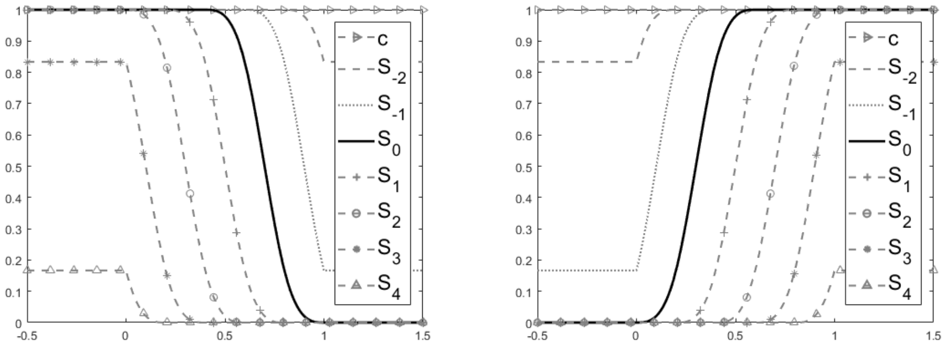
$$\pi^{(j)} = \pi^{\varepsilon}_{h;\theta}(X^{(j)})\,, \tag{2.39}$$

with same θ, h and ε, for all $j = 1, \ldots, C$. In addition to $\widehat{h}$, we introduce a parameter $\widehat{\theta}$ and ambiguity $\widehat{\varepsilon}$. The empirical prices are

$$\widehat{\pi}^{(j)} = (1 - \widehat{\theta}) \sum_{i=1}^{n} \left(x^{(j)}_{[i]} \cdot \int_{\frac{i-1}{n}}^{\frac{i}{n}} \widehat{h}(v)\, dv \right) \pm \widehat{\varepsilon}\, ||\widehat{h}||_\infty\,, \tag{2.40}$$

where the ambiguity premium is positive (negative) for increasing (decreasing) $\widehat{h}$. If $\widehat{h}$ is given as a step function, then we substitute (2.37) in (2.40) and calculate the corresponding empirical prices as

$$\widehat{\pi}^{(j)} = (1 - \widehat{\theta}) \sum_{k=1}^{l} \frac{\lambda_k}{n} \cdot \left(\sum_{i=(k-1)L+1}^{L\cdot k} x^{(j)}_{[i]} \right) \pm \widehat{\varepsilon}\, ||\widehat{h}||_\infty\,, \tag{2.41}$$



(a) Increasing (b) Decreasing

Figure 2.8: Base of cubic monotone (a) decreasing and (b) increasing splines of dimension 8.

where $L = n/l$. Alternatively, if $\widehat{h}$ is a spline, the empirical prices are

$$\widehat{\pi}^{(j)} = (1-\widehat{\theta}) \sum_{i=1}^{n} x_{[i]}^{(j)} \cdot \left(\sum_{k=-2}^{L} \lambda_k \, A_{ik} \right) \pm \widehat{\varepsilon} ||\widehat{h}||_\infty \,, \tag{2.42}$$

where $A_{ik} = \int_{\frac{i-1}{n}}^{\frac{i}{n}} S_k(v)\, dv$ and $S_k(v)$ for all $k = -2, \ldots, L$.

In order to recover h, θ and ε, we solve

$$\begin{aligned} \min \quad & \sum_{j=1}^{C} \left(\hat{\pi}^{(j)} - \pi^{(j)} \right)^2 \\ \text{s.t.} \quad & (\widehat{h} \text{ distortion density}) \\ & \widehat{\theta} < 1;\ \widehat{\varepsilon} \geq 0\,. \end{aligned} \tag{P2}$$

The problem (P2) is to be solved separately for step functions and splines. In this paper, (P2) also allows for decreasing distortion densities. For this reason, we also solve (P2) separately imposing $\widehat{h}$ to be increasing or decreasing. The constraints for the increasing case are in Escobar and Pflug 2018 [25]. If $\widehat{h}$ is decreasing and is given as a step function, we simply impose decreasing steps. Otherwise, if $\widehat{h}$ is a decreasing spline, then we change the base of splines to decreasing ones. The base of decreasing splines, say $\{\tilde{S}_{-2}, \tilde{S}_{-1}, \tilde{S}_0, \ldots, \tilde{S}_L\}$, is easily obtained from the base of increasing splines. Take $\tilde{S}_k(v) = S_k(1-v)$, for all $k = -2, \ldots L-1$, and $\tilde{S}_L = S_L$ (the constant 1 is included in the base), where S_k are the increasing splines in Escobar and Pflug 2018 [25]. Both, the decreasing and increasing bases are shown in Figure 2.8.

We recall the density constraints of $\widehat{h}$. For $\widehat{h}$ as a step function we have

$$\sum_{k=1}^{l} \lambda_k = l\,. \tag{2.43}$$

Otherwise, if $\widehat{h}$ is a spline we have

$$\sum_{k=-2}^{L} \lambda_k \, a_k = 1 \,, \tag{2.44}$$

with $a_k = \sum_{i=1}^{n} A_{ik}$. For the solution of (P2), we define a new parameter, say $\mu_k = (1-\theta)\,\lambda_k$, for all $k = 1, \ldots, l$ ($k = -2, \ldots, L$, when splines are considered). Since we solve (P2) with constraints on μ, (2.43) translates to

$$\sum_{k=1}^{1} \mu_k = (1-\theta) \cdot \sum_{k=1}^{l} \lambda_k = (1-\theta)\, l \,,$$

which is always positive for $\theta < 1$. In the case of splines, (2.44) translates to

$$\sum_{k=-2}^{L} \mu_k \, a_k = (1-\theta) \cdot \sum_{k=-2}^{L} \lambda_k \, a_k = 1-\theta \,,$$

which is always positive for $\theta < 1$. Concerning the ambiguity, we define $\delta = \pm\varepsilon \, ||h||_\infty$, and impose $\delta \geq 0$ for increasing $\widehat{h}$, or $\delta \leq 0$ for decreasing $\widehat{h}$. As a final technicality, we recall the solution of (P1) is supported by Prop. 6 in Escobar and Pflug [25]. By introducing, θ and ε, this result no longer holds and we need to bound θ and ε. As for the number of steps or splines, we will choose 10 steps and 8 elements in the base of splines, precisely shown in Figure 2.8.

2.4.1 Groups of futures

We solve (P2) for different groups of futures presented in Section 2.3.2. There is a trade-off between the number of contracts in each group and the number of distinguishable groups. Each group must have enough contracts to estimate the weights for the step and spline distortion, the shifting parameter and the ambiguity radius. At the same time, each group must be distinct from each other, in the sense that, they should be well categorized among unique characteristics. In this paper, we decide to separate futures among the following categories.

- Profile: base futures are separated from peak futures.
- Delivery season: futures will be separated according to the season of its delivery. The seasons are winter (October to March) and summer (April to September).
- Time-to-maturity: futures traded 1, 2, 3, 4, 5 and 6 months ahead delivery will be separated among each other.

By clustering futures according to these categories, we are able to test whether the estimated parameters are significantly different in each group. Since the estimations are in groups we will estimate an average pattern that persists over the years. Figure 2.6 also shows a priori the differences among these categories.

Since we aim to recover persistent behaviour over the years, the training data we choose needs to be rather homogeneous among these categories. Futures with delivery in 2008 and 2009 are strongly linked to the financial crisis. In Figure 2.6 we observe futures with delivery at the beginning of 2008 were not significantly high, but from January 2009 futures traded during 2008 to deliver in the next months suffer a peak to almost 100 €/MWh and drop down to 30 €/MWh. We also notice the large forecast errors (Figure 2.7) at the beginning of 2009. Additionally, the liquidity of this market starts to grow after 2009 as we mentioned above. Therefore, our training set includes all futures with delivery from 2010 to 2016. The last year of our sample data 2017, will be reserved for out-of-sample results.

For instance, one of these groups contains all base futures with delivery in winter traded one month ahead. Thus, all the contracts in this group are base futures $\mathbb{F}(\tau(1,m),m)$, where the delivery month m is every winter month included in the training set. To recover θ, h and ε for his group, we take the samples of $X_{\tau(1,m)}$ for every winter month in the training set and solve (P2).

Table 2.1 shows the different groups, each with 42 (= 6 months per year × 7 years) different contracts. For each of them, we recover a distortion density, denoted by $\widehat{h}_k^W$ or $\widehat{h}_k^S$, where the superscript indicates whether the delivery is in winter (W) or summer (S); and the subscript denotes time-to-maturity (1 to 6 months ahead). The distinction between base and peak profiles will be clear from the context. For our training sample, we aim at recovering the main characteristics of futures prices and their risk premia driven by three different factors, a shifting parameter, risk aversion and ambiguity in the model. Our goal shall not be confused with forecasting futures.

Profile	Base/Peak		
Season for m	Summer/ Winter		
Time-to-maturity (months)	1	$\cdots$	6
Contracts	$\mathbb{F}(\tau(1,m),m)$	$\cdots$	$\mathbb{F}(\tau(6,m),m)$

Table 2.1: Grouping of contracts for base and peak futures.

2.5 Identification results in power markets

On the following, we present the results of solving (P2) for each group of futures and infer conclusions on the price formation of monthly futures in the EEX.

For the upcoming results, recall that the sign of θ, the shape of h, and the value of ε intend to explain different aspects of monthly futures price formation. If the estimated shifting parameter (for one group) is $\widehat{\theta} < 0$, then our samples (given solely with spot prices information) are in general underestimating the expected monthly averages. In that case, this correction factor adds a positive premium by distorting the empirical distributions. The opposite is derived if $\widehat{\theta} > 0$. The estimated $\widehat{h}$ takes the shifted distributions of the pertaining group and distorts them. If $\widehat{h}$ is increasing, the expectations are larger with the distortion than without it. If $\widehat{h}$ is decreasing, then the expectations are smaller with the distortion than without it. In general, for increasing distortion densities, the more weight the distortion density puts to higher quantiles, the larger the final price (as the final price lies in the upper tail of the shifted distribution). For decreasing distortion densities, the more weight $\widehat{h}$ puts to lower quantiles, the lower the final price. Recall Figure 2.3 and Figure 2.1, where different values for θ and different distortion densities are compared. Finally, an ambiguity premium is added if $\widehat{h}$ is increasing. The ambiguity premium is negative when $\widehat{h}$ is decreasing. The size of this extra load depends on the ambiguity radius and the tightest upper bound of $\widehat{h}$. As a final remark, if $\widehat{\theta} = 0$ we do not shift the distributions, if $\widehat{h}(v) = 1$ (the constant 1), there is no distortion, and if $\widehat{\varepsilon} = 0$ we do not include an ambiguity premium. When all these values are taken, the price is simply the expectation of our empirical distributions, see (2.31).

2.5.1 Base futures

For each group of base futures, we show the solutions to solving (P2) with step and spline functions. Table 2.2 presents their estimated shifting parameters, ambiguity radii and indicates the monotonicity of their estimated distortion densities. Table 2.3 specifies the corresponding weights when $\widehat{h}$ is defined as a step function (heights of steps) and when it is defined as a spline (weights for the combination of splines). The distortion densities we estimate are never decreasing. Therefore, the ambiguity premium is always positive and it is also calculated for every group in Table 2.2. Since the splines are increasing, the weights in Table 2.3 are for the base of increasing splines. For each group, the best fit between step and spline distortion densities is illustrated in Figure 2.9.

Step		Winter				Summer			
label	k	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon}$	$\widehat{\varepsilon}\cdot\|\|\widehat{h}\|\|_\infty$	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon}$	$\widehat{\varepsilon}\cdot\|\|\widehat{h}\|\|_\infty$
-1B	1	$\nearrow$	0.1263	0.1177	1.177	$\nearrow$	0.1488	0.1542	1.0716
-2B	2	$\nearrow$	0.1243	0.1347	1.3472	$\nearrow$	0.1802	0.1108	1.1076
-3B	3	$\nearrow$	0.0620	0.2179	1.3634	$\nearrow$	0.1301	0.4059	1.1762
-4B	4	$\nearrow$	-0.0083	1.0227	1.4301	$\nearrow$	0.0446	1.2529	1.3662
-5B	5	$\nearrow$	-0.0696	1.4758	1.6887	$-$	0.0091	1.4456	1.4456
-6B	6	$-$	-0.0927	1.9706	1.9706	$-$	-0.0030	1.7652	1.7652

(a)

Spline		Winter				Summer			
label	k	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon}$	$\widehat{\varepsilon}\cdot\|\|\widehat{h}\|\|_\infty$	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon}$	$\widehat{\varepsilon}\cdot\|\|\widehat{h}\|\|_\infty$
-1B	1	$\nearrow$	0.1350	0.0600	1.1809	$\nearrow$	0.1499	0.1394	1.0724
-2B	2	$\nearrow$	0.1312	0.0761	1.3569	$\nearrow$	0.1780	0.0808	1.1067
-3B	3	$\nearrow$	0.0670	0.1169	1.3668	$\nearrow$	0.1314	0.3711	1.1788
-4B	4	$\nearrow$	-0.0107	1.0222	1.4287	$\nearrow$	0.0373	1.2873	1.3591
-5B	5	$\nearrow$	-0.0637	0.8977	1.6947	$-$	0.0091	1.4456	1.4456
-6B	6	$-$	-0.0927	1.9706	1.9706	$-$	-0.0030	1.7652	1.7652

(b)

Table 2.2: For each group, we show the monotonicity of the estimated distortion density, increasing ($\nearrow$) or constant ($-$), the shifting parameter, the ambiguity radius and the ambiguity premium. Estimations are separated for (a) step distortion densities and (b) splines distortion densities.

For each group, we interpret the values of our estimations. We start by looking

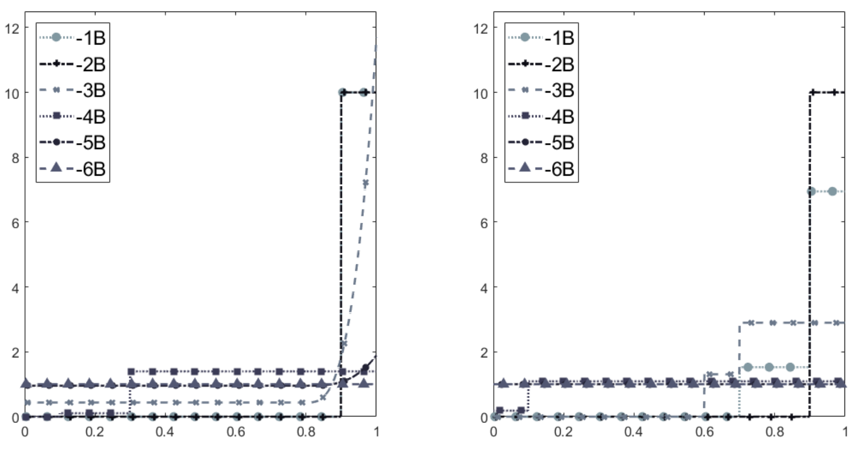
Step		weights									
	k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
W	1	0	0	0	0	0	0	0	0	0	10
	2	0	0	0	0	0	0	0	0	0	10
	3	0.416	0.416	0.416	0.416	0.416	0.416	0.416	0.416	0.416	6.256
	4	0	0.106	0.106	1.398	1.398	1.398	1.398	1.398	1.398	1.398
	5	0.964	0.964	0.964	0.964	0.964	0.964	0.964	0.964	1.144	1.144
	6	1	1	1	1	1	1	1	1	1	1
S	1	0	0	0	0	0	0	0	1.526	1.526	6.948
	2	0	0	0	0	0	0	0	0	0	10
	3	0	0	0	0	0	0	1.308	2.897	2.897	2.897
	4	0.186	1.090	1.090	1.090	1.090	1.090	1.090	1.090	1.090	1.090
	5	1	1	1	1	1	1	1	1	1	1
	6	1	1	1	1	1	1	1	1	1	1

(a)

Spline		weights							
	k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
W	1	0	0	0	0	0	0	0.245	116.819
	2	0	0	0	0	0	0	1.627	98.854
	3	0.437	0	0	0	0	0	0	67.542
	4	0	0	0.113	1.285	0	0	0	0
	5	0.953	0	0	0	0	0	0	5.607
	6	1	0	0	0	0	0	0	0
S	1	0	0	0	0	0	0	9.231	0
	2	0	0	0	0	0	0	4.728	58.531
	3	0	0	0	0	0.235	2.942	0	0
	4	0	0.585	0.470	0	0	0	0	0
	5	1	0	0	0	0	0	0	0
	6	1	0	0	0	0	0	0	0

(b)

Table 2.3: Estimated weights for the (a) steps and (b) for the combination of splines. Estimations are per groups of base futures. Weights are rounded up to 3 decimals.



(a) Winter (b) Summer

Figure 2.9: Estimated distortion densities for each group of base futures with delivery months in (a) winter and (b) summer.

at futures with delivery in winter. The pertaining shifting parameters decrease as time-to-delivery increases, and they change their sign from 4 months ahead. By looking at (2.40), we see how the shifting parameters for 1 and 2 months ahead decrease each sample point to approximately 12% of their value. Recall those positive shifting parameters reduce the empirical means and variances in the pertaining groups. This is the case for futures that deliver in the next 3 months. Our samples (using information from the spot market up to trading time) are overestimating the distribution range of the random variables of monthly averages over winter months (for groups -1B, -2B and -3B). When delivery takes place in further months (from 4 to 6), the shifting parameters are negative. The opposite is concluded for these groups. For futures with delivery in summer, the sign of θ is only negative for the longest time-to-delivery (6 months). Only for this case, a positive correction premium is added, but smaller than the positive premium added in winter. We increase the mean and variance of the empirical distributions in this group. Otherwise, $\widehat{\theta}$ is positive, reaching its maximum value for -2B. Judging from their values, we subtract a larger quantity for these cases in summer than in winter (14-18% vs. 12%). Altogether, the correction premia make delivering in summer cheaper than in winter, where for front months our expectations are decreased and for further deliveries are increased. When delivery is close, traders have a better understanding of other uncertainties that may influence the average spot prices at delivery. Figure 2.4 shows how coal prices (one of the main sources of electricity generation in Germany) are below spot prices. Hence, our samples with spot prices information may be overestimating production costs for these cases and we can see the correction premium as a wholesale discount factor. Although many uncertainties, which are not incorporated in the information of spot prices, may influence futures prices, we are only able to recover a persistent and general pattern over the years 2010-2016. Moreover, if we look at the ambiguity premia (Table 2.2), the longer the time-to-delivery, the more ambiguity premium is incorporated. Traders face more uncertainty of scenarios when delivery is far from the trading time, then an ambiguity premium and a correction premium are added for these groups (the ambiguity premia are positive since all the distortion densities are increasing).

Our pricing rule (2.18) applies the distortion premium to the shifted distributions. The distortion densities together with the shifting parameter will estimate a persistent behaviour present in the risk premia. Figure 2.9 shows the distortion densities for each group. We notice that the estimated distortion densities distort higher quantiles of our shifted samples for groups of futures that deliver in the front months (-1B and -2B). When time-to-delivery increases, the distortion densities gradually distort also smaller quantiles, thus their heights decrease (since they need to integrate 1), giving lower prices. For longer time-to-delivery, they become the constant 1, i.e., distortion is not applied in the pricing. In particular, we obtain this for delivery in winter in -6B and sooner for delivery in summer in -5B and -6B. Precisely, for delivery in winter, $\widehat{h}_1^W$ and $\widehat{h}_2^W$ correspond to the upper $\mathrm{AV@R}_{0.9}$. For 3 months ahead, the weighting is smoother but the distortion is also concentrated on higher quantiles. Subsequently, $\widehat{h}_4^W$ is again a step function, that gradually distorts the quantiles of the initial distributions of this group. The estimated spline $\widehat{h}_5^W$ is close to the constant one and slightly distorts high quantiles. Finally, $\widehat{h}_6^W$ is the constant one and no distortion is applied. With this evolution on the distortion densities, we see how risk aversion changes with time-to-delivery. For delivery months in summer, the largest price is attained at $\widehat{h}_2^S$, which corresponds to the $\mathrm{AV@R}_{0.9}$. The distortion densities $\widehat{h}_1^S$ and $\widehat{h}_3^S$ also distort high quantiles, although their upper bounds are smaller than in winter (for the same labels). The upper bound of $\widehat{h}_4^S$ drops close to one and it is smaller than the one for winter. Lastly, $\widehat{h}_5^S$ and $\widehat{h}_6^S$ do not distort the shifted samples. In summary, we can recover classic distortion premium principles as the average value-at-risk. We also see that the distortion densities give higher prices for delivery in the front months and prices are lower when time-to-delivery increases. Prices increase with demand and are lower when there is more supply. Therefore, for front months, consumers have more risk aversion than producers and try to lock in the price for short time-to-deliveries. On the other hand, producers have more incentives for locking in the price and secure selling their production for further delivery months. In these cases, prices tend to approach the expectation (after the incorporation of a shift). We also see how prices are higher in winter than in summer, from the shifting parameters and the upper bounds of the distortion densities.

We calculate the prices with the distortion densities in Figure 2.9 and the shifting parameters and ambiguity radii estimated. If the distortion density is a step function, the estimated prices are calculated as (2.41), otherwise, if the distortion density is a spline, they are given by (2.42). The comparison between the estimated prices and the observed futures is shown in Figure 2.10. Recall that these parameters are estimated with our in-sample set, which covers futures with delivery from 2010 to 2016. With the same shifting parameters, distortion densities and ambiguity radii estimated, we also calculate prices for 2017 (the out-of-sample set), see Figure B.1 in our Appendix B.7. To show the evolution over time, we put together the estimations for winter and summer and separate them by time-to-maturity. The residuals of the estimations are shown in a box-plot for each group, see Figure B.2 in Appendix B.7. Above each box-plot, we also place the out-of-sample residuals. Appendix B.7 also shows the goodness of fit of (2.31) with our estimations for each time-to-delivery (see Table B.1). This Appendix also contains results of the regressions fitting (2.31), see Table B.2. There we see the shifted distortion principle (estimated including ambiguity) applied to our samples, explain from 93% to 75% the futures prices as time-to-delivery increases.

With the estimated prices, we also compute the estimated ex-ante risk premia as in (2.32). Figure 2.11 compares them to the ex-ante risk premia calculated with the observed futures, following (2.4). In both cases, we use our model for average spot prices. For this illustration, we put together the estimations for summer and winter and separate them by time-to-delivery. With the interactions of the shifting parameters and distortion functions, we recover a seasonal behaviour present in the ex-ante risk premia. Futures with delivery in winter have a higher risk premium than futures that deliver in summer. We also capture a seasonal change of sign for the front months. We estimate that negative risk premia are more likely in summer months than in winter. As time-to-delivery increases, the ex-ante risk premia tend to have less negative values and the magnitude increases, but we also notice the presence of extreme negative ex-ante risk premia (straight line) for some futures with long time-to-delivery in 2011. Figure 2.13 shows the empirical distributions for every month in 2011 calculated with the information given 6 months ahead of delivery. We observe futures prices fall in general above the empirical mean and only for July to October, futures are below, which explains the negative ex-ante risk premia. Spot prices increase at the beginning of 2011, therefore our expectations assume this trend for spot prices in 6 months. The estimated ex-ante risk premia

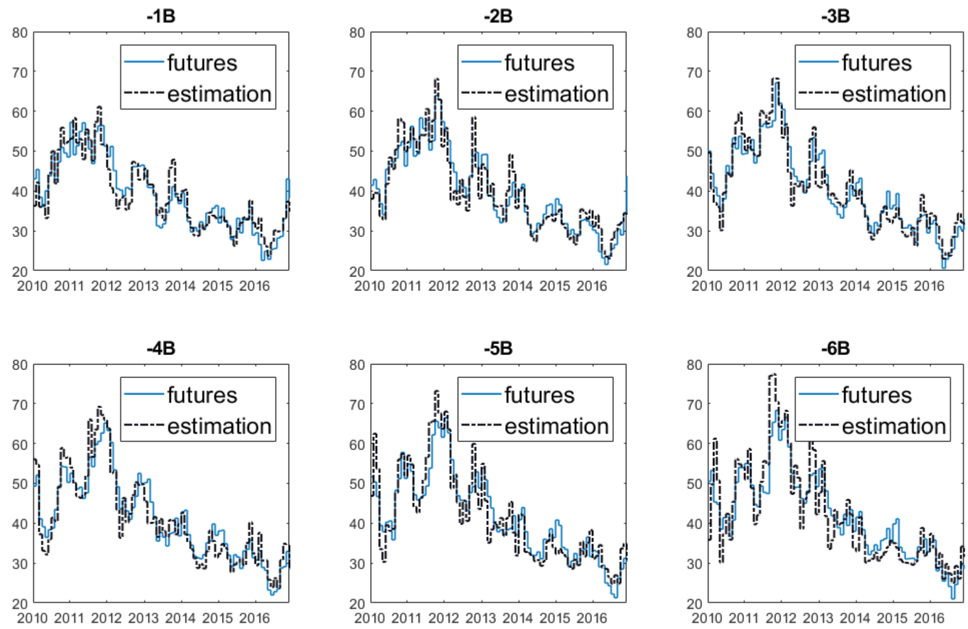


Figure 2.10: Comparison between observed monthly base futures prices and estimated prices. Futures are separated by time-to-maturity.

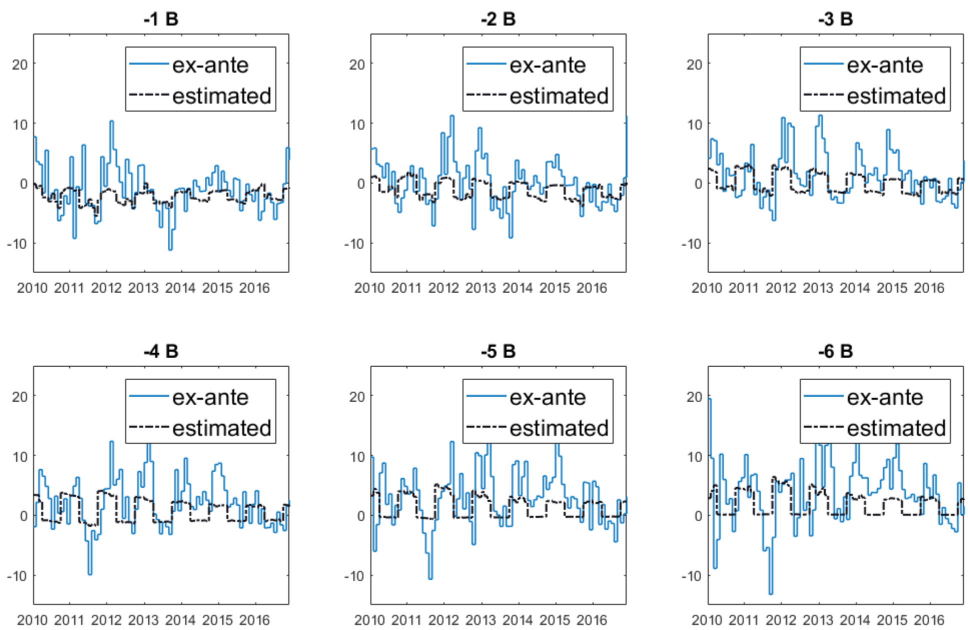


Figure 2.11: Comparison of the ex-ante risk premia calculated with our model using our estimated parameters (dashed line), versus the ex-ante risk premia calculated with the observed futures (line).

do not capture the large negative magnitudes as this is not a persistent event over the years.

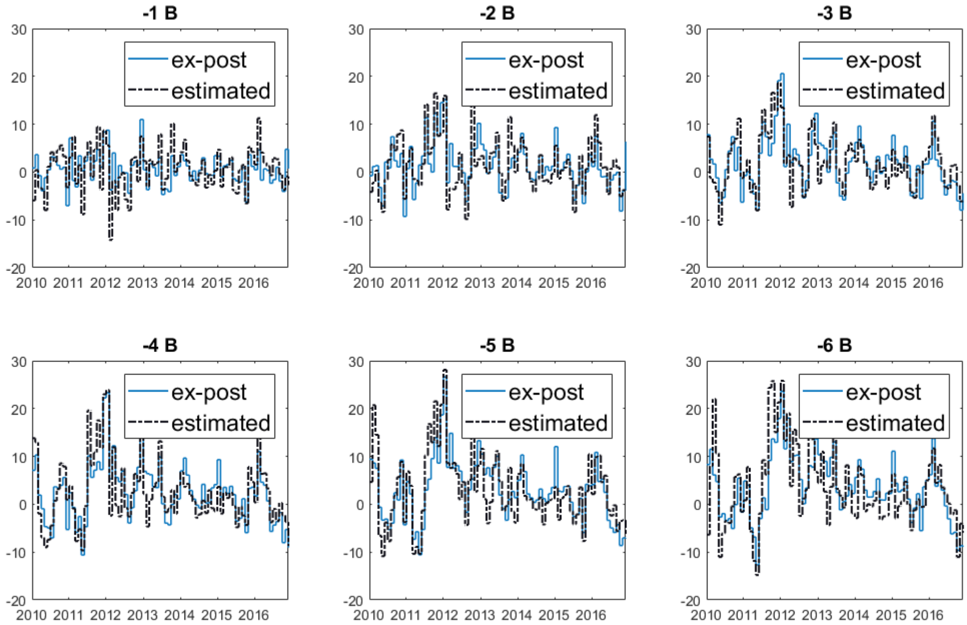


Figure 2.12: Comparison of the ex-post risk premia calculated with our model using our estimated parameters (dashed line), versus the ex-post risk premia calculated with observed base futures (line).

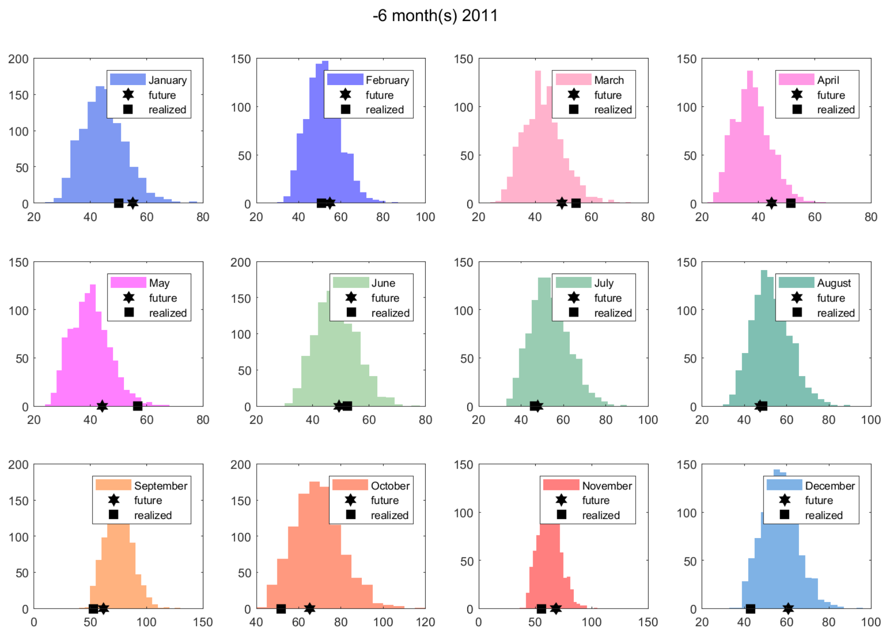


Figure 2.13: Empirical distributions for the average spot prices in 2011 with information 6 months ahead each delivery. For each histogram we show the value of the base futures that delivers in that month (★) and the realized average spot prices (■).

We also compare the observed ex-post risk premia (2.5) with the estimated ex-post premia (see (2.33)) in Figure 2.12. Recall that we can decompose the estimated ex-post risk premia as the sum of the estimated ex-ante risk premia, ambiguity premium and forecast error (see (2.34)). Then, the estimated seasonality, magnitude and change of the ex-ante risk premia are present in the dynamics of the ex-post risk premia. The volatility of the ex-post risk premia can be explained by the forecast errors (see Figure 2.7) with the ambiguity premium, which adds larger quantities as time-to-maturity increases.

2.5.2 Peak futures

Similarly, we show the corresponding solutions for peak futures. Table 2.4 shows the monotonicity of the estimated distortion densities, shifting parameters and ambiguity radii for all groups, where $\widehat{h}$ is a step function and a spline. Figure 2.14 shows the best fit for each group of peak futures between steps and splines and Table 2.5 contains the weights of the distortion densities. It is noticeable that for front months, the estimated distortion densities are constant 1 and then their shape becomes decreasing. $\widehat{h}$ is constant (one) also for further delivery in winter. Since the distortion densities are never increasing, the weights in Table 2.5 for the splines are for the base of decreasing splines.

Step		Winter			Summer		
label	k	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon} = \widehat{\varepsilon} \cdot \lvert\lvert\widehat{h}\rvert\rvert_\infty$	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon} = \widehat{\varepsilon} \cdot \lvert\lvert\widehat{h}\rvert\rvert_\infty$
-1P	1	$-$	0.0908	1.5989	$-$	0.1435	1.3736
-2P	2	$-$	0.0549	1.7944	$-$	0.1301	1.4414
-3P	3	$-$	0.0247	1.8841	$\searrow$	-0.0521	≈ 0
-4P	4	$-$	-0.0008	2.0452	$\searrow$	-0.0807	≈ 0
-5P	5	$\searrow$	-0.3102	≈ 0	$\searrow$	-0.2045	≈ 0
-6P	6	$\searrow$	-0.3833	≈ 0	$\searrow$	-0.2553	≈ 0

(a)

Spline		Winter			Summer		
label	k	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon} = \widehat{\varepsilon} \cdot \lvert\lvert\widehat{h}\rvert\rvert_\infty$	$\widehat{h}$	$\widehat{\theta}$	$\widehat{\varepsilon} = \widehat{\varepsilon} \cdot \lvert\lvert\widehat{h}\rvert\rvert_\infty$
-1P	1	$-$	0.0908	1.5989	$-$	0.1435	1.3736
-2P	2	$-$	0.0549	1.7944	$-$	0.1301	1.4414
-3P	3	$\searrow$	-0.2092	≈ 0	$-$	0.1116	1.6053
-4P	4	$-$	-0.0008	2.0452	$\searrow$	-0.0750	≈ 0
-5P	5	$\searrow$	-0.3131	≈ 0	$\searrow$	-0.2290	≈ 0
-6P	6	$\searrow$	-0.4104	≈ 0	$\searrow$	-0.2798	≈ 0

(b)

Table 2.4: Monotonicity of the estimated distortion density, decreasong ($\searrow$) or constant ($-$), shifting parameter, and ambiguity radius for each group of base futures. Estimations are separated for (a) step distortion densities and (b) splines distortion densities.

Regarding the shifting parameters corresponding to the best fits, $\widehat{\theta} > 0$ up to 3 months ahead and corrects the initial distributions by lowering the empirical means and variances. From this time-to-delivery, the shifting parameters add variance and increase the means. The same pattern is estimated for delivery in summer. Since the distortion densities for the front months (1 to 3) are constant, peak futures in

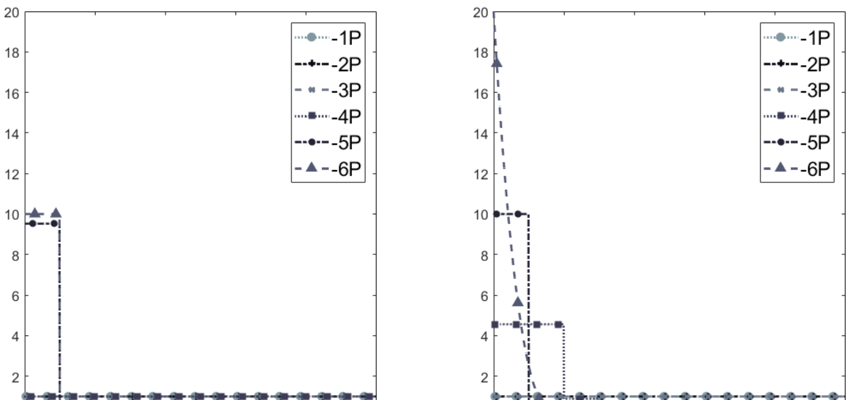
Step		weights									
	k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
W	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	1	1	1
	4	1	1	1	1	1	1	1	1	1	1
	5	9.524	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053
	6	10	0	0	0	0	0	0	0	0	0
S	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1
	3	5.08	4.92	0	0	0	0	0	0	0	0
	4	4.56	4.56	0.88	0	0	0	0	0	0	0
	5	10	0	0	0	0	0	0	0	0	0
	6	10	0	0	0	0	0	0	0	0	0

(a)

Spline		weights							
	k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
W	1	1	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0
	3	0.01	0	0	0	0	0	0	118.804
	4	1	0	0	0	0	0	0	0
	5	0.11	0	0	0	0	0	0	106.832
	6	0	0	0	0	0	0	0	120
S	1	1	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0
	3	1	0	0	0	0	0	0	0
	4	0	0	0	0	0	1.729	4.443	0
	5	0	0	0	0	0	0	0	120
	6	0	0	0	0	0	0	0	120

(b)

Table 2.5: Estimated weights for the (a) steps and (b) for the combination of splines. Estimations are per groups of peak futures. Weights are rounded up to 3 decimals.



(a) Winter (b) Summer

Figure 2.14: Estimated distortion densities for each group of peak futures with delivery months in (a) winter and (b) summer.

these groups are explained by expectations of the shifted distortions (smaller mean and variance) plus the ambiguity radii.

On the other hand, for delivery in 4 to 6 months, we estimate $\widehat{\theta} < 0$. If we only look at delivery in winter, $\widehat{\theta}_4^W \approx 0$ and its corresponding distortion density is again the constant 1, then futures in this group are also estimated as expectations of our empirical distributions plus an ambiguity premium of 2 €/MWh. For delivery in winter in 5 and 6 months, the shifting parameters increase each sample point for at least 30% of their value. The correction premium adds a positive value for these groups. Concerning delivery in summer in 4 months, the correction premium adds greater value than in winter. For delivery in 5 and 6 months, the correction premium is also positive. Although the shifting parameters for further deliveries (4 to 6) are negative and add a positive premium, the estimated distortion densities are decreasing. When distortion densities concentrate most of their mass to small quantiles, the final prices fall in the lower tail of the (shifted) distributions. For these cases, the ambiguity is non-significant, which means we do not need to subtract any extra quantity (the ambiguity premium). Finally, we notice that $\widehat{h}_6^W$ prices with the lower $\mathrm{AV@R}_{0.1}$. Small prices mean large presence of producers when trading peak futures with delivery further in time. Since production is more flexible than demand, we can justify generators are more interested in these products. Even though prices lie in the lower tails, since the range of the empirical distributions takes place in higher values than for base samples, the final prices are still higher than for base futures (compare Figure 2.13 to Figure 2.18).

The estimated prices are compared to the observed peak futures in Figure 2.15. Out-of-sample estimations and the residuals are shown in the Appendix B.7 (see Figure B.1 and Figure B.3). We observe that residuals for peak futures are larger than for base futures, in particular when the time-to-delivery is 4, 5 and 6 months. See Table B.1 for the normalized mean square errors in the Appendix B.7. In this Appendix we also show the results of fitting (2.31) for peak futures. The results of the regression show that up to 3 months ahead (when ambiguity is different than zero), the shifted distortion principle applied to our samples explains around 80% of peak futures. From 4 to 6 months ahead, the regression estimates significant intercept which increases with time-to-delivery. Since the ambiguity radii were estimated equal to zero for these groups (as they have to be negative), our model needs an even larger increasing trend as the one we estimated.

Figure 2.16 illustrates the estimated ex-ante risk premia (given by (2.32)) versus the ex-ante risk premia from the market (calculated with our model). From our estimations, we identify a seasonal pattern present in the ex-ante risk premia for peak futures. We estimate a higher magnitude for futures with delivery in winter than in summer. Moreover, our estimations also recover an increasing trend of the ex-ante risk premia over the years 2010-2016. Likewise, for base futures, we notice significantly negative values of the ex-ante risk premia (straight line) in the second half of 2011 for delivery in 4, 5 and 6 months. Increasing coal prices at the beginning of 2011 resulted in larger values of spot prices, and hence in monthly averages (see Figure 2.4). Figure 2.18 shows our empirical distributions for peak monthly averages in 2011, given with spot prices information 6 months before delivery. We observe how peak futures are in the lower tail of our empirical distributions in July, August, September and October. For the other months, futures prices in general fall close to the mean or below.

Figure 2.17 shows the estimated ex-post risk premia (see (2.33)) versus the observed

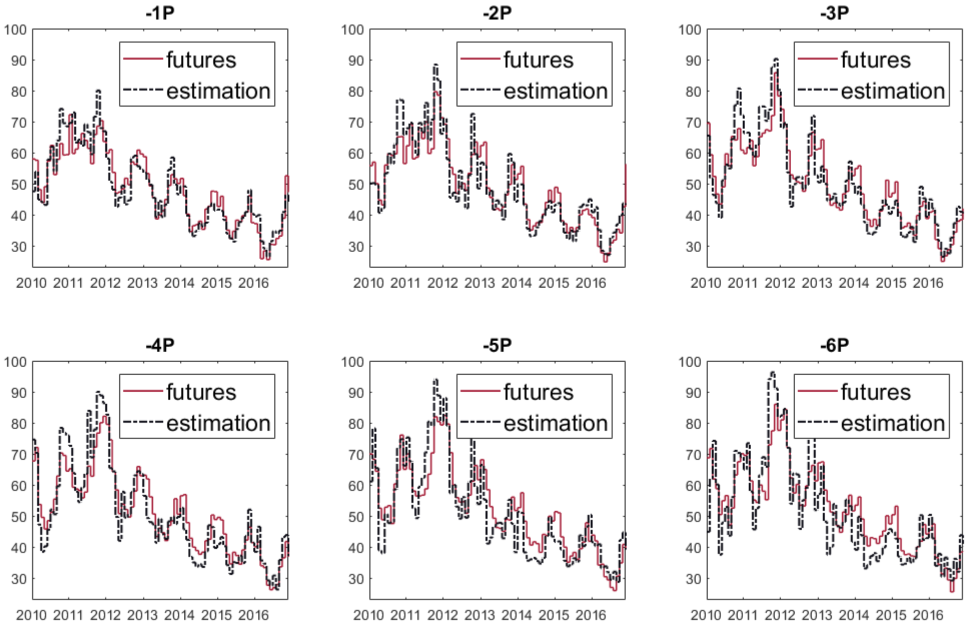


Figure 2.15: Comparison between observed monthly peak futures prices and estimated prices. Futures are separated by time-to-maturity.

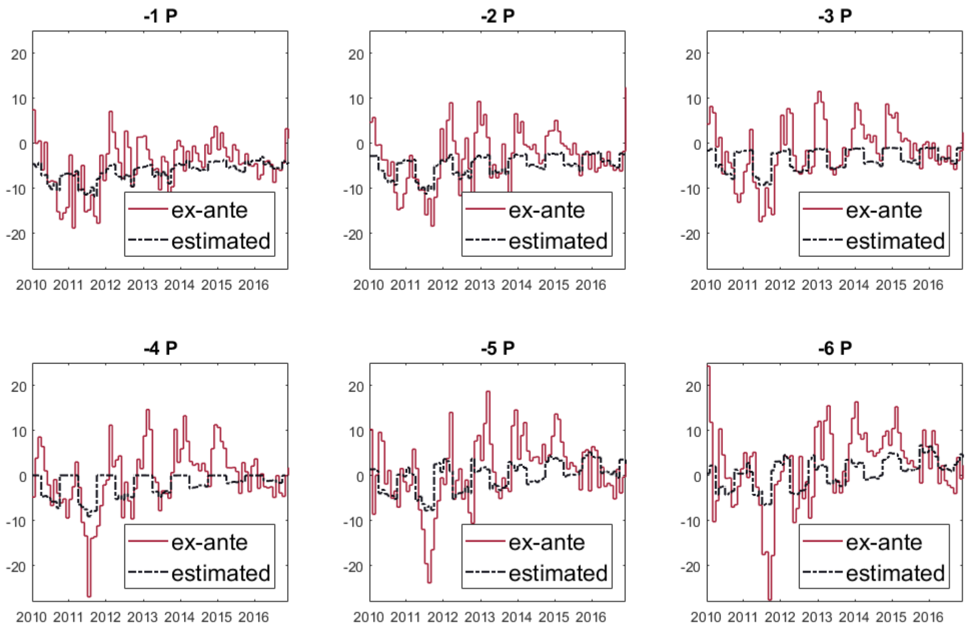


Figure 2.16: Comparison of the ex-ante risk premia calculated with our model using our estimated parameters (dashed line), versus the ex-ante risk premia calculated with the observed futures (line).

ex-post risk premia. The ex-post risk premia are the sum of the ex-ante risk premia, the forecast error and the ambiguity premium. Therefore, the ex-post risk premia also have seasonal pattern and its variability comes from the forecast errors and the

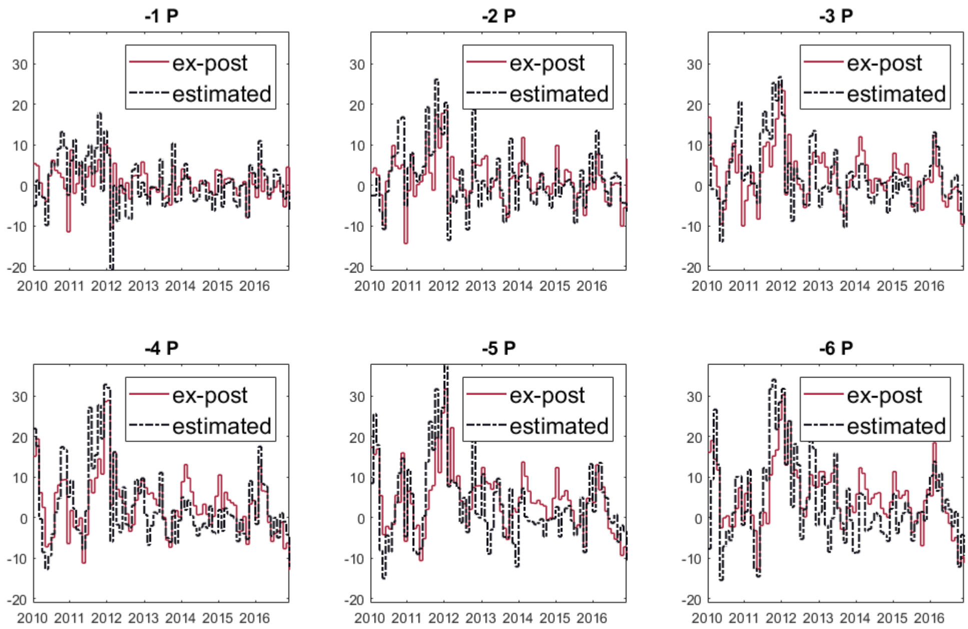


Figure 2.17: Comparison of the ex-post risk premia calculated with our model using our estimated parameters (dashed line), versus the ex-post risk premia calculated with observed peak futures (line).

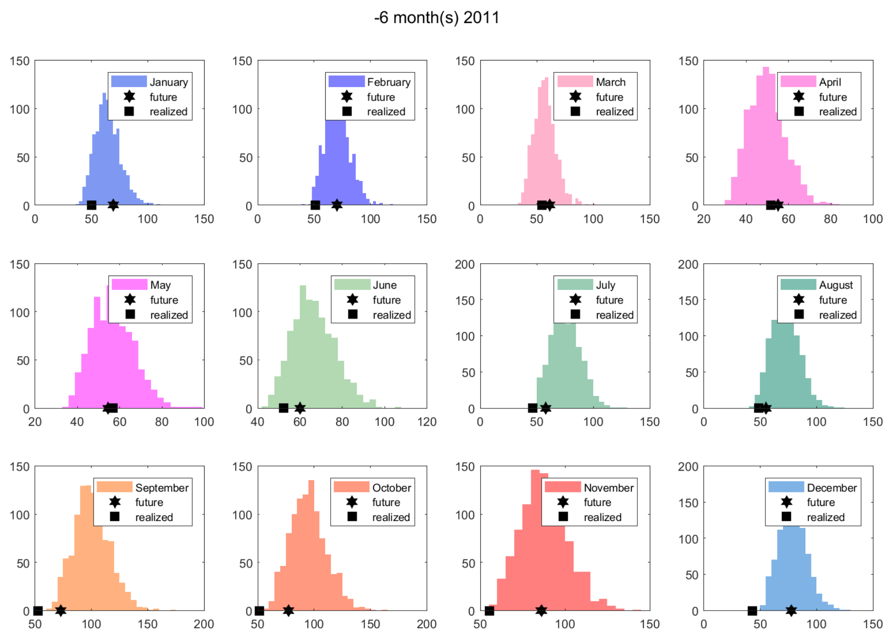


Figure 2.18: Histograms of peak monthly averages for every month in 2011. The samples were calculated with spot prices 6 months before delivery. We compare our empirical distributions with the peak futures (★) and the realized average spot prices (■).

ambiguity premium. For peak futures, our estimations have larger residuals than for base futures, in particular for futures traded 4 to 6 months ahead (see Figure B.3). Also, recall for these time-to-deliveries the ambiguity premium is zero. We can also see the forecast errors for peak futures are higher than for base futures (see Figure 2.7) in the years 2010-2012. Since each group of peak futures contains contracts that deliver along all these years, the overestimation of our model of monthly averages led to large residuals for the identification of peak futures. The overestimation over the years 2010-2012 is displayed in the ex-ante risk premia in Figure 2.16.

2.6 Conclusions

This paper aims at studying the gap between electricity futures and the spot market. We propose to explain futures with three different quantities: a correction premium, a distortion premium, and an ambiguity premium. The primary goal of this paper is the identification of these three different quantities as drivers of futures prices and consequently of the risk premia. Since we aim to explain the mechanism of futures, we specify a model for spot prices exclusively estimated with spot prices information.

The identification of these different premia is applied to different groups of monthly futures. With these groups, we test the differences between base and peak futures, presence of seasonality and dependence on time-to-delivery. For each group, we propose non-parametric distortion densities as step and spline functions.

The shifting parameter from the information premium corrects our empirical distributions and it is closely related to the moments of the distributions: mean and variance. We also relate this correction factor to production costs. Since production costs are cheaper than spot prices (coal prices), the correction premium corrects our monthly expectation for front months. As time-to-delivery increases, the information premium gives more variability and increases the expected average spot prices given with our model.

The distortion premium explains risk preferences with distortion functions. Large prices are given with increasing distortion densities and low prices with decreasing ones. With our estimations, we find that consumers are more risk averse for futures that delivery in the front months, while producers prefer to lock in the prices for further periods. For base futures, the risk aversion of consumers can be explained with the upper average value-at-risk $\mathrm{AV@R}_{0.9}$. The longer the time-to-delivery the less steep the distortion densities are. For these cases, producers trade futures with values closer to the expectations (including the shift corrections). Although for peak futures, the distortion premium also gives larger prices for front months than for further months, we observe producers take more risk for these contracts when delivery lies further in time. For some cases, we recover the lower average-value-at-risk $\mathrm{AV@R}_{0.1}$. For base futures, the ambiguity premium increases with time-to-delivery, the further the delivery period the more uncertain about the model and hence we add a larger premium. The ambiguity also increases for peak futures but only up to 3 months ahead.

The interaction of the shifting parameter and the distortion densities play the main role in the explanation of the ex-ante risk premia. We recover a seasonal pattern, where the ex-ante risk premia are larger in winter than in summer. The magnitude of the ex-ante risk premia for base futures is significantly positive and close to zero in summer. The estimated ex-ante risk premia for peak futures also have seasonal behaviour. The magnitude is larger than for base futures, and we estimate an increasing trend over the years. For peak futures, the ex-ante risk premia are more likely to have large negative values.

Overall, with these three terms, we can explain different risk preferences for producers and consumers behind the traded prices of base and peak futures in the EEX. In general, our estimations capture better the price dynamics for base futures (from 93% to 75% as time-to-delivery increases) than for peak futures (for front months we explain around 80%). Moreover, we can recover persistent characteristics of base and peak futures and their risk premia.

Acknowledgements We are thankful for the funding from Adolf Øens Donasjonsfond, Energizing new computational frontier. Debora Daniela Escobar acknowledges the financial support from the Gaspard-Monge program for Optimization and Operations Research (FMJH/PGMO) with number 2017-0094. Michael Schürle's research is part of the SCCER CREST, which is financially supported by the Swiss Commission for Technology and Innovation (CTI). Florentina Paraschiv and Michael Schürle would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support during the programme "Mathematics of Energy Systems". This work was supported by EPSRC grant no. EP/R014604/1.

CHAPTER 3

Distributionally robust optimization with multiple time scales: valuation of a thermal power plant

Joint work with Wim V. Ackooij, Martin Glanzer and Georg Ch. Pflug

Abstract

The valuation of a real option is preferably done with the inclusion of uncertainties in the model, since the value depends on future costs and revenues, which are not perfectly known today. The usual value of the option is defined as the maximal expected (discounted) profit one may achieve under optimal management of the operation. However, also this approach has its limitations, since quite often the models for costs and revenues are subject to model error. Under a prudent valuation, the possible model error should be incorporated into the calculation. In this paper, we consider the valuation of a power plant under ambiguity of probability models for costs and revenues. The valuation is done by stochastic dynamic programming and on top of it, we use a dynamic ambiguity model for obtaining the prudent minimax valuation. For the valuation of the power plant under model ambiguity we introduce a distance based on the Wasserstein distance. Another highlight of this paper is the multiscale approach, since decision stages are defined on a weekly basis, while the random costs and revenues appear on a much finer scale. The idea of bridging stochastic processes is used to link the weekly decision scale with the finer simulation scale. The applicability of the introduced concepts is broad and not limited to the motivating valuation problem.
Keywords— Model ambiguity · Distributionally robust decision making · Multistage stochastic optimization · Multiscale stochastic optimization · Dynamic programming · Wasserstein distance · Nested distance.

3.1 Introduction

Since the deregulation of the energy market, the question of how to determine the value of a power plant can be asked. The traditional approach of valuing it within a given portfolio of other assets in a coordinated way against one's customer load is one possibility. A second approach is to adopt the ideas of real option pricing in finance. In the first case one ends up with models resembling unit commitment (e.g., [92]) but at a long time scale. Although the actual operation of the power plant can be presented in great detail, it will be harder to incorporate other features in the model. This will typically be the case for uncertainty, where one ends up with multi-stage mixed-integer programs which are not easily solved. One can also argue that it is unreasonable to model the system as fully coordinated. In contrast, when modelling the power plant as a real option, thus operating it in the face of a set of market signals, the setting becomes that of perfect competition. Uncertainty is also naturally modelled, but it comes at the expense of modelling the plant as an independent production unit and thus with less realism in that sense.

However, the price of the real option may well serve as a financial reference base between two parties. For example between the power plant owner and a trading entity actually operating on the market. Taking the option pricing perspective, it must be emphasized that energy markets are by far not as "granular" as the equity markets. For instance, on the electricity market one cannot buy a contract of delivery for a given hour 6 months from now. The classical pricing-hedging duality argument is thus not feasible. Moreover, when operating a power plant, generation will be bound locally by a given power output level. This can be either the result of ramping conditions or minimum up/down times. It is therefore reasonable to try to model the power plant with sufficient realism for the above discrepancies to be minimal. This is the stance that we have taken in the current work. It will lead us to consider a multiscale stochastic program in the sense of [45], i.e., a multistage stochastic optimization problem where each stage itself is subdivided into a given set of time instants.

To account for uncertainty, we start out with a set of typical stochastic models for underlying prices, which are based on multi-factor models (e.g., [17]) driven by Brownian motions. Clearly, such (commonly used) idealized modelling assumptions are rather unrealistic. It is thus the aim and the core part of the present paper to relax such strong assumptions by computing distributionally robust solutions to the studied operational problem and to investigate how the resulting valuation deviates when considering model ambiguity. Distributionally robust optimization is a field which has recently gained a lot of popularity in the literature (see [76, pp. 232–233] for a review of different approaches). In particular, ambiguity sets based on distance concepts between probability measures (such as the Wasserstein distance) are well-supported by theory and frequently applied (e.g., [21, 28, 40, 46, 78]). However, to the best of our knowledge, the effects of distributional robustness in (especially multistage) real-world applications, have not been investigated yet.

In order to solve the formulated problem numerically, the given uncertainty model will be discretized on a scenario lattice. The multiscale structure could then simply mean that uncertainty is lost within a given stage (cf., e.g., [68]). More advanced approaches do consider some uncertainty (e.g., the so-called multi-horizon approach originally suggested in [57] and subsequently studied and applied in [66, 84, 86, 103, 113]), but the resulting paths do not necessarily connect with subsequent elements in the scenario tree/lattice. Hence, the multi-horizon approach is not appropriate for the present problem, as the key requirement of two time scales which may be assumed to run completely independent from each other, is not given. Indeed, we deal with two different granularities associated with one and the same stochastic process reflecting the evolution of the underlying market prices. A framework for such situations, where sub-stage paths in the lattice are carefully connected, has recently been proposed in [45]. We test the multiscale stochastic programming approach suggested in [45] in the context of the present real-world application.

Although the resulting ideas will be illustrated through the power plant real option framework, their potential usage is readily seen to be beyond this specific application. In terms of contributions we can can state:

- For the application of real option pricing, we investigate more reasonable exercise patterns. In order to keep computational burden low, this naturally leads to multiscale stochastic programs. We also consider model ambiguity to mitigate the fairly ideal models for market prices. From a high-level perspective, we thus extend the literature on real-world applications of dealing with two fundamental problems in stochastic programming, namely the problem of time scales with multiple granularities as well as the problem of model ambiguity.

- With respect to multistage model ambiguity, we propose a new concept based on the Wasserstein distance. It is tailored with a computational intention, namely in such a way that (on a discrete scenario tree/lattice) the applicability of a classical backward dynamic programming recursion can be maintained. In particular, the suggested framework leads to solutions that are robust w.r.t. model misspecification in a ball around each conditional transition probability distribution. The size of these balls may be controlled uniformly by a single input parameter. We also link the concept to the nested distance in such a way that it inherits a favourable stability property of the latter.

- In the context of Wasserstein ambiguity sets, we propose a state-dependent metric as a basis for the Wasserstein distance. Thereby we account for more realistic worst-case scenarios. We discuss that the well-appreciated statistical motivation for using Wasserstein balls is not invalidated by doing so.

The paper is organized as follows. Section 3.2 describes the valuation model and the uncertainty model. As typical for real-world energy applications, a sound mathematical framework reflecting all peculiarities of the problem requires carefulness in all details. The underlying uncertainty factors are modelled by a continuous time stochastic process. However, in the light of the nature of the decision problem, we will eventually apply a stochastic dynamic programming algorithm which operates backwards in (discrete) time. To prepare for the computational solution, we therefore discuss all discretization steps required by the multiscale stochastic programming framework that we adopt. Section 3.3 is dedicated to model ambiguity. We introduce and discuss a new concept which is tailor-made for incorporating model ambiguity into dynamic stochastic optimization models on discrete structures. All numerical experiments and aspects of the computational solution algorithm are given in Sect. 3.4. Section 3.5 concludes. Some technical details and examples are deferred to the Appendix C.

3.2 The model

Our valuation problem belongs to the class of discrete time sequential decision problems with finite horizon T, decisions u_t, state variables z_t, and a Markovian driving process ξ_t:

$$\begin{aligned} \max_{\{u_t\}_{t=0}^{T-1}} \mathbb{E} & \left[\sum_{t=0}^{T-1} \beta_t h_t(z_t, u_t, \xi_t) \right] \\ \text{s.t. } & z_{t+1} = g_t(z_t, u_t, \xi_t) \quad \forall t = 0, \dots, T-1 \\ & u_t \in \mathcal{U}_t(z_t) \text{ a.s. } \forall t = 0, \dots, T-1, \;\; u_t = u_t(z_t, \xi_t), \\ & z_t \in \mathcal{Z}_t \text{ a.s. } \forall t = 0, \dots, T-1 \end{aligned} \tag{3.1}$$

Here T is the number of decision stages and $g_t(z_t, u_t, \xi_t)$ is the state transition function. The driving stochastic process ξ_t is assumed to belong to $L_1(\Omega_t, \mathcal{F}_t; \mathbb{R}^m)$ and the feasible decision variables at stage t are defined by the set $\mathcal{U}_t(z_t) \subseteq \mathbb{R}^m$. The set of all reachable state variables is denoted by $\mathcal{Z}_t \subseteq \mathbb{R}^{d_1}$. The stage-wise profit function $h_t : \mathbb{R}^{d_1} \times \mathbb{R}^m \times \mathbb{R}^{d_2} \to \mathbb{R}$ is continuous and satisfies the following growth condition:

$$|h_t(z, u, x)| \leq K \cdot (1 + \|z\| + \|u\| + \|x\|) \,,$$

for all $(z, u, x) \in \mathbb{R}^{d_1+m+d_2}$ and some constant K. We choose the discount factor $\beta_t = \beta^t$ for some constant $\beta \in (0, 1]$ throughout the paper. Any decision u_t to be made at time t may only depend on the current state z_t and the most recent observation of exogenous information ξ_{t-1}. This is the non-anticipativity condition. The initial conditions for the random process ξ and the state vector z are that ξ_0 and z_0 are assumed to be constant.

In our application, the decisions u_t represent the weekly electricity production plan for a thermal power plant. The latter is characterized by many technical constraints, such as minimum up/down times or ramping constraints. Fine grain constraints can be incorporated into the model by increasing the dimension of the state vector and accounting for the number of hours the plant has been offline/online. Such state-representations of constraints on generation assets have received attention in the literature (see, e.g., [36, 37, 67] and the references therein). Finer granularity of the time dimension and/or the state variable would result in a significant increase of time steps T (and reduction of the time step size Δt) as well as an increase in the dimension of z_t. For this reason, we introduce here the idea of a multiscale model: While the production decisions u_t are made on a weekly scale, the production costs and revenues are calculated on a finer time scale. To make the dynamic optimization algorithm tractable, we make the assumption that the decisions, i.e. the production profiles, must be chosen from a pre-specified set with finite cardinality. The profiles are set up such that they reflect realistic operating conditions and key choices, such as generating at minimal stable generation (MSG) at off peak hours.

Just prior to presenting the specific instantiation of (3.1), let us emphasize once more that the idea of subdividing a "stage" to mitigate (the curse of) dimensionality goes largely beyond the presented application. Typical other energy problems with similar mechanisms are cascaded reservoir management problems (e.g., see the extensive discussion in [91] as well as [1, 16, 26, 27, 34, 83, 112]).

3.2.1 Instantiation of the problem: the valuation model

In our instantiation of problem (3.1) the time horizon is spanned by T weeks. Each week $t = 1, ..., T$ is subdivided into S equally sized blocks of hours. With respect to our earlier introduced notation, we now present the following specific versions:

- the price process $\xi_{t,s} = (\xi^e_{t,s}, \xi^f_{t,s}, \xi^c_{t,s}) \in \mathbb{R}^3$ represents the electricity price in GBP per megawatt hour (£/MWh), the fuel price in USD per tonne (\$/tonne(fuel)) and the CO_2 allowances price in EUR per tonne (€/tonne(carbon)), for each block s within week t, for $s = 0, \ldots, S$. The information up to stage t is the information up to $\xi_{t,0}$. The values within the weeks are $\xi_{t,s}$ for $s = 1, \ldots, S$ with the convention that $\xi_{t,S} = \xi_{t+1,0}$ coincides with the initial prices of the next stage. With this convention we ensure continuity of prices in between weeks. In this way, the information up to stage t is to be understood as the information up to the value $\xi_{t,0}$.
- the control $u_t = \{u_{t,s}\}_{s=0}^{S-1} \in \mathcal{U} \subseteq \mathbb{R}^S_+$ represents the production profile vector for week t, where $u_{t,s}$ is given in megawatt (MW) and denotes the production at block s . Before the beginning of intermediate values of week t, we determine u_t . Then, u_t is $\xi_{t,0}-$ measurable.
- the state vector z_t is two-dimensional, i.e., $z_t = (x_t, y_t)$ with
  - $x_t \in \mathbb{R}_+$ representing the amount of CO_2 allowances (measured in tonnes of carbon), that are left for week t .
  - $y_t \in \mathbb{Z}_+$ representing the number of hours the power plant was offline before the beginning of week t.

The objective function $h_t : \mathbb{R} \times \mathbb{Z}_+ \times \mathbb{R}^S_+ \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$h_t\left([x_t, y_t], u_t, \xi_{t,0}\right) = \mathbb{E}\left[\sum_{s=0}^{S-1} f_s(x_t, y_t, u_t, \xi_{t,s}) \middle| \xi_{t,0}\right] . \tag{3.2}$$

The profit at each block s within week t is defined as follows:

$$f_s(x_t, y_t, u_t, \xi_{t,s}) = \begin{cases} (\xi^e_{t,0} - H_1\, \xi^f_{t,0})\, u_{t,0}\, \Delta s - f^{\mathrm{CO}_2}(x_t, \bar{u}_t, \xi^c_{t,0}), & \\ \quad -f^{\mathrm{start}}(0, y_t, u_t, \xi^e_{t,0}, \xi^f_{t,0}) - f^{\mathrm{tr}}(u_t) & \text{if } s = 0 \\ (\xi^e_{t,s} - H_1\, \xi^f_{t,s})\, u_{t,s}\, \Delta s - f^{\mathrm{start}}(s, y_t, u_t, \xi^e_{t,s}, \xi^f_{t,s}) & \text{if } s > 0, \end{cases}$$

where $\bar{u}_t := \sum_{s=0}^{S-1} u_{t,s}$. Costs incurred are based on the following component functions:

- $f^{\mathrm{CO}_2} : \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ gives the cost of buying more CO_2 allowances at the beginning of week t (before the values within week t are known);
- $f^{\mathrm{start}} : \mathbb{Z}^2_+ \times \mathbb{R}^S_+ \times \mathbb{R}^2_+ \to \mathbb{R}_+$ gives the start-up cost if the power plant has been offline prior to (one of its arguments) block s;
- $f^{\mathrm{tr}} : \mathbb{R}^S_+ \to \mathbb{R}_+$ represents fuel transportation costs linked to a selected production profile at the beginning of week t.

Table 3.1 summarizes constants used above or in the sequel.

Notation	Unit	Description
H_1	(£/\$)·(tonne(fuel)/MWh)	$H_2 \cdot H_7$
H_2	£/\$	exchange rate
H_3	£/€	exchange rate
H_4	tonne(carbon)/MWh	$H_7 \cdot J$
H_5	MWh/GJ	
H_6	tonne(fuel)/GJ	$H_5 \cdot H_7$
H_7	tonne(fuel)/MWh	heat rate
J	tonne(carbon)/tonne(fuel)	amount of CO_2 emitted due to fuel burnt

Table 3.1: Description of the constants.

The way in which each state variable is updated will be described now. First we will focus on the variables regarding the CO_2 allowances. Although in the past, a given set of allowances was allocated for free, in principle, they are now obtained from a non modelled auction process. Within our model, the variable I_t will represent the number of additional CO_2 allowances received from the regulator at the beginning of week t (measured in tonnes of carbon). Note that this variable will typically be equal to zero but sometimes it will take a relatively high value. The latter happens exactly at the rare events when new allowances are obtained.

Now, $H_4\, \bar{u}_t\, \Delta s$ is the amount of generated CO_2 during stage t. Hence, together with x_t being the remaining stock level and I_t the "inflows", the amount of allowances one needs to buy at stage t is $\alpha_t = [x_t + I_t - H_4\, \bar{u}_t\, \Delta s]_-$.[1] In the case where α_t is

[1] We use here the definition $[f(x)]_- := -\min\{f(x), 0\}$ for the negative part of a function.

positive, we follow a procurement strategy based on a low/middle/high price range partition resulting from some pre-market analysis. Prices in the interval $[\underline{b}, \overline{b}]$, with $0 \leq \underline{b} < \overline{b}$, are considered middle range. This is formalized as follows:

$$A(x_t, \bar{u}_t, \xi^c_{t,0}) = \alpha_t \cdot \left(1 + C_\alpha \cdot \min\left\{\max\{\overline{b} - \xi^c_{t,0}, 0\}/(\overline{b} - \underline{b}), 1\right\}\right), \qquad (3.3)$$

where C_α is a constant that determines the size of the extra amount to be bought.

Recall, our implicit assumption is that new allowances are always bought before the prices within week t are known. The cost of buying more certificates for week t is then given by

$$f^{\mathrm{CO_2}}(x_t, \bar{u}_t, \xi^c_{t,0}) = A(x_t, \bar{u}_t, \xi^c_{t,0}) \cdot H_3\, \xi^c_{t,0}\,,$$

The amount x_{t+1} of remaining allowances after the previous purchase, is updated as follows:

$$g^{(1)}(x_t, u_t, \xi^c_{t,0}) = A(x_t, \bar{u}_t, \xi^c_{t,0}) + x_t + I_t - H_4\, \bar{u}_t\, \Delta s\,.$$

The second state variable accounts for start-ups and related costs. The latter depend on the amount of time the power plant was offline. In our model this time frame will be partitioned into C different intervals of hours denoted by $(c_j, c_{j+1}]$, (c_C, ∞), for $j = 1, \ldots, C-1$, $c_1 = 0$, over which the start-up costs are assumed to be constant. The associated costs are in terms of power, fuel burnt and extra costs. Depending on y_t and the chosen profile u_t, one can readily figure out in which interval each start-up of u_t falls.

The induced start-up costs at block s within week t are given by:

$$f^{\mathrm{start}}(s, y_t, u_t, \xi^e_{t,s}, \xi^f_{t,s}) = \overline{W}_s(y_t, u_t)\, \xi^e_{t,s} + \overline{B}_s(y_t, u_t)\, H_2\, H_6\, \xi^f_{t,s} + \overline{E}_s(y_t, u_t)\,, \qquad (3.4)$$

where

- $\overline{W}_s(y_t, u_t)$ is the amount of works power (MWh) for a start-up at s;
- $\overline{B}_s(y_t, u_t)$ is the amount of solid fuel burnt (GJ) during a start-up at s;
- $\overline{E}_s(y_t, u_t)$ denotes engineering and imbalance costs (£) during a start-up at s.

The updated state y_{t+1} is given by:

$$g^{(2)}(u_t) = (S - \max\{s : u_{t,s} \neq 0\}) \cdot \Delta s\,,$$

where $\max\{\emptyset\} = 0$.

As a further cost factor, we account for the fuel transportation costs associated to each profile:

$$f^{\mathrm{tr}}(u_t) = C^{tr}\, H_7\, \bar{u}_t\, \Delta s\,, \qquad (3.5)$$

where the unit transportation cost (in £ per tonne of fuel) is given by the constant factor C^{tr}.

3.2.2 Underlying price processes

To model the underlying uncertainties, i.e., the stochastic price-evolution of electricity, fuel and CO_2 allowances, we postulate a version of a classical two-factor model. The latter are commonly used for the modeling of commodity markets (cf. [17, 30, 33, 81]). More specifically, in our model the electricity price behaviour is governed by a long term and a short term factor, whereas fuel and CO_2 allowances prices evolve according to a one factor model. In summary, we get a three-dimensional geometric Brownian motion model driven by four correlated one-dimensional Brownian components $B^{e,\mathrm{sh}}, B^{e,\mathrm{lo}}, B^f, B^c$. In particular, the dynamics of the underlying stochastic process F are described by the SDE

$$\begin{pmatrix} dF^e_{t,t}/F^e_{t,t} \\ dF^f_{t,t}/F^f_{t,t} \\ dF^c_{t,t}/F^c_{t,t} \end{pmatrix} = \begin{pmatrix} \sigma^{\mathrm{e,sh}}_{\mathrm{t}} & 0 & 0 & \sigma^{\mathrm{e,lo}}_{\mathrm{t}} \\ 0 & \sigma^{\mathrm{f}}_{\mathrm{t}} & 0 & 0 \\ 0 & 0 & \sigma^{\mathrm{c}}_{\mathrm{t}} & 0 \end{pmatrix} \begin{pmatrix} dB^{e,\mathrm{sh}}_t \\ dB^f_t \\ dB^c_t \\ dB^{e,\mathrm{lo}}_t \end{pmatrix}, \tag{3.6}$$

where the superscripts sh and lo refer to short-term and long-term, respectively. Volatility is allowed to be time-dependent but deterministic. The double-index notation $F_{t,t}$ expresses the fact that we model the spot price as a special case of the forward price. In particular, the forward price $F_{0,t}$ (as observed in the market at time 0) will enter the solution of (3.6) at time t. In this way, we account for the well-known seasonality (peak-hours and off-peak-hours) inherent in electricity prices. Figure 3.1 visualizes this typical effect. To avoid notational clutter, we henceforth write ξ_t with one index for the spot price as a short hand for $F_{t,t}$. Note that we are dealing with a continuous time stochastic model here. The index notation should not be confused with the discrete-time multiscale indexes used in the valuation model; the context will always make clear what is meant.

Regarding the dependence structure between the underlying assets, we allow for a time dependent correlation matrix

$$\rho_t = \begin{pmatrix} 1 & \varrho^{\mathrm{e^{sh},f}}_{\mathrm{t}} & \varrho^{\mathrm{e^{sh},c}}_{\mathrm{t}} & \varrho^{\mathrm{e^{sh},e^{lo}}}_{\mathrm{t}} \\ \varrho^{\mathrm{e^{sh},f}}_{\mathrm{t}} & 1 & \varrho^{\mathrm{f,c}}_{\mathrm{t}} & \varrho^{\mathrm{e^{lo},f}}_{\mathrm{t}} \\ \varrho^{\mathrm{e^{sh},c}}_{\mathrm{t}} & \varrho^{\mathrm{f,c}}_{\mathrm{t}} & 1 & \varrho^{\mathrm{e^{lo},c}}_{\mathrm{t}} \\ \varrho^{\mathrm{e^{sh},e^{lo}}}_{\mathrm{t}} & \varrho^{\mathrm{e^{lo},f}}_{\mathrm{t}} & \varrho^{\mathrm{e^{lo},c}}_{\mathrm{t}} & 1 \end{pmatrix}.$$

Using the (lower triangular) matrix L_t resulting from a Cholesky decomposition of ρ_t, we may replace the Brownian factors $[dB^{e,\mathrm{sh}}_t, dB^f_t, dB^c_t, dB^{e,\mathrm{lo}}_t]^\top$ by the matrix-vector product $L_t \times [dW^{(1)}_t, dW^{(2)}_t, dW^{(3)}_t, dW^{(4)}_t]^\top$, such that the underlying prices are driven by independent Wiener processes $W^{\mathrm{s}}_1, W_2, W_3, W^{\mathrm{l}}_1$. Multiplying the volatility matrix in (3.6) with L_t, we can write the model in the form

$$\begin{pmatrix} d\xi^e_t/\xi^e_t \\ d\xi^f_t/\xi^f_t \\ d\xi^c_t/\xi^c_t \end{pmatrix} = \begin{pmatrix} a_{11}(t) & a_{12}(t) & a_{13}(t) & a_{14}(t) \\ a_{21}(t) & a_{22}(t) & 0 & 0 \\ a_{31}(t) & a_{32}(t) & a_{33}(t) & 0 \end{pmatrix} \begin{pmatrix} dW^{(1)}_t \\ dW^{(2)}_t \\ dW^{(3)}_t \\ dW^{(4)}_t \end{pmatrix}. \tag{3.7}$$

The non-zero components of the above coefficient matrix involve nasty terms with combinations of the various correlations. The precise parameters can be found in the Appendix C.

The solution of SDEs of such a form as in (3.7) is well known to be of the geometric Brownian motion type (e.g., see [70, pg. 62]). In particular, the random vector $\xi_t = [\xi_t^e, \xi_t^f, \xi_t^c]$ follows a three-dimensional log-normal distribution. The corresponding parameters can again be found in the Appendix C.

3.2.2.1 Discretization & the associated bridge process

For our numerical solution framework, which is discussed in detail in Sect. 3.4, the process ξ will first be discretized in all decision stages. Then, an approximate solution of the problem will be computed by stochastic dynamic programming with a backward recursion. In each decision stage, the algorithm relies on the expected profit/loss associated with any decision to be made for the upcoming observation blocks of the following week. To compute such values, we will exploit the structure of the valuation model, the uncertainty model and the backwards recursion. In particular, we are able to compute the expected profits by an analytical formula.

Let us start with the discretization step. To account for the two different time scales explained in Section 3.2.1 above, namely the weekly decision scale and the much finer intra-week observation scale, we use the notation $\xi_{t,s}$, where $t = 0, \ldots, T$ runs in weeks and $s = 0, \ldots, S$ in hour-blocks of equal size Δs (such that $t + S \cdot \Delta s = t + 1$) . Considering the fact that intra-week data of fuel and CO_2 allowances prices typically show – if even available – a rather stable evolution with low fluctuations, we assume those prices to be constant from Monday to Sunday. On the contrary, the electricity price dimension is truly stochastic even on a fine time-scale. Looking at the expected profit function (3.2) at some block s during a week t, it turns out that (on the basis of our assumptions) the problem boils down to the expected value of the electricity price $\xi_{t,s}^e$ given both the initial value $\xi_{t,0}^e$ as well as the final value $\xi_{t,S}^e$ of week t . This is due to the fact that the function $h_t(\cdot)$ is linear in $\xi_{t,s}^e$. Mathematically speaking, we are left with the computation of the conditional expected value at time $t + s \cdot \Delta s$ of the stochastic bridge process linking the values $\xi_{t,0}^e$ and $\xi_{t,S}^e$, for all $s = 1, \ldots, S$. All other parts can be computed in a straightforward way.

The one-dimensional process $\xi_{t,s}^e$ follows a univariate lognormal distribution. Thus, its transition density δ is available in analytical terms and the transition density of the associated bridge process can be computed explicitly. Let an *initial* value η_1 of the process at the beginning of some week t and a *final* value η_2 at the end of that week be given (i.e., $\xi_{t,0}^e = \eta_1$ and $\xi_{t,S}^e = \eta_2$). Then, the bridge process transition density, at time $s \in [0, S]$, is given by

$$\begin{aligned}\delta\left(x, t + s \cdot \Delta s | \eta_1, t, \eta_2, t + 1\right) &= \frac{\delta\left(\eta_2, t + 1 | x, t + s \cdot \Delta s\right) \cdot \delta(x, t + s \cdot \Delta s | \eta_1, t)}{\delta(\eta_2, t + 1 | \eta_1, t)} \\ &= \frac{1}{\sqrt{2\pi\hat{\sigma}_{s|t}^2 x}} \exp\left(-\frac{\left(\log(x) - \log(\eta_1) - \hat{\mu}_{s|t}\right)^2}{2\hat{\sigma}_{s|t}^2}\right) ,\end{aligned}$$

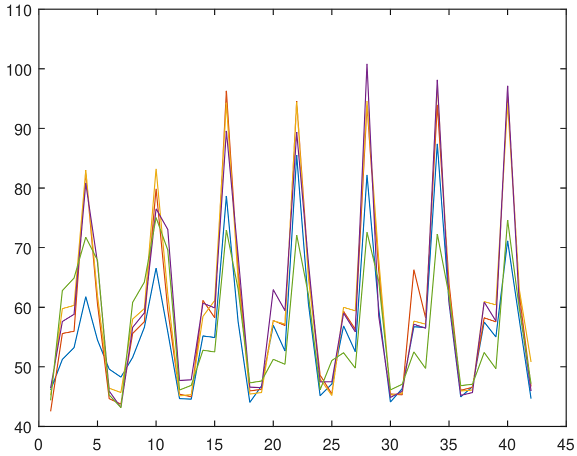


Figure 3.1: Typical structure of electricity forward prices. Each curve represents one week of EFA block power prices (in £/MWh, observed every four hours), from Saturday, 3 am, until Friday, 11 pm. Five weeks of forward price data from the beginning of February to the first week of March, 2017.

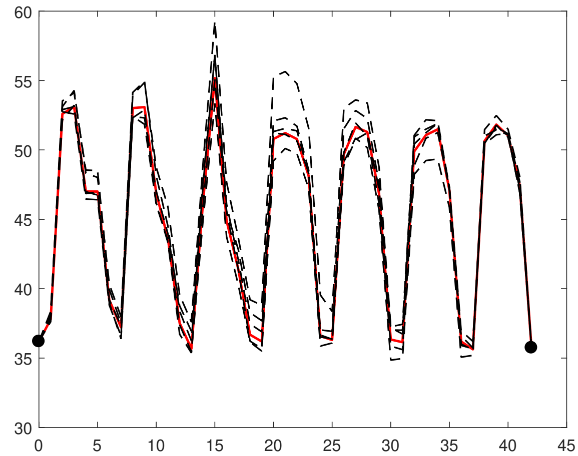


Figure 3.2: Electricity forward price data (solid line) versus 5 simulated trajectories (dashed lines). Simulation based on the bridge process dynamics.

where

$$\hat{\mu}_{s|t} = \frac{\int_t^{t+s\cdot\Delta s} \sigma^2(u)\, du}{\int_t^{t+1} \sigma^2(u)\, du} \log\left(\left(\frac{F^e_{0,t}}{F^e_{0,t+1}}\right)\cdot\frac{\eta_2}{\eta_1}\right),$$

$$\hat{\sigma}_{s|t} = \frac{\left(\int_t^{t+s\cdot\Delta s} \sigma^2(u)\, du\right)\left(\int_{t+s\cdot\Delta s}^{t+1} \sigma^2(u)\, du\right)}{\int_t^{t+1} \sigma^2(u)\, du}.$$

In particular, we get for the conditional expectation

$$\mathbb{E}\left[\xi^e_{t,s} \middle| \xi^e_{t,0} = \eta_1, \xi^e_{t,S} = \eta_2\right] = \eta_1 \cdot \left(\frac{F^e_{0,s\cdot\Delta s}}{F^e_{0,t}} \right) \cdot \exp\left(\hat{\mu}_{s|t}\right) \, . \tag{3.8}$$

Let us emphasize that the above analytical tractability is not due to our restriction of the intra-week stochasticity to one dimension (see [45] for a treatment of the more general multi-dimensional case). This restriction is purely motivated by data.

Figure 3.2 illustrates a set of sample paths from the bridge process, which starts and ends in the forward prices corresponding to two consecutive weeks. The intermediate forward prices are shown for comparison of the seasonal behaviour.

3.3 Ambiguity for dynamic stochastic optimization models

It is an application of classical stochastic dynamic programming theory to solve (3.1) backwards in time on the basis of the following recursion scheme:

$$\begin{aligned} V_t(z_t, \xi_t) = \max_{u_t \in \mathcal{U}_t(z_t)} & h_t(z_t, u_t, \xi_t) + \beta\, \mathbb{E}[V_{t+1}(z_{t+1}, \xi_{t+1}) | \xi_t] \\ & \text{s.t. } z_{t+1} = g_t(z_t, u_t, \xi_{t+1}) \, , \end{aligned} \tag{3.9}$$

where $V_T(z_T, \xi_T) \equiv 0$, z_0 and ξ_0 are given.

Let ξ be a Markovian process defined on a finite state space $\Xi_0 \times \cdots \times \Xi_T$, where on each Ξ_t there is a distance d_t. Let the cardinality of Ξ_t be N_t with $N_0 = 1$ (typically nondecreasing in t). Then the transition matrices P_t, $t = 0, \ldots, T-1$ are of the form $N_t \times N_{t+1}$, where the $i-$th row of the matrix P_t is denoted by $p_t(i)$, for all $i = 1, \ldots, N_t$. Notice that each row $p_t(i)$ describes a probability measure on the metric space (Ξ_{t+1}, d_{t+1}).

Let $\xi^i_t \in \Xi_t$ be given. Then, the conditional probability to transition to $\xi^j_{t+1} \in \Xi_{t+1}$ is given by the j-th element of the row vector $p_t(i)$, denoted as $p_t(i,j)$ for $j = 1, \ldots, N_{t+1}$ and $i = 1, \ldots, N_t$. In this discrete case, the objective of the recursion in (3.9) can be written as

$$V_t(z_t, \xi^i_t) = \max_{u_t \in \mathcal{U}_t(z_t)} h_t(z_t, u_t, \xi^i_t) + \beta \sum_{j=1}^{N_{t+1}} p_t(i,j) \cdot V_{t+1}\left(g_t(z_t, u_t, \xi^j_{t+1}), \xi^j_t\right) \, . \tag{3.10}$$

3.3.1 A new concept: uniform Wasserstein distances

In order to consider model ambiguity, we look for alternative transition matrices Q_t, which are close to a given matrix P_t . Let us first recall the general definition of the

Wasserstein distance for discrete models.

Definition 3.1. *Let $P = \sum_{i=1}^{n} P_i \, \delta_{\xi^i}$ and $Q = \sum_{j=1}^{\tilde{n}} Q_j \, \delta_{\tilde{\xi}^j}$ be two discrete measures sitting on the points $\{\xi^1, \ldots, \xi^n\} \subset \Xi$ and $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^{\tilde{n}}\} \subset \tilde{\Xi}$, respectively. Then, the Wasserstein distance between P and Q is defined as*

$$\mathfrak{W}(P, Q) := \min_{\pi_{ij}} \sum_{i,j} \pi_{ij} \cdot D_{ij} \,,$$

where $\pi = \sum_{i,j} \pi_{i,j} \cdot \delta_{\xi^i, \tilde{\xi}^j}$ is a probability measure on $\Xi \times \tilde{\Xi}$ with marginals P and Q, and where D_{ij} is a distance between the resp. atoms.

We will now measure the closeness between (discrete) multistage models $\mathbb{P}$ and $\mathbb{Q}$ by a uniform Wasserstein distance concept. The rows of alternative matrices Q_t are denoted by $q_t(i)$, $i = 1, \ldots, N_t$. The measure $q_t(i)$ is sitting on at most N_{t+1} points in Ξ_{t+1} and is such that $\mathrm{supp}(q_t(i)) \subseteq \mathrm{supp}(p_t(i))$. Then, we define the distance

$$\mathfrak{W}^{\infty}(\mathbb{P}, \mathbb{Q}) = \max_{0 \le t \le T-1} \max_{1 \le i \le N_t} \mathfrak{W}(p_t(i), q_t(i)) \,, \tag{3.11}$$

which can be interpreted as a uniform version of scenariowise Wasserstein distances. An ε ball around $\mathbb{P}$ is characterized by the fact that all members $\mathbb{Q}$ satisfy

$$\mathfrak{W}(p_t(i), q_t(i)) \le \varepsilon \,,$$

for all $t = 0, \ldots, T-1$ and all $i = 1, \ldots, N_t$.

When introducing ambiguity into the model, we would like to solve problem (3.1) wherein the objective function is replaced with:

$$\max_{\{u_t\}_{t=0}^{T-1}} \min_{\mathbb{Q} \,:\mathfrak{W}^{\infty}(\mathbb{P}, \mathbb{Q}) \le \varepsilon} \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=0}^{T-1} \beta_t h_t(z_t, u_t, \xi_t) \right] ,$$

where $\mathbb{E}_{\mathbb{Q}}$ denotes the expectation with respect to the measure $\mathbb{Q}$.

Our choice of the multistage distance makes it possible to keep the decomposed structure of the backward recursion, which reads:

$$\begin{aligned} V_t(z_t, \xi_t^i) = \max_{u_t} \min_{\mathfrak{W}(p_t(i), q_t(i)) \le \varepsilon} & \; h_t(z_t, u_t, \xi_t^i) + \beta \sum_{j=1}^{N_{t+1}} q_t(i,j) \; V_{t+1}(z_{t+1}, \xi_{t+1}^j) \\ \text{s.t. } & \; z_{t+1} = g_t(z_t, u_t, \xi_{t+1}) \,. \end{aligned} \tag{3.12}$$

Hence, the ambiguity approach just extends the max model to a maximin model. The ambiguous model can also be seen as a risk adverse model in contrast to the basic risk neutral model. If the distance is not of the decomposable form, then the backward recursion does not decompose scenariowise and one has to find all optimal decisions in one very big stagewise but not scenariowise decomposed algorithm. However, decomposability is the key feature of successful methods for dynamic decision problems. Hence, our concept is strongly motivated by its favourable computational properties. However, as we will discuss now, under a mild regularity condition (in the sense that it will always hold for discrete models, which are the basis of the whole computational framework) it can still be shown that optimal solutions are close if the underlying models are close w.r.t. the uniform Wasserstein distance.

The general distance concept for stochastic processes (including their discrete representation in the form of scenario trees) is the nested distance introduced in [74, 75] as a multistage generalization of the classical Wasserstein distance.[2] In our case we have a Markov process which can be seen as a lattice process. Notice that a lattice can be interpreted as a compressed form of a tree. It can always be "unfolded" to a tree representing the same filtration structure, by splitting each node according to the number of incoming arcs. Thus, all results applying for trees do hold for lattices as well. The uniform Wasserstein distance introduced above is given by the maximum Wasserstein distance over all conditional transitions. The subsequent stability result holds.

Proposition 3.1. *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be two discrete Markovian probability models defined on the filtered space $(\Omega, \sigma(\xi))$. Assume the following Lipschitz condition regarding $\mathbb{P}$ to hold for all $t = 0, \ldots, T-1$ and all values ξ_t^i, ξ_t^j, where $i, j = 1, \ldots, N_t$:*

$$\mathfrak{W}\left(P_t(\cdot|\xi_t^i), P_t(\cdot|\xi_t^j)\right) \leq K_t \cdot \|\xi_t^i - \xi_t^j\|,$$

for $K_t \in \mathbb{R}$. Consider the generic multistage stochastic optimization problem

$$v(\mathbb{Q}) := \inf \mathbb{E}^{\mathbb{Q}}[c(x, \xi)],$$

where the (nonanticipative) decisions x lie in some convex set and where the function $c(\cdot,\cdot)$ is convex in x and 1-Lipschitz w.r.t. ξ. Then the relation

$$\left| v(\mathbb{P}) - v(\tilde{\mathbb{P}}) \right| \leq \mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}}) \leq K \cdot \mathfrak{W}^\infty(\mathbb{P}, \tilde{\mathbb{P}})$$

holds, where $\mathsf{dl}(\cdot,\cdot)$ denotes the nested distance, and where the constant K is given by

$$K := \sum_{t=0}^{T-1} \prod_{i=t+1}^{T} (1 + K_j).$$

Proof. The first inequality is a well-known result from [75, Th. 11]. The statement then follows readily from [76, Lem. 4.27], by using $\mathfrak{W}^\infty(\mathbb{P}, \tilde{\mathbb{P}})$ as a uniform bound for $\mathfrak{W}(P_t(\cdot|\xi_t), \tilde{P}_t(\cdot|\xi_t))$, over all t. □

Remark 3.1. *Notice that for discrete Markov chain models the assumption in Prop. 3.1 always holds, as one can simply choose the ergodic coefficient*

$$K_t = \max_{\xi_t^i \neq \xi_t^j} \frac{\mathfrak{W}\left(P_t(\cdot|\xi_t^i), P_t(\cdot|\xi_t^j)\right)}{\|\xi_t^i - \xi_t^j\|}.$$

Remark 3.2. *In the above construction, all models contained in the ambiguity set share exactly the same tree structure and node values. Thus, one might conjecture*

[2]The general definition of the nested distance can be found in the Appendix C.

at a first glance that it would be possible to bound the nested distance by a simple sum of the stagewise maximum of conditional Wasserstein distances, weighted by the number of subtrees at the respective stage. A simple example in the Appendix C shows that such a construction does not work in general.

3.3.2 State-dependent distances

In practice, the worst-case model for an upcoming period may often depend on the current state. In the model considered in the present paper (cf. Sect. 3.2.1), we decide only at the beginning of each stage about the procurement of additional CO_2 allowances. In particular, we restrict ourselves not to buy any if the current stock is sufficient for whatever we may do during the subsequent week; regardless of their market price. If we neglect this consideration when searching for the optimal distributionally robust production profile, the worst case may reflect a variation in the CO_2 allowances price dimension which in fact will not have an impact on our optimal decision. Thus, we modify the distance on the underlying three-dimensional space by projecting to the electricity price and the fuel price dimension only, given that our stock of CO_2 allowances is sufficient. Otherwise, we keep the usual L_1 norm. More formally, we define

$$D\left([\xi_t^e, \xi_t^f, \xi_t^c], [\tilde{\xi}_t^e, \tilde{\xi}_t^f, \tilde{\xi}_t^c]\right) := \begin{cases} w^e|\xi_t^e - \tilde{\xi}_t^e| + w^f|\xi_t^f - \tilde{\xi}_t^f| & \text{if } \alpha_t = 0 \\ w^e|\xi_t^e - \tilde{\xi}_t^e| + w^f|\xi_t^f - \tilde{\xi}_t^f| + w^c|\xi_t^f - \tilde{\xi}_t^f| & \text{if } \alpha_t > 0\,, \end{cases} \tag{3.13}$$

with α_t defined in Sect. 3.2.1 and positive weights w^e, w^f, w^c. Notice that D is not a distance, as it does not separate points. However, this fact does not entail any restrictions for our considerations.

When basing the uncertainty model on historical observations, there is a strong statistical argument for using balls w.r.t. the Wasserstein distance as ambiguity sets (cf. [28]). In particular, large deviations results are available (see [12, 35] for the case of the Wasserstein distance and [46] for the case of the nested distance) which provide probabilistic confidence bounds for the true model being contained in the ambiguity set around the (smoothed) empirical measure. Observe that such results are not invalidated by the state-dependency that we introduce: it is evident that a given confidence bound is directly inherited if one neglects some dimension. Notice however that a general state-dependent weighting of the dimensions would require a more careful treatment.

3.4 A case study

In the following, we will test the framework elaborated in Section 3.2.1 and Section 3.3 for a specific power plant. For the present application, each week t is subdivided in $S = 42$ blocks, where each block has 4 hours. We solve the problem for a quarter ahead, thus we take the horizon to be $T = 13$ weeks.

In this case, the control variables are given vectors of dimension $S = 42$. The set of production profiles we use for this case study consists of 10 different production schedules. This set is denoted by $\mathcal{U} = \{u^{(i)}\}_{i=1}^{10}$ (see Figure 3.3) and it will remain constant for every stage.

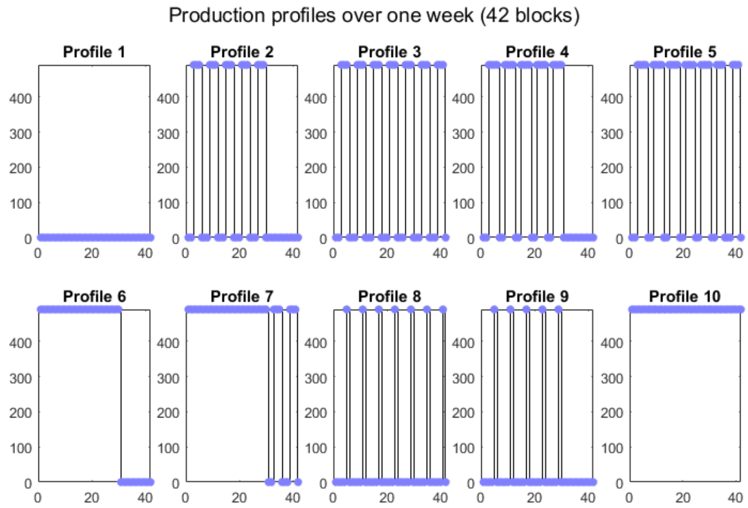


Figure 3.3: All different profiles in $\mathcal{U}$. Production is given in MW.

The discrete evolution of prices is given as a lattice process (ξ) defined on $\Xi_0 \times \cdots \times \Xi_T$, where each space Ξ_t has N_t elements correspondent to the number of nodes, i.e., $\Xi_t = \{\xi_t^1, \ldots, \xi_t^{N_t}\}$ and each node $\xi_t^i = (\xi_t^{e,i}, \xi_t^{f,i}, \xi_t^{c,i}) \in \mathbb{R}^3$ for all $i = 1, \ldots, N_t$ and all $t = 0, \ldots, T$. As explained in Section 3.3, P_t will denote the probability transition matrices from stage t to stage $t+1$ of dimensions $N_t \times N_{t+1}$. The description of the lattice construction will be explained in Section 3.4.1.1.

The profit function h_t at stage t is defined by the expected profit during the upcoming week (see (3.2)). Profits at every block s in the week are quantified by the functions f_s, for $s = 0, \ldots, S-1$. Costs of buying additional allowances and transportation costs are quantified only at the beginning of each week, while start-up costs need to be assigned at each block s. We proceed to the description of the state variables. The costs of buying new allowances depend on the strategy A defined in (3.3). For its computation we consider $[\underline{b}, \overline{b}] = [4.4, 9.6]$, where the latter values were obtained by applying a simple quantile rule to the available data set. Moreover, we set $C_\alpha = 2$.

For the partition of the amount of available CO_2 allowances x_t, we consider different possible values from 0 tonne(carbon) to 10^5 tonne(carbon), and we also take into account the allowances needed for each profile. All in all, the partition of state x_t has 16 different elements.

For every state in the partition, an example of the procurement strategy is shown in Figure 3.4 for different prices. Note that we illustrate this example when we choose full production, i.e., $u^{(10)}$. The strategy when choosing a different profile is similar, the only change is that we do not need to buy as many allowances as with $u^{(10)}$.

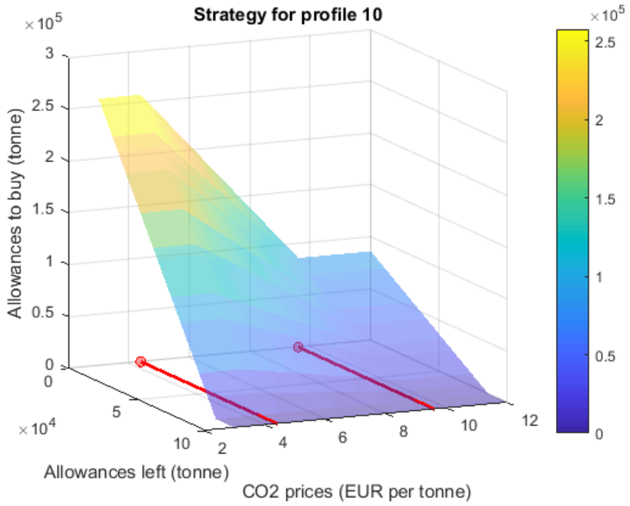


Figure 3.4: If we choose profile $u^{(10)}$, the strategy A to follow is illustrated for different prices $\xi^c_{t,0}$ and all possible left allowances x_t in the partition. The horizontal lines indicate the lower and upper bounds for the prices $\underline{b}$ and $\overline{b}$, respectively.

The second state variable y_t describes the hours the power plant was off since the last time it was on. The costs associated with restarting the production depend on y_t and the chosen profile u_t. The cost function is a step function given in Table 3.2 below.

offline hours	Solid fuel burnt ($\overline{B}$) in GJ	Works power ($\overline{W}$) in MWh	Engineering costs (EC)	Imbalance costs (IC)
$(0, 20]$	412	14	10000	1600
$(20, 36]$	454	20.5	10800	1600
$(36, 48]$	596	30	11800	1600
$(48, 65]$	596	40.5	12800	1600
$(65, \infty)$	629	60.9	13800	1600

Table 3.2: Classification of offline hours and associated start-up costs.

Once a profile is chosen for any week, the first time the profile is different than zeros is where we consider initial start up costs. Then, for the initial start up costs we consider the hours the power plant was offline before week t starts (i.e., y_t) in addition to the hours the chosen profile is off before it starts producing. For the rest of the blocks the costs will only depend on the profile. As for the notation of the elements in f^{start}, see 3.4, $\overline{E}_s$ is the sum of the last two columns of Table 3.2. We illustrate the initial start up costs for a profile that is on in the beginning of the week for all possible values of y_t (see Figure 3.5) .

Given that we have a finite set of profiles, we can calculate the value of y_t for each profile. The partition of y_t will have these values and the limit numbers of the classes in Table 3.2.

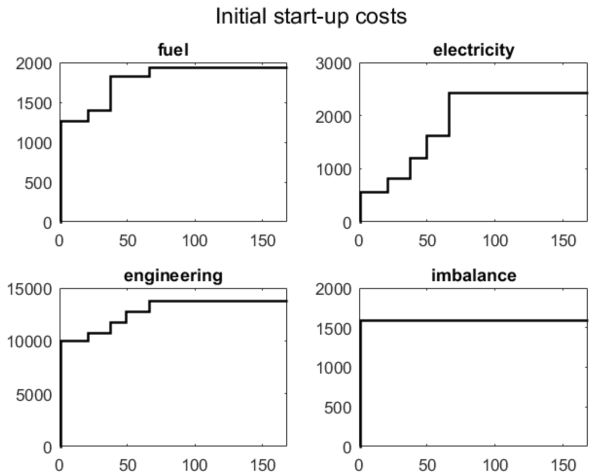


Figure 3.5: Initial start-up costs (£) for all possible initial states $y_t \in \{0, 1, \ldots 168\}$, when the profile is on in the beginning of the week. The costs are calculated for a specific node with electricity price 40 €/MWh and fuel price 90 \$/tonne(fuel).

Regarding the transportation cost function in (3.5), we assume $C^{tr} = 40$ (£/tonne(fuel)).

Finally, the values of the constants in Table 3.1 are $H_2 = 0.78$ (£/\$), $H_3 = 0.9$ (£/€), $H_5 = 0.0975$ (MWh/GJ), $H_7 = 0.45$ (tonne(fuel)/MWh) and $J = 2.31$ (tonne(carbon)/tonne(fuel)).

3.4.1 The solution algorithm

We numerically solve the power plant valuation problem by a stochastic dynamic programming algorithm. A lattice structure is used as a discrete representation of the uncertainty model in all decision stages.

3.4.1.1 On the lattice construction

The state-of-the art approach for the construction of scenario lattices is based on optimal quantization techniques (cf. [6, 61]). For a given number of discretization points, such methods select the optimal locations as well as the associated probabilities in such a way that the Wasserstein distance (or some other distance concept for probability measures) with respect to a (continuous) target distribution is minimized. For the present study, we have implemented a stochastic approximation algorithm for the quantization task, following [76, Algorithm 4.5]. Referring to the latter algorithm, we first applied the iteration step (ii) in order to find the atoms of all marginal distributions, separately for each stage. We then formed a lattice out of these sets of points by fixing the structure of allowed transitions and then applied step (iv) to determine all conditional transition probabilities. Eventually, this also determines the absolute probabilities of each node in the lattice. The Wasserstein distance of order two has been used as a target measure for the minimization. We use a ternary lattice, i.e., each node has (at most) three successors with a positive transition probability.

As for the notation, recall that we distinguish between stages (weeks) and intra-week blocks. Decisions are taken only at each stage but the profit will be calculated taking into account the random evolution of the prices during the entire week. The discretized process in each stage t and node i is denoted as $\xi_t^i = (\xi_t^{e,i}, \xi_t^{f,i}, \xi_t^{c,i})$, and the values of the process within week t starting in node i, are denoted by $\xi_{t,s}^i = (\xi_{t,s}^{e,i}, \xi_{t,s}^{f,i}, \xi_{t,s}^{c,i})$, for $s = 0, \dots S$. At $s = 0$, $\xi_{t,0}^i = \xi_t^i$ takes the value of node i; and at $s = S$, $\xi_{t,S}^i = \xi_{t+1}^j$ takes the value of node j in the next stage with probability $p_t(i,j)$, for $j = 1, \dots, N_{t+1}$. Note that if the lattice structure does not make a link between two nodes in consecutive stages, then the probability of such a transition will be zero.

3.4.2 Computing the expected profit between two decisions

The valuation of the power plant is obtained by solving (3.10) backwards in time from $t = T$ to $t = 0$. In this section, we specify how to solve (3.10) and its robust version (3.12) at any stage t. We specifically concentrate on the calculation of the expected profit within each week given the current node and a successor node.

As discussed in Section 3.2.2.1, electricity prices within weeks will be modeled by the bridge process that we described. Fuel and CO_2 allowances prices are assumed to remain constant between stages.

We start now with the computation of the expected profit, as defined in (3.2). In classical stochastic dynamic programming problems, h_t exclusively depends on the values observed at time t . In contrast, in Sect. 3.2.1 we instantiated (3.1) in such a way that the function h_t is defined as an expected value of the the random profits within week t. Hence, given the values of node i at stage t (at block $s = 0$), as well as initial states (x_t, y_t); the weekly profit h_t will be calculated as follows

$$\begin{aligned} h_t([x_t, y_t], u_t, \xi_{t,0}^i) &= \mathbb{E}\left[\sum_{s=0}^{S-1} f_s(x_t, y_t, u_t, \xi_{t,s}^i) \middle| \xi_t^i = \xi_{t,0}^i\right] \\ &= \sum_{j=1}^{N_{t+1}} \left(\sum_{s=0}^{S-1} \mathbb{E}\left[f_s(x_t, y_t, u_t, \xi_{t,s}^i) \middle| \xi_t^i = \xi_{t,0}^i,\, \xi_{t+1}^j = \xi_{t+1,0}^j \right] \right) \cdot p_t(i,j)\,. \end{aligned}$$

We define

$$h_t^j(x_t, y_t, u_t, \xi_{t,0}^i) = \sum_{s=0}^{S-1} \mathbb{E}\left[f_s(x_t, y_t, u_t, \xi_{t,s}^i) \middle| \xi_t^i = \xi_{t,0}^i,\ \xi_{t+1}^j = \xi_{t+1,0}^j \right],$$

for all $j = 1, \ldots, N_{t+1}$. Then, $h_t([x_t, y_t], u_t, \xi_{t,0}^i) = \sum_{j=1}^{N_{t+1}} h_t^j(x_t, y_t, u_t, \xi_{t,0}^i) \cdot p_t(i,j)$. We compute now h_t^j as follows

$$\begin{aligned} h_t^j(x_t, y_t, u_t, \xi_{t,0}^i) = & \sum_{s=0}^{S-1} (u_{t,s}\,\Delta s - \overline{W}_s(y_t, u_t)) \cdot \mathbb{E}[\xi_{t,s}^{e,i} | \xi_t^e = \xi_{t,0}^{e,i},\ \xi_{t+1}^e = \xi_{t+1,0}^{e,j}] \\ & - \xi_{t,0}^{f,i} \sum_{s=0}^{S-1} (H_1 u_{t,s} \Delta s + H_2\, H_6\, \overline{B}_s(y_t, u_t)) \\ & - A(x_t, \bar{u}_t, \xi_{t,0}^{c,i})\, H_3\, \xi_{t,0}^{c,i} - \sum_{s=0}^{S-1} \overline{E}_s(y_t, u_t) - H_7\, C^{tr}\, \bar{u}_t\, \Delta s\,. \end{aligned}$$

At stage T we set the terminal condition $V_T = 0$. Given an initial state (x_0, y_0), going backwards in time from $t = T-1$ to $t = 0$, we obtain the power plant value at $t = 0$. The latter is calculated with respect to the baseline multistage model $\mathbb{P}$ and will be denoted as $\nu_0(\mathbb{P})$. The policy associated with $\nu_0(\mathbb{P})$ is denoted with $u_{\mathbb{P}}^*$. It can be represented as a probabilistic tree of profiles.

If we incorporate ambiguity in the lattice process using the uniform Wasserstein distance, for all $0 \le t \le T-1$ we solve:

$$\begin{aligned} V_t([x_t, y_t], \xi_t^i) = \max_{u_t} \min_{\mathfrak{W}(p_t(i), q_t(i)) \le \varepsilon} \{ \\ \sum_{j=1}^{N_{t+1}} \left[h_t^j(x_t, y_t, u_t, \xi_{t,0}^i) + \beta V_{t+1}([g^{(1)}(x_t, u_t, \xi_{t,1}^c), g^{(2)}(u_t)], \xi_{t+1}^j) \right] q_t(i,j) \}. \end{aligned}$$

The optimal value is reached at $t = 0$ and it will be denoted as $\nu_0(\mathcal{Q}^\varepsilon)$, where $\mathcal{Q}^\varepsilon$ denotes the ambiguity set defined as

$$\mathcal{Q}^\varepsilon = \{ \mathbb{Q} \,:\, \mathfrak{W}^\infty(\mathbb{P}, \mathbb{Q}) \le \varepsilon \}\,.$$

A worst-case model $\mathbb{Q}^{\varepsilon *}$ is any multistage probability model contained in $\mathcal{Q}^\varepsilon$ such that $\nu_0(\mathcal{Q}^\varepsilon) = \nu_0(\mathbb{Q}^{\varepsilon *})$. More concrete, the optimal value is reached at a saddle point $(u_{\mathbb{Q}^{\varepsilon *}}^*, \mathbb{Q}^{\varepsilon *})$ where $u_{\mathbb{Q}^{\varepsilon *}}^*$ is the policy associated with the worst-case model.

At each node i, the objective function of the minimization problem is linear in $q_t(i)$ under linear constraints. Define

$$c_j = h_t^j(x_t, y_t, u_t, \xi_{t,0}^i) + \beta V_{t+1}([g^{(1)}(x_t, u_t, \xi_{t,1}^c), g^{(2)}(u_t)], \xi_{t+1}^j))\,,$$

for $j = 1, \ldots, N_{t+1}$. Then, the minimization problem can be written as

$$
\begin{aligned}
\min_{q_t(i,j),\,\pi_{k,\,l}} \quad & \sum_{j=1}^{N_{t+1}} c_j \cdot q_t(i,j) \\
\text{s.t.} \quad & \sum_{k=1}^{N_{t+1}} \pi_{k,\,l} = q_t(i,l) \quad \forall l = 1,\dots,N_{t+1} \\
& \sum_{l=1}^{N_{t+1}} \pi_{k,\,l} = p_t(i,k) \quad \forall k = 1,\dots,N_{t+1} \\
& \sum_{k,\,l} D_{k\,l}\, \pi_{k,\,l} \le \varepsilon \\
& \sum_{k,\,l} \pi_{k,\,l} = 1 \\
& \pi_{k,\,l} \ge 0,\, \forall k,\, l = 1,\dots,N_{t+1}\ ,
\end{aligned}
$$

where $D_{k\,l} = D\left([\xi^{e,k}_{t+1},\, \xi^{f,k}_{t+1},\, \xi^{c,k}_{t+1}], [\xi^{e,l}_{t+1},\, \xi^{f,l}_{t+1},\, \xi^{c,l}_{t+1}]\right)$ is the distance between nodes k and l at stage $t+1$, as defined in (3.13) with specific weights $w^e = 1$, $w^f = H_1$ and $w^c = H_3 \cdot H_4$.

3.4.3 Impact of model ambiguity

3.4.3.1 The value of the power plant

We describe the optimal valuation of the power plant when we compute the iterative system of backward equations in (3.10) and (3.12). We assume that the initial state x_0 provides enough allowances to execute any of the profiles and the power plant was not offline before we start, i.e., $y_0 = 0$. Moreover, the terminal condition for both problems is set to be $V_T = 0$. With the baseline model we obtain an expected profit of approximately $\nu_0(\mathbb{P}) = 2.3 \cdot 10^6$ (£). The optimal decision at $t = 0$ is to turn off the power plant by choosing $u^{(1)}$. The valuation of the power plant including ambiguity is obtained for different radii $\varepsilon \in [0, 2]$. The different optimal values $\nu_0(\mathcal{Q}^\varepsilon)$ are illustrated in Figure 3.6 with respect to the ambiguity level. We observe the valuation of the power plant decreases when the ambiguity radius is higher. For $\varepsilon = 2$, the valuation of the power plant decreases to $\nu_0(\mathcal{Q}^\varepsilon) = 6.8 \cdot 10^5$.

In order to get an insight into the change of prices in the ambiguity model, we report the changes of electricity prices for the worst case models $\mathbb{Q}^{\varepsilon *}$ in Table 3.3. Let B_0 be a vector containing the stagewise expectations of electricity prices with respect to the baseline model and let B_ε be the vector containing the expectations with respect to each worst-case model $\mathbb{Q}^{\varepsilon *}$. The percentage of change at each stage t in the prices are denoted with the parameter θ_t, such that $B_\varepsilon(t) = (1 - \theta_t)B_0(t)$. We observe that the largest change in prices is given in stage 12, where electricity prices decrease up to 30% for $\varepsilon = 2$.

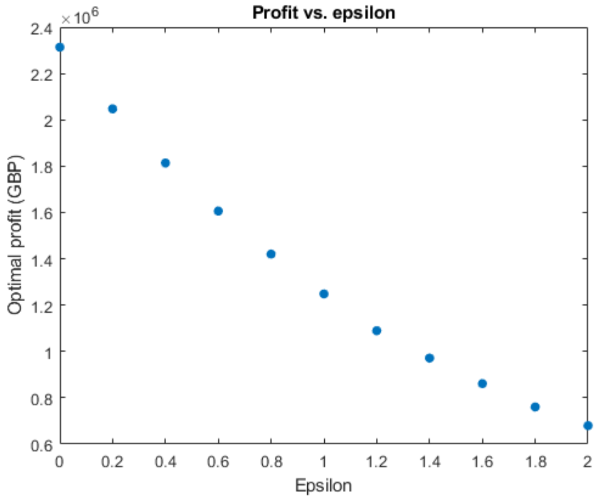


Figure 3.6: The impact of the ambiguity radius ε on the optimal profit over 13 weeks.

ε	Stage-wise change in allowances prices								
	1	2	3	6	7	8	11	12	13
0.4	0.0072	0.0034	0.0094	0.0179	0.0121	0.0109	0.0387	0.0692	0.0442
1	0.0179	0.0183	0.0311	0.0437	0.0410	0.0179	0.0806	0.1633	0.0932
1.4	0.0274	0.0066	0.0316	0.0413	0.0532	0.0263	0.1287	0.2511	0.1121
2	0.0415	0.0173	0.0493	0.0678	0.0892	0.0359	0.1879	0.2991	0.1277

Table 3.3: Percentage of change in electricity prices for every stage and different ambiguity radii.

3.4.3.2 Forward in time

With the iterative solution of the backwards equations we eventually obtain an initial optimal profile at $t = 0$, namely $u_0^* = u^{(1)}$. With this initial decision we go forward in time and create a probabilistic tree $u_{\mathbb{P}}^*$ of the optimal decisions together with their profits. Starting with the given states and the optimal profile at $t = 0$, the updated states at stage $t = 1$ are completely determined by the knowledge of x_0, y_0 and u_0^*. The choice of the optimal profile in $t = 1$, for each node $i = 1, \ldots, N_1$, will be made by looking at the nearest location of the updated states in the grid and taking the correspondent profile chosen in the backwards algorithm. We proceed in this way until we obtain all the optimal profiles at stage $T - 1$. Eventually, we obtain a probabilistic tree with 3^T possible paths. Following the same procedure, we calculate the probabilistic tree of optimal profiles $u_{\mathbb{Q}^{\varepsilon *}}^*$ for each worst-case model $\mathbb{Q}^{\varepsilon *}$, and the corresponding profits under the worst-case models for different radii ε.

Starting with $u^{(1)}$ is optimal for all models at $t = 0$. For the subsequent stages the choices of optimal profiles change. Figure 3.7 shows the stagewise distribution of the optimal profiles chosen with the baseline model $\mathbb{P}$. Figure 3.8 shows the changes of profile choices when we incorporate ambiguity in the model.

With no ambiguity there is a probability greater than 0 to choose full production in stages 6, 7, 8, 10 and 12. When we start increasing the radius of ambiguity these chances drop to 0. The larger the ambiguity radius is the more we choose to be offline or not to produce in the weekends choosing profiles like $u^{(2)}$, $u^{(4)}$, $u^{(6)}$. A different option, but with less probability is not to produce in peak hours, by choosing $u^{(3)}$ or $u^{(5)}$.

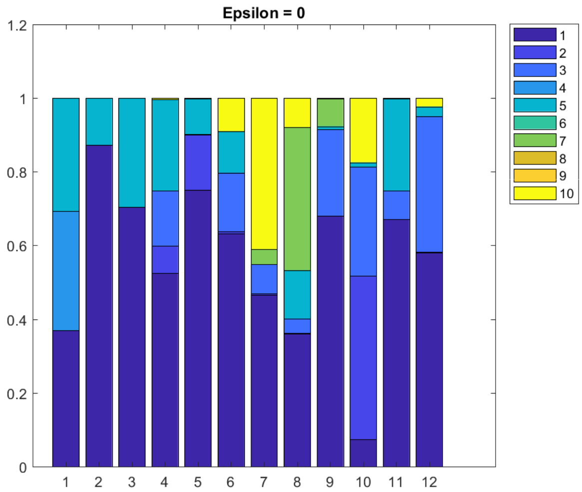
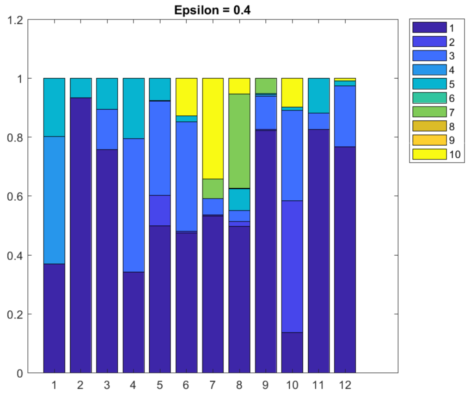


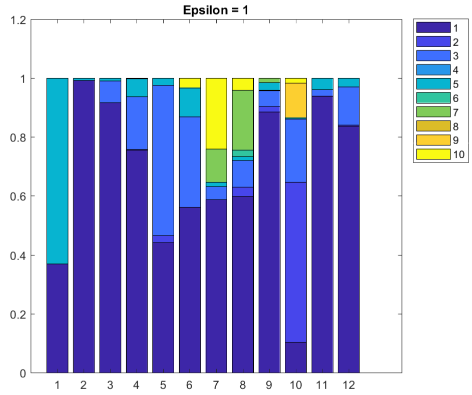
Figure 3.7: Stagewise distribution of the optimal profiles $u^*_{\mathbb{P}}$.

Given the optimal profiles for the baseline model $\mathbb{P}$ and the alternative models $\mathbb{Q}^{\varepsilon*}$ we can calculate the profits we make along the decision tree. To be precise, we denote by $i_\tau \in N_\tau$ any node index at stage $\tau = 0, \ldots, T$. Since $N_0 = 1$, a possible path to follow forward in time up to stage t, will go through any sequence of nodes $(1, i_1, \ldots, i_{t-1}, i_t)$. The profit from i_τ to $i_{\tau+1}$ is the profit made in stage τ and is written as $h_\tau^{i_\tau, i_{\tau+1}}$. This profit is obtained with probability $p_\tau(i_\tau, i_{\tau+1})$. Therefore, the accumulated profit until stage $t-1$ is $(h_0^{0,i_1} + \cdots + h_{t-1}^{i_{t-1},i_t})$ with probability $p_0(1, i_1) \cdots p_{t-1}(i_{t-1}, i_t)$, when we end in node $i_t \in N_t$. Figure 3.9 shows the accumulated profits following the tree of optimal profiles $u^*_{\mathbb{P}}$ as well as the distribution of the final profits.

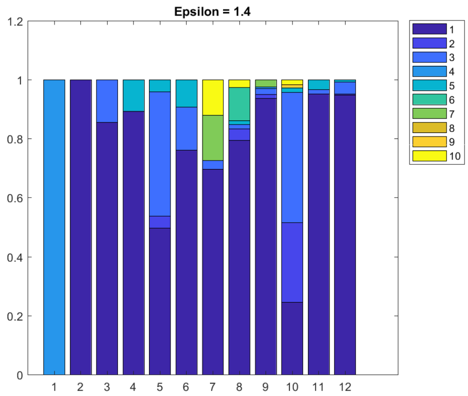
If we include ambiguity, then the optimal profits change as well as their probabilities. Figure 3.10 shows the profit trees together with the final distribution with respect to the correspondent alternative model. We observe that for larger ε the alternative models put more weight at lower profits.



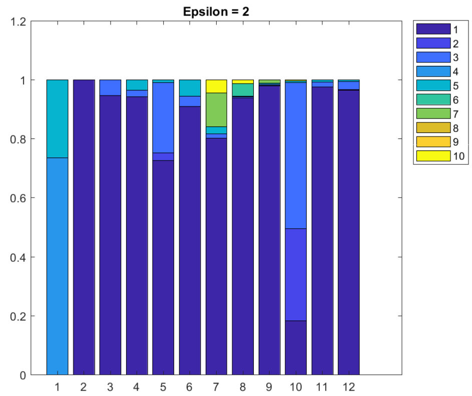
(a) $\varepsilon = 0.4$.



(b) $\varepsilon = 1$.



(c) $\varepsilon = 1.4$.



(d) $\varepsilon = 2$.

Figure 3.8: Stage distribution of the optimal profile with respect the ambiguity radius. For each ε we plot the distribution of $u^*_{\mathbb{Q}^{\varepsilon *}}$.

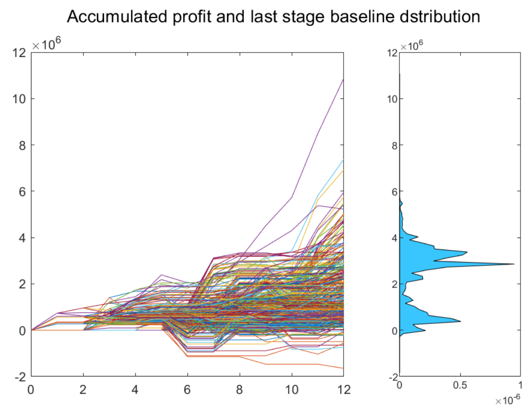
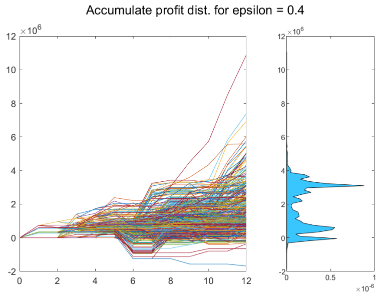
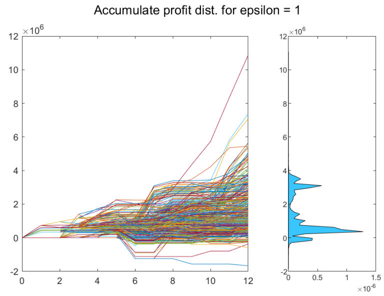


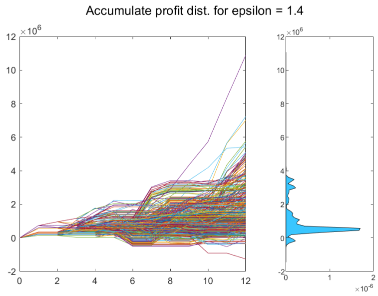
Figure 3.9: On the left the profit tree by following the optimal profiles and on the right the distribution of the final profits obtained following every path.



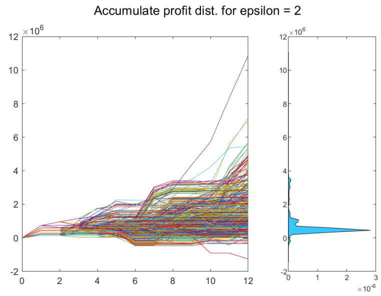
(a) $\varepsilon = 0.4$.



(b) $\varepsilon = 1$.



(c) $\varepsilon = 1.4$.



(d) $\varepsilon = 2$.

Figure 3.10: Stagewise accumulated profit distribution for each solution $u^*_{\mathbb{Q}^{\varepsilon*}}$.

3.5 Conclusions

In this paper, we have shown how a realistic valuation of a power plant can be done by solving a multistage Markovian decision problem. The value is defined as the (discounted) expected net profit, that one can get from the operation of the plant, if an optimal production plan is implemented. In this valuation process, all relevant purchasing costs and selling prices are included in the model. The number of feasible production plans is finite and thus a discrete multistage optimization problem has to be solved. We use the classical backward algorithm for the Markovian control problem and a forward algorithm for determining an estimate of the achievable profit and its distribution. The novelty of the paper is twofold. First, we adopt a multiscale approach, where decisions are made on a coarser scale than costs are calculated. This allows us to keep the computational effort tractable. Second, we do not only consider the baseline model for the random factors, but rather a set of models (the ambiguity set) which are close to the baseline model. This allows to incorporate the fact that probability distributions for future costs and revenues are not known precisely. The more models, and especially the more unfavourable models are included in the ambiguity set, the smaller is the robust value of the plant. We demonstrate how the final value under model ambiguity depends on the degree of uncertainty about the correct price and cost model. Our distance model for the ambiguity set depends on the state of the system, taking into account that what is close for two price vectors depends also on the fact whether these prices are relevant for the state at hand. We also noticed that the optimal production strategy not only depends on the degree of ambiguity, but also gets more diversified for larger ambiguity, in contrast to some bang-bang solutions in unambiguous models.

Acknowledgements The authors would like to acknowledge financial support from the Gaspard-Monge program for Optimization and Operations Research (FMJH/
PGMO) project with number 2017-0094H and title "Incorporating model error in the management of electricity portfolios".

Conclusions

In this thesis, robust pricing is analysed in three different contexts. The first chapter studies the theoretical aspects and identification of the distortion premium for insurance pricing. The second chapter focuses on understanding the price mechanism of electricity futures in terms of distortion functions and ambiguity premia. Lastly, the central topic of the third chapter is the incorporation of model ambiguity for the valuation of a real option.

From the theoretical aspects, we conclude there is a natural metric to use for the calculation of robust distortion premia. This is the Wasserstein distance. Our continuity results of the distortion premium w.r.t. the choice of distributions, provide upper bounds for the robust distortion premium. Under similar assumptions, we characterize the worst case solutions that reach these upper bounds. For particular cases, we are also able to approximate the worst case distributions. Additionally, we characterize the situations under which the robust distortion premia are unbounded. By varying the underlying distance and the order of the Wasserstein distance, we can also find bounds for the robust distortion premia for more cases. We conclude the ambiguity premium does not depend on the initial distribution but only on the radii of the ambiguity sets and the distortion functions. In this paper, we also present an empirical procedure to estimate distortion functions for a group of contracts. We successfully recover well-known examples as the average value-at-risk and the power distortion premium with non-parametric functions. Finally, we discuss possible generalizations of the average value-at-risk.

The second Chapter is devoted to monthly electricity futures, its price formation and their risk premia. We propose to explain the dynamics of futures in terms of the distortion premium, a correction factor and an ambiguity premium. For this purpose, we recover distortion functions, shifting parameters and ambiguity radii from observed futures prices. We identify these parameters for groups of base and peak futures. We test season dependence and the influence of time-to-delivery. For base futures, we can find increasing ambiguity premia as futures are traded further from the delivery period. The estimated shifting parameters reduce the expected monthly averages with our model for futures that deliver in the next months. On the contrary, for futures that deliver in further months, the shifting parameter increases our expectations. The distortion densities capture different risk preferences according to the time-to-delivery. In general, we observe that consumers are more risk averse for short time-to-deliveries than producers, while producers are more concerned about fixing a price to deliver electricity in further months than consumers. The interaction of the shifting parameter and the distortion functions exhibit a seasonal pattern in the ex-ante risk premia. They are larger in winter than in summer and we also recover an increasing trend for the ex-ante risk premia for peak futures. The ex-post risk premia also have a seasonal pattern and are larger when time-to-delivery increases. Our estimations start from initial distributions of the monthly averages at delivery. They are given with a regime-switching model for spot prices that takes into account their main features. Our conclusions are derived from futures prices traded in the EEX over 2010-2017.

To enclose the study of robust pricing in different contexts, the last Chapter evaluates a real option under model ambiguity. We evaluate an overall profit of a thermal power plant which decides weekly on its production. Different prices of electricity, fuel and CO2 certificates influence the profits and decisions taken. Firstly, we discretize and quantize the uncertainties in a lattice process. We also propose a bridge process that makes possible the incorporation of change in prices within weeks. Secondly, we propose a distance between lattice processes to define ambiguity sets. We evaluate this multistage decision problem (including a multiscale model) backwards in time. Our distance is consistent with the backwards algorithm and the robust valuation reduces to stage-wise optimization. Although the ambiguity radius is the same for all stages, we define a state-dependent metric as an underlying metric of the Wasserstein distance. Our solution methods calculate an optimal decision tree of production and its respective profit tree. The more uncertain we are on the prices the less the overall profit. We also observe stage-wise distributions of optimal productions under different levels of ambiguity. Increasing the ambiguity radius, decreases the optimal robust production levels.

References

[1] E. K. Aasgård, G. S. Andersen, S. Fleten, and D. Haugstvedt. Evaluating a stochastic-programming-based bidding model for a multireservoir system. *IEEE Transactions on Power Systems*, 29(4):1748–1757, July 2014. ISSN 0885-8950. doi: 10.1109/TPWRS.2014.2298311.

[2] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.

[3] P. Albrecht. Premium calculation without arbitrage? a note on a contribution by g. venter. *ASTIN Bulletin: The Journal of the IAA*, 22(2):247–254, 1992.

[4] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[5] A. Balbás, J. Garrido, and S. Mayoral. Properties of distortion risk measures. *Methodology and Computing in Applied Probability*, 11(3):385, 2009.

[6] V. Bally and G. Pagès. A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6):1003–1049, 12 2003. doi: 10.3150/bj/1072215199. URL `https://doi.org/10.3150/bj/1072215199`.

[7] F. E. Benth and T. Meyer-Brandis. The information premium for non-storable commodities. *Journal of Energy Markets*, 2(3):111–140, 2009. doi: 10.21314/JEM.2009.021.

[8] F. E. Benth and C. Sgarra. The risk premium and the esscher transform in power markets. *Stochastic Analysis and Applications*, 30(1):20–43, 2012.

[9] F. E. Benth, Á. Cartea, and R. Kiesel. Pricing forward contracts in power markets by the certainty equivalence principle: explaining the sign of the market risk premium. *Journal of Banking & Finance*, 32(10):2006–2021, 2008.

[10] H. Bessembinder and M. L. Lemmon. Equilibrium pricing and optimal hedging in electricity forward markets. *the Journal of Finance*, 57(3):1347–1382, 2002.

[11] L. Bloechlinger, K. Frauendorfer, and F. Trojani. *Power prices: a regime-switching spot/forward price model with Kim filter estimation.* PhD thesis, Citeseer, 2008.

[12] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2007. ISSN 0178-8051.

[13] K. Borch. The utility concept applied to the theory of insurance. *ASTIN Bulletin: The Journal of the IAA*, 1(5):245–255, 1961.

[14] A. Bublitz, D. Keles, and W. Fichtner. An analysis of the decline of electricity spot prices in europe: Who is to blame? *Energy Policy*, 107:323–336, 2017.

[15] D. W. Bunn and N. Karakatsani. Forecasting electricity prices. *London Business School*, 1, 2003.

[16] C. Cervellera, V. Chen, and A. Wen. Optimization of a large-scale water reservoir network by stochastic dynamic programming with efficient state space discretization. *European Journal of Operational Research*, 171(3):1139–1151, 2006.

[17] L. Clelow and C. Strickland. *Energy Derivatives: Pricing and Risk Management*. Lacima Group, 2000.

[18] D. Denneberg. Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin: The Journal of the IAA*, 20 (2):181–190, 1990.

[19] P. Diko, S. Lawford, and V. Limpens. Risk premia in electricity forward prices. *Studies in Nonlinear Dynamics & Econometrics*, 10(3), 2006.

[20] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970.

[21] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu. Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric. *IEEE Transactions on Power Systems*, 33(5):4924 – 4936, 02 2018.

[22] P. Embrechts, C. Klüppelberg, and T. Mikosch. Modelling extremal events. 1997.

[23] A. Energiebilanzen. Bruttostromerzeugung in deutschland ab 1990 nach energieträgern. *Accessed on 20/08/2018 from URL*. URL `http://www.ag-energiebilanzen.de/index.php`.

[24] D. Erni. *Day-ahead electricity spot prices: fundamental modelling and the role of expected wind electricity infeed at the European energy exchange*. 2012.

[25] D. D. Escobar and G. C. Pflug. The distortion principle for insurance pricing: properties, identification and robustness. *Annals of Operations Research*, Dec 2018. ISSN 1572-9338. doi: 10.1007/s10479-018-3119-1. URL `https://doi.org/10.1007/s10479-018-3119-1`.

[26] L. F. Escudero, J. L. de la Fuente, C. Garcia, and F. J. Prieto. Hydropower generation management under uncertainty via scenario analysis and parallel computation. *IEEE Transactions on Power Systems*, 11(2):683–689, May 1996. ISSN 0885-8950. doi: 10.1109/59.496139.

[27] L. F. Escudero, F. J. Quintana, and J. Salmeron. CORO, a modeling and an algorithmic framework for oil supply, transformation and distribution optimization under uncertainty. *European Journal of Operational Research*, 114(3): 638–656, 1999. URL `https://EconPapers.repec.org/RePEc:eee:ejores:v:114:y:1999:i:3:p:638-656`.

[28] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.

[29] M. Everts, C. Huber, and E. Blume-Werry. Politics vs markets: how german power prices hit the floor. *The Journal of World Energy Law & Business*, 9 (2):116–123, 2016.

[30] C.-O. Ewald, A. Zhang, and Z. Zong. On the calibration of the Schwartz two-factor model to WTI crude oil options and the extended Kalman filter. *Annals of Operations Research*, Jan 2018. ISSN 1572-9338. doi: 10.1007/s10479-018-2770-x. URL `https://doi.org/10.1007/s10479-018-2770-x`.

[31] A. Eydeland and K. Wolyniec. *Energy and power risk management: New developments in modeling, pricing, and hedging*, volume 206. John Wiley & Sons, 2003.

[32] E. F. Fama and K. R. French. Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of business*, pages 55–73, 1987.

[33] W. Farkas, E. Gourier, R. Huitema, and C. Necula. A two-factor cointegrated commodity price model with an application to spread option pricing. *Journal of Banking & Finance*, 77(C):249–268, 2017. doi: 10.1016/j.jbankfin.2016.1. URL `https://ideas.repec.org/a/eee/jbfina/v77y2017icp249-268.html`.

[34] S.-E. Fleten, D. Haugstvedt, J. A. Steinsbø, M. Belsnes, and F. Fleischmann. Bidding hydropower generation: Integrating short- and long-term scheduling. MPRA Paper 44450, University Library of Munich, Germany, 2011. URL `https://ideas.repec.org/p/pra/mprapa/44450.html`.

[35] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707, 2015. URL `https://hal.archives-ouvertes.fr/hal-00915365`.

[36] A. Frangioni and C. Gentile. Solving non-linear single-unit commitment problems with ramping constraints. *Operations Research*, 54(4):767–775, 2006.

[37] A. Frangioni, C. Gentile, and F. Lacalandra. Solving Unit Commitment Problems with General Ramp Contraints. *International Journal of Electrical Power and Energy Systems*, 30:316–326, 2008.

[38] E. Furman and R. Zitikis. Weighted premium calculation principles. *Insurance: Mathematics and Economics*, 42(1):459–465, 2008.

[39] E. Furman, R. Wang, and R. Zitikis. Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking & Finance*, 83:70–84, 2017.

[40] R. Gao and A. J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

[41] H. Geman. *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy.* John Wiley & Sons, 2009.

[42] U. Genschel and W. Q. Meeker. A comparison of maximum likelihood and median-rank regression for weibull estimation. *Quality Engineering*, 22(4): 236–255, 2010.

[43] R. Gibson and E. S. Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *The Journal of Finance*, 45(3):959–976, 1990.

[44] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of mathematical economics*, 18(2):141–153, 1989.

[45] M. Glanzer and G. Ch. Pflug. Multiscale stochastic programming. *Available on Optimization Online*, 2019. URL `http://www.optimization-online.org/DB_HTML/2019/03/7098.html`.

[46] M. Glanzer, G. Ch. Pflug, and A. Pichler. Incorporating statistical model error into the calculation of acceptability prices of contingent claims. *Mathematical Programming*, 174(1):499–524, 2019. ISSN 1436-4646. doi: 10.1007/s10107-018-1352-7. URL `https://doi.org/10.1007/s10107-018-1352-7`.

[47] M. J. Goovaerts, R. Kaas, R. J. Laeven, and Q. Tang. A comonotonic image of independence for additive risk measures. *Insurance: Mathematics and Economics*, 35(3):581–594, 2004.

[48] C. Gourieroux and W. Liu. Sensitivity Analysis of Distortion Risk Measures. Technical report, 2006.

[49] F. Greselin and R. Zitikis. From the classical gini index of income inequality to a new zenga-type relative measure of risk: A modellers perspective. *Econometrics*, 6(1):1–20, 2018. URL `https://EconPapers.repec.org/RePEc:gam:jecnmx:v:6:y:2018:i:1:p:4-:d:128699`.

[50] M. Grewal and A. Andrews. Kalman filtering: Theory and practice with matlab 4 th edition. 2015.

[51] L. Han and M. Neumann. Effect of dimensionality on the nelder–mead simplex method. *Optimization Methods and Software*, 21(1):1–16, 2006.

[52] L. Hirth. What caused the drop in european electricity prices? a factor decomposition analysis. *The Energy Journal*, Volume 39(Number 1), 2018. URL `https://EconPapers.repec.org/RePEc:aen:journl:ej39-1-hirth`.

[53] P. J. Huber. *Robust Statistics*, pages 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_594. URL `https://doi.org/10.1007/978-3-642-04898-2_594`.

[54] R. Huisman and C. De Jong. Option pricing for power prices with spikes. *Energy power risk management*, 7(11):12–16, 2003.

[55] N. L. Johnson, S. Kotz, and N. Balakrishnan. Continuous univariate distributions. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 159(2):343, 1996.

[56] E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the fatou property. In *Advances in mathematical economics*, pages 49–71. Springer, 2006.

[57] M. Kaut, K. T. Midthun, A. S. Werner, A. Tomasgard, L. Hellemo, and M. Fodstad. Multi-horizon stochastic programming. *Computational Management Science*, 11(1-2):179–193, 2014.

[58] D. 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, 2012.

[59] R. Kiesel, R. Rühlicke, G. Stahl, and J. Zheng. The wasserstein metric and robustness in risk management. *Risks*, 4(3):32, 2016.

[60] S. Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer, 2001.

[61] N. Löhndorf and D. Wozabal. Gas storage valuation in incomplete markets. *Available on Optimization Online*, 2019. URL `http://www.optimization-online.org/DB_HTML/2017/02/5863.html`.

[62] F. A. Longstaff and A. W. Wang. Electricity forward prices: a high-frequency empirical analysis. *The journal of finance*, 59(4):1877–1900, 2004.

[63] C. Luan. Insurance premium calculations with anticipated utility theory. *ASTIN Bulletin: The Journal of the IAA*, 31(1):23–35, 2001.

[64] J. J. Lucia and E. S. Schwartz. Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of derivatives research*, 5(1): 5–50, 2002.

[65] J. J. Lucia and H. Torró. On the risk premium in nordic electricity futures prices. *International Review of Economics & Finance*, 20(4):750–763, 2011.

[66] F. Maggioni, E. Allevi, and A. Tomasgard. Bounds in multi-horizon stochastic programs. *Annals of Operations Research*, pages 1–21, 2019.

[67] M. G. Martinez, A. L. Diniz, and C. Sagastizábal. A comparative study of two forward dynamic programming techniques for solving local thermal unit commitment problems. In *16th PSCC Conference, Glasgow Scotland*, pages 1–8, 2008.

[68] V. Moriggia, M. Kopa, and S. Vitali. Pension fund management with hedging derivatives, stochastic dominance and nodal contamination. *Omega*, 2018. ISSN 0305-0483. doi: https://doi.org/10.1016/j.omega.2018.08.011. URL `http://www.sciencedirect.com/science/article/pii/S0305048318300471`.

[69] H. T. Nguyen, U. H. Pham, and H. D. Tran. On some claims related to choquet integral risk measures. *Annals of Operations Research*, 195(1):5–31, 2012.

[70] B. Oksendal. *Stochastic Differential Equations*. Springer, 5th edition, 2000.

[71] F. Paraschiv, D. Erni, and R. Pietsch. The impact of renewable energies on eex day-ahead electricity prices. *Energy Policy*, 73:196–210, 2014.

[72] F. Paraschiv, S.-E. Fleten, and M. Schürle. A spot-forward model for electricity prices with regime shifts. *Energy Economics*, 47:142–153, 2015.

[73] G. Ch. Pflug. Subdifferential representations of risk measures. *Mathematical programming*, 108(2-3):339–354, 2006.

[74] G. Ch. Pflug. Version-independence and nested distributions in multistage stochastic optimization. *SIAM Journal on Optimization*, 20(3):1406–1420, 2010. doi: 10.1137/080718401.

[75] G. Ch. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012. doi: 10.1137/110825054.

[76] G. Ch. Pflug and A. Pichler. *Multistage stochastic optimization*. Springer, 2014.

[77] G. Ch. Pflug and W. Romisch. *Modeling, measuring and managing risk*. World Scientific, 2007.

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

[79] A. Pichler. *Distance of probability measures and respective continuity properties of acceptability functionals*. PhD thesis, uniwien, 2010.

[80] A. Pichler. The natural banach space for version independent risk measures. *Insurance: Mathematics and Economics*, 53(2):405–415, 2013.

[81] D. R. Ribeiro and S. D. Hodges. A two-factor model for commodity prices and futures valuation. In *EFMA 2004 Basel Meetings Paper*, 2004.

[82] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10(1):51–74, 2006.

[83] S. Séguin, S.-E. Fleten, P. Côté, A. Pichler, and C. Audet. Stochastic short-term hydropower planning with inflow scenario trees. *European Journal of Operational Research*, 259(1):1156–1168, 2017.

[84] P. Seljom and A. Tomasgard. The impact of policy actions and future energy prices on the cost-optimal development of the energy system in Norway and Sweden. *Energy Policy*, 106(C):85–102, 2017. doi: 10.1016/j.enpol.2017.04.0. URL `https://ideas.repec.org/a/eee/enepol/v106y2017icp85-102.html`.

[85] H. A. Shawky, A. Marathe, and C. L. Barrett. A first look at the empirical relation between spot and futures electricity prices in the united states. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 23(10): 931–955, 2003.

[86] C. Skar, G. Doorman, G. A. Pérez-Valdés, and A. Tomasgard. A multi-horizon stochastic programming model for the European power system. Censes working paper 2/2016, NTNU Trondheim, 2016. ISBN: 978-82-93198-13-0.

[87] S. Stoft. Power system economics. *Journal of Energy Literature*, 8:94–99, 2002.

[88] H. Tsukahara. Estimation of distortion risk measures. *Journal of Financial Econometrics*, 12(1):213–235, 2013.

[89] S. Vallender. Calculation of the wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18(4):784–786, 1974.

[90] W. van Ackooij, D. D. Escobar, M. Glanzer, and G. Ch. Pflug. Distributionally robust optimization with multiple time scales: valuation of a thermal power plant. *Available on Optimization Online*. URL `http://www.optimization-online.org/DB_HTML/2019/04/7157.html`.

[91] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. Joint chance constrained programming for hydro reservoir management. *Optimization and Engineering*, 15:509–531, 2014.

[92] W. van Ackooij, I. Danti Lopez, A. Frangioni, F. Lacalandra, and M. Tahanan. Large-scale unit commitment under uncertainty: an updated literature survey. *Annals of Operations Research*, 271(1):11–85, 2018.

[93] G. G. Venter. Premium calculation implications of reinsurance without arbitrage. *ASTIN Bulletin: The Journal of the IAA*, 21(2):223–230, 1991.

[94] A. E. D. Veraart and L. A. M. Veraart. Risk premia in energy markets. CREATES Research Papers 2013-02, Department of Economics and Business Economics, Aarhus University, Jan. 2013. URL `https://ideas.repec.org/p/aah/create/2013-02.html`.

[95] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[96] A. Vinel and P. A. Krokhmal. Certainty equivalent measures of risk. *Annals of Operations Research*, 249(1-2):75–95, 2017.

[97] J. Von Neumann and O. Morgenstern. Theory of games and economic behavior, 2nd rev. 1947.

[98] J. Von Neumann and O. Morgenstern. *Theory of games and economic behavior (commemorative edition)*. Princeton university press, 2007.

[99] S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54, 1995.

[100] S. Wang. Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA*, 26(1):71–92, 1996.

[101] S. S. Wang. A class of distortion operators for pricing financial and insurance risks. *Journal of risk and insurance*, 67(1):15–36, 2000.

[102] S. S. Wang, V. R. Young, and H. H. Panjer. Axiomatic characterization of insurance prices. *Insurance: Mathematics and economics*, 21(2):173–183, 1997.

[103] A. S. Werner, A. Pichler, K. T. Midthun, L. Hellemo, and A. Tomasgard. Risk measures in multi-horizon scenario trees. In *Handbook of risk management in energy production and trading*, pages 177–201. Springer, 2013.

[104] R. Weron. *Modeling and forecasting electricity loads and prices: A statistical approach*, volume 403. John Wiley & Sons, 2007.

[105] R. Weron, M. Bierbrauer, and S. Trück. Modeling electricity prices: jump diffusion and regime switching. *Physica A: Statistical Mechanics and its Applications*, 336(1-2):39–48, 2004.

[106] J. L. Wirch and M. R. Hardy. Distortion risk measures: Coherence and stochastic dominance. In *International Congress on Insurance: Mathematics and Economics*, pages 15–17, 2001.

[107] D. Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47, 2012.

[108] D. Wozabal. Robustifying convex risk measures for linear portfolios: A non-parametric approach. *Operations Research*, 62(6):1302–1315, 2014.

[109] M. E. Yaari. The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, pages 95–115, 1987.

[110] A. A. Yavuz. Estimation of the shape parameter of the weibull distribution using linear regression methods: Non-censored samples. *Quality and Reliability Engineering International*, 29(8):1207–1219, 2013.

[111] V. R. Young. Premium principles. *Wiley StatsRef: Statistics Reference Online*, 2014.

[112] L. Zéphyr, P. Lang, and B. F. Lamond. Controlled approximation of the value function in stochastic dynamic programming for multi-reservoir systems.

Computational Management Science, 12(4):539–557, Oct 2015. ISSN 1619-6988. doi: 10.1007/s10287-015-0242-1. URL https://doi.org/10.1007/s10287-015-0242-1.

[113] S. Zhonghua, R. Egging, D. Huppmann, and A. Tomasgard. A Multi-Stage Multi-Horizon Stochastic Equilibrium Model of Multi-Fuel Energy Markets. Censes working paper 2/2016, NTNU Trondheim, 2015. ISBN: 978-82-93198-15-4.

[114] D. Zwillinger. *CRC standard mathematical tables and formulae.* Chapman and Hall/CRC, 2002.

APPENDIX A

Further remarks on the properties of the distortion premium, identification and robustness

A.1 Properties of the generalized distortion premium

We consider here the generalized distortion premium

$$R(X) = \int_0^1 \nu(\mathrm{AV@R}_\alpha(X))\, k(\alpha)\, d\alpha, \tag{A.1}$$

where $X \in \mathcal{L}^1$, ν a convex, monotone Lipschitz function and k a non-negative weight function on [0,1], which satisfies $\int_0^1 (1-\alpha)^{-1}\, k(\alpha)\, d\alpha < \infty$. Clearly, $X \mapsto R(X)$ is convex and monotone, but is positively homogeneous and/or translation equivariant only if ν is a multiple of the identity. To see this, consider the subdifferential of R at $Y \in \mathcal{L}^1$ is

$$Z_Y = \int_0^1 \nu'(AV@R_\alpha(Y))(1-\alpha)^{-1} \mathbb{1}_{Y > F_Y^{-1}(\alpha)}\, k(\alpha)\, d\alpha \quad \in \mathcal{L}^\infty, \tag{A.2}$$

where F_Y is the distribution function of Y. Notice that $\mathbb{E}(Y \cdot Z_Y)$ depends only on the distribution function F_Y. After some calculation, one finds that

$$\mathbb{E}(Y \cdot Z_Y) = \int_0^1 \nu'(AV@R_\alpha(Y)) \cdot AV@R_\alpha(Y)\, k(\alpha)\, d\alpha.$$

Finally, based on the subdifferential, one gets a dual representation

$$R(X) \;=\; \sup_{Y \in \mathcal{L}^1} \{\mathbb{E}(X \cdot Z_Y)$$

$$- \int_0^1 [\nu'(AV@R_\alpha(Y))AV@R_\alpha(Y) - \nu(AV@R_\alpha(Y))]\, k(\alpha)\, d\alpha\},$$

where Z_Y is given by (A.2).

It is well known (see Pflug and Römisch 2007 [77]) that R is positively homogeneous only if

$$\int_0^1 [\nu'(AV@R_\alpha(Y))AV@R_\alpha(Y) - \nu(AV@R_\alpha(Y))]\, k(\alpha)\, d\alpha = 0,$$

when it is finite. This implies that $\nu(x) = \gamma \cdot x$ for some $\gamma > 0$. R is translation equivariant, if in addition the expectation of the dual multiplier Z_Y is one, which in happens only if $\int_0^1 \gamma\, k(\alpha)\, d\alpha = 1$.

A.2 On different underlying metrics for the Wasserstein distance

There is a whole family of distances on $\mathbb{R}$, which are generalizations of d_1. Set for $x, y \geq 0$, $d_p(x,y) = |x^p - y^p|$. The Wasserstein distance of order 1 with distance d_p is

$$WD_{1,d_p}(F,G) = \int_0^1 |(F^{-1}(v))^p - (G^{-1}(v))^p|\, dv.$$

Lemma A.1. *Notice that for $p \geq 1$*

$$WD_{p,d_1}(F,G) \leq [WD_{1,d_p}(F,G)]^{1/p}.$$

Proof. By the subadditivity of $x \mapsto x^p$ on $\mathbb{R}_{\geq 0}$ one has that $|x-y|^p \leq |x^p - y^p|$ and therefore

$$WD_{p,d_1}(F,G) = \left[\int_0^\infty |F^{-1}(v) - G^{-1}(v)|^p\, dv\right]^{1/p} \leq [WD_{1,d_p}(F,G)]^{1/p}.$$

□

Remark A.1. *This argument also shows that if F has finite $p-$moments and if $WD_{1,d_p}(F,G) < \infty$ (and a fortiori if $WD_{p,d_1}(F,G) < \infty$), then also G has finite $p-$moments. On the other hand, if both F and G have finite $p-$moments, then*

$$WD_{1,d_p}(F,G) \leq p \cdot WD_{p,d_1}(F,G)(1 + \|F^{-1}\|_p^{p-1} + \|G^{-1}\|_p^{p-1})$$

(see Lemma 2.19 in [76]). Therefore, imposing conditions on WD_{1,d_p} or on WD_{p,d_1} leads to quite similar results.

APPENDIX B

Further remarks on recovering distortion functions in power markets

B.1 On the shifted distortion premium

Incorporating ambiguity in the distortion premium with baseline F_θ, the shifted distribution of a cdf F with $\theta < 1$, can be seen as the robust distortion premium with baseline F with a different ambiguity radius. Note that

$$\begin{aligned}
\mathrm{WD}(F_\theta, G) &= \int_0^1 |F_\theta^{-1}(v) - G^{-1}(v)|\, dv \\
&= (1-\theta) \int_0^1 |F^{-1}(v) - G^{-1}(v)|\, dv \\
&= (1-\theta)\mathrm{WD}(F, G)\,.
\end{aligned}$$

Then for $\theta < 1$ and h an increasing (decreasing) distortion density we have

$$\begin{aligned}
\pi_h^\varepsilon(F_\theta) &= \sup\{\pi_h(G) \;:\; \mathrm{WD}(F_\theta, G) \le \varepsilon\} && \text{(inf)} \\
&= \sup\{\pi_h(G) \;:\; \mathrm{WD}(F, G) \le \varepsilon/(1-\theta)\} && \text{(inf)} \\
&= \pi_h^{\varepsilon/(1-\theta)}(F)\,.
\end{aligned}$$

B.2 On the risk premia in the power market

As a remark, the risk premia defined in Bessembinder and Lemmon 2002 [10], as a linear combination of the variance and unstandardised skewness is, in other words assuming the forward price to be the insurance premium:

$$\pi(X) = \mathbb{E}(X) + \alpha\, Var(X) + \gamma\, Skew(X), \tag{B.1}$$

where $Skew(X)$ is the unstandardised skewness. We can think of (B.1) as an extension of the Variance premium principle (2.8). Recall the premium does not fulfill translation equivariance neither positive homogeneous. Furthermore, the future price defined in (B.1) can also be linked to utility functions as we will show. Let V be 3 times differentiable and increasing function defined in $\mathbb{R}_+$. Consider a Taylor expansion of V at $a = \mathbb{E}(X)$ of degree 3 as an approximation of V and write

$$V(X) \approx V(a) + V'(a)(X-a) + \frac{V''(a)}{2!}(X-a)^2 + \frac{V^{(3)}(a)}{3!}(X-a)^3. \tag{B.2}$$

Then take expected values in both sides of (B.2)

$$\mathbb{E}(V(X)) \approx V(a) + \frac{V''(a)}{2!}Var(X) + \frac{V^{(3)}(a)}{3!}Skew(X). \tag{B.3}$$

Now we apply V^{-1} to both sides of (B.3):

$$V^{-1}(\mathbb{E}(V(X))) \approx V^{-1}\left(V(a) + h_1\frac{V''(a)}{2!} + h_2\frac{V^{(3)}(a)}{3!}\right),$$

where $h_1 = Var(X)$ and $h_2 = Skew(X)$ are small. Finally, we do the Taylor expansion of V^{-1} around $V(a)$ with a small step given by $h_1V''(a)/2! + h_2V^{3)}(a)/3!$ obtaining

$$V^{-1}(\mathbb{E}(V(X))) \approx a + (V^{-1})'(V(a))\left(h_1\frac{V''(a)}{2!} + h_2\frac{V^{(3)}(a)}{3!}\right). \tag{B.4}$$

Recall the certainty equivalence principle, the left-hand side of (B.4) is the certainty equivalence premium principle

$$\pi(X) = V^{-1}(\mathbb{E}(V(X))) \approx \mathbb{E}(X) + \frac{V''(a)}{2V'(a)}Var(X) + \frac{V^{(3)}(a)}{3!V'(a)}Skew(X).$$

Hence, for any disutility function V we can find the coefficients $\alpha = V''(a)/(2\,V'(a))$ (the risk-aversion coefficient) and $\gamma = V^{(3)}(a)/(3! \cdot V'(a))$ for (B.1).

B.3 On carbon emissions for coal and gas

Coal and gas market prices were adjusted (divided) by assumed power plant efficiencies of 0.36 and 0.49131, respectively, which correspond to the proportion of energy contained in the fuel that is converted into electricity. The coal price used here is the price of the CIF ARA steam coal front month futures, quoted in USD/mt, converted into EUR with the exchange rate of the corresponding trading day and multiplied by 0.144 for the conversion of metric tons to MWh. The gas price is observed on the NCG spot market, quoted in EUR/MWh. As carbon dioxide emissions must be covered by emission allowances, the costs of the required certificates (European

Emission Allowance, EUA) are also taken into account by multiplying the spot market price of one EUA contract, quoted in EUR/mt, with CO2 adjustment factors of 0.95 (coal) and 0.42 (gas), respectively. The latter represent the number of metric tons of carbon dioxide emitted to produce one MWh of electricity. The efficiency and adjustment factors in the above calculations were taken from the Bloomberg terminal, where they are used for the derivation of dark and spark spreads.

B.4 Data

We use a sample period of five years for the estimation of the spot price model introduced in Section 2.3 and the temperature model (B.10). Model calibrations are updated monthly. Spot prices from the day-ahead auction for the German/Austrian price zone are downloaded from `https://mis.eex.com`. Daily temperature data for three German weather stations (B.-Tempelhof, Hannover-Langenhagen, München-Flughafen) are obtained from the Climate Data Center of *Deutscher Wetterdienst*, `ftp://ftp-cdc.dwd.de/pub/CDC/`. For all three we construct the average daily temperature as $Tp_i := 0.5 \cdot (Tp_i^n + TP_i^x)$, where Tp_i^n is the minimum and Tp_i^x the maximum temperature for station $i = 1, \ldots, 3$ on a certain day. Then the daily temperature for the calculation of *CDD* and *HDD* as defined in Appendix B.5 is calculated as mean $Tp := Tp_1 + Tp_2 + Tp_3$ of these three observations.

B.5 Seasonality component

The estimation of the spot price model defined in (2.28) requires first the identification of the seasonal component s_t from equation (2.23), which itself can be decomposed into two parts:

$$s_t = f2y_{d(t)} \cdot f2d_t, \tag{B.5}$$

$f2y_d$ ("factor-to-year") reflects the seasonal pattern of daily prices that are typically larger in winter than in summer and higher on workdays than on weekends or holidays. It is defined as the relative weight of the base price, i.e., the average of all hours, of the day d compared to the annual average, where $d := d(t)$ maps time t to a specific calendar day:

$$f2y_d = \frac{base_d}{\frac{1}{D}\sum_{k=d-D+1,\ldots,d} base_k}. \tag{B.6}$$

Herein $base_k := \frac{1}{24}\sum_{\{t|d(t)=k\}} spot_t$ is the daily base price, the denominator represents the moving average of the base prices from the previous 12 months and $D = 365$ (or $D = 366$ if a leap day occurred in the corresponding period).

$f2d_t$ ("factor-to-day") as the second component in (B.5) defines the weight of the price of a particular hour compared to the daily base price,

$$f2d_t = \frac{spot_t}{base_{d(t)}}. \tag{B.7}$$

It reflects the intra-day seasonality of spot prices due to different demand in individual hours.

For the model calibration both factors, $f2y_d$ defined in (B.6) and $f2d_t$ defined in (B.7), are calculated from historical spot price data. Outliers, i.e., prices beyond plus/minus three standard deviations from the annual average, as well as negative prices are eliminated beforehand. Note that this applies only to the estimation of the seasonal component, not to the subsequent estimation of remaining stochastic components in (2.28) since extreme observations are represented by spikes. The variation of $f2y_d$ is explained by a multiple regression model:

$$f2y_d = a^y + \sum_{i=1}^{6} b_i^y D_{d,i} + \sum_{i=1}^{12} c_i^y M_{d,i} + d^y CDD_d + e^y HDD_d + \epsilon_d, \tag{B.8}$$

The explanatory data are dummy variables $D_1, \ldots, D_6$ for the different weekdays and $M_1, \ldots, M_{12}$ for the different months. A dependency on temperature is taken into account by cooling degree days $CDD := \max\{Tp - 18.3, 0\}$ and heating degree days $HDD := \max\{18.3 - Tp, 0\}$, where Tp represents the daily average temperature in centigrade. The intra-day seasonal pattern differs very much between workdays, Saturdays and Sundays, and depends also on the time of the year. Therefore, we estimate the following regression model separately for workdays in different months as well as for Saturdays and Sundays in different quarters, which leads to 20 different parameter sets:

$$f2d_t = a^p + \sum_{i=1}^{23} b_i^p H_{t,i} + \epsilon_t, \qquad p = 1, \ldots, 20, \tag{B.9}$$

where $H_1, \ldots, H_{23}$ represent dummy variables for the individual hours of the day. For details on the estimation of both factors, we refer to Blöchlinger 2008 [11].

After estimation of the regression models (B.8) and (B.9), the values of the *explained* seasonality component $\widehat{s}_t$ are derived from equation (B.5) using the explained values $\widehat{f2y}_d$ and $\widehat{f2d}_t$. We use historic temperature data for the estimation of (B.6). In this way, the stochastic component that is derived as difference between observed $f2y_d$ and explained $\widehat{f2y}_d$ includes only variations caused by deviations between demand and supply that are *not* related to temperature. To take also the latter into account, we model temperature variation over the year by an additional model based on Fourier series:

$$Tp_d = a^{Tp} + \sum_{i=1}^{3} b_1^{Tp} \cos\left(\frac{2\pi i}{365} dn_d\right) + \sum_{i=1}^{3} b_2^{Tp} \sin\left(\frac{2\pi i}{365} dn_d\right) + \epsilon_d, \tag{B.10}$$

where dn_d is the number of days on date d since the beginning of the year. It turns out that residuals in (B.10) can be modeled by an AR(1) process. For our spot price simulations, we generate in a first step trajectories for the stochastic daily temperature based on this approach. Then realizations of $f2y_d$ and, in combination with the deterministically calculated $f2d_t$, scenarios of the seasonality component s_t are obtained for future time points. Finally, the latter are combined with simulations of the stochastic components in (2.28).

In this respect, our approach differs also from Paraschiv et al. 2015 [72], where the $\widehat{f2y}_d$ are estimated using explained values from the temperature model (B.10) rather than historical data, and temperature variations are not modeled as additional stochastic factor. Instead, variations that carry over from the previous day caused by temperature are largely explained by the coefficient for lag 24 in the autoregressive process for the base regime in model (2.21).

B.6 Model estimation

B.6.1 Outline of Kalman filter

As outlined in section 2.3, the logarithm of the adjusted spot price Y_t follows the Ornstein-Uhlenbeck process (2.27) around a non-observable price level, which itself follows the stochastic process (2.25). Such systems can be estimated efficiently with a Kalman filter, e.g., see Grewal and Andrews 2015 [50] for a thorough introduction. Before we describe the estimation procedure for the specific model used here, we give in the sequel a brief summary of the basic idea of Kalman filtering. For the sake of notational convenience, we restrict ourselves to the one-dimensional case, i.e., all quantities and the parameters $a_j, b_j, c_j, d_j \in \mathbb{R}$ in the following equations are scalars.

A sequence of realizations $y_1, y_2, \ldots, y_N$ of random variable Y_t is observed at discrete time points $t_1, t_2, \ldots, t_N$, where the index $j = 1, 2, \ldots, N$ refers to time t_j. It is assumed that y_j depends (linearly) on the unobservable state variable $x_j \in \mathbb{R}$ and is biased with some normally distributed noise v_j with mean zero and variance φ_j:

$$y_j = a_j + b_j x_j + v_j, \qquad v_j \sim \mathcal{N}(0, \varphi_j). \tag{B.11}$$

We call (B.11) *measurement equation.* It links an observation with the latent state variable x_j, which is a linear combination of the previous state x_{j-1} and some normally distributed noise w_j with variance ψ_j as described by the *transition equation*:

$$x_j = c_j + d_j x_{j-1} + w_j, \qquad w_j \sim \mathcal{N}(0, \psi_j), \tag{B.12}$$

The Kalman Filter is a recursive estimator and consists conceptually of two distinct phases. Given that the state of the system is known up to step $j-1$, first a *prediction* x_j^- of the state estimate in the new step k and its variance P_j^- are computed:

$$x_j^- = c_j + d_j x_{j-1}^+, \tag{B.13}$$
$$P_j^- = d_j P_{j-1}^+ d_j + \psi_j, \tag{B.14}$$

where x_{j-1}^+ and P_{j-1}^+ are the results from the previous iteration. The *update* step starts with an observation of the measurement y_j that allows to calculate the error ϵ_j of the prediction x_j^- in (B.13) and its variance R_j:

$$\epsilon_j = y_j - a_j - b_j x_j^-, \tag{B.15}$$
$$R_j = b_j P_j^- b_j + \varphi_j. \tag{B.16}$$

Then the Kalman gain K_j is given by

$$K_j = P_j^- b_j / R_j, \tag{B.17}$$

which allows to improve the estimate of the state variable and its variance:

$$x_j^+ = x_j^- + K_j \epsilon_j, \tag{B.18}$$
$$P_j^+ = (1 - K_j b_j) P_j^-. \tag{B.19}$$

The Kalman gain obtained from (B.17) defines the weight of the current measurement relative to the prediction (B.13) for the improvement of the state estimate. According to (B.18) the impact of the new measurement increases with the variance P_j^- of the prediction x_j^- from (B.14) and decreases with the variance R_j of the error ϵ_j from the current measurement defined in (B.16). The variance P_j^+ of the updated state estimate x_j^+ in (B.19) determines to what extent the current measurement influences the next prediction. While the *prediction* and the *update* phase usually alternate, the update may also be skipped and multiple prediction steps performed once an observation is not available.

B.6.2 Application to spot price model

In order to apply the outlined Kalman filter procedure to the estimation of the model introduced in section 2.3, we consider time-discrete versions of the continuous-time processes (2.25) and (2.27). Note that lower case letters are used to distinguish in the notation between the discrete and continuous time cases:

$$y_j = y_{j-1} \cdot \exp(-\kappa_{(j)} \Delta t_j) + x_j \left(1 - \exp(-\kappa_{(j)} \Delta t_j)\right) + v_j, \tag{B.20}$$
$$x_j = x_{j-1} + (\mu - 0.5 \cdot \eta^2)\, \Delta t_j + w_j, \tag{B.21}$$

with $\Delta t_j := t_j - t_{j-1}$. Here $\kappa_{(j)} := \kappa_{\beta(t_j)}$ specifies the value of the parameter at the time t_j of the j-th observation, and likewise the indices of the other parameters must be understood. By comparison with equations (B.11) and (B.12), we find that the Kalman filter can be applied with

$$a_j = y_{j-1} \cdot \exp(-\kappa_{(j)} \Delta t_j), \qquad b_j = 1 - \exp(-\kappa_{(j)} \Delta t_j),$$
$$c_j = (\mu - 0.5 \cdot \eta^2)\, \Delta t_j, \qquad d_j = 1.$$

The variances φ_j of the noise v_j in (B.20) and of the noise ψ_j in (B.21) are

$$\varphi_j = \frac{\sigma^2_{(j)}}{2\kappa_{(j)}}\left(1 - \exp(-2\kappa_{(j)}\Delta t_j)\right), \qquad \psi_j = \eta^2 \Delta t_j. \tag{B.22}$$

For the identification of the unobservable state variables x_j (logarithm of price level), the initial state estimate x_0^+ and its variance P_0^+ must be set to starting values. The latter are additional parameters and result from the model estimation. During the alternating execution of *prediction* and *update* steps, extreme observations are classified as spikes and excluded from the estimation of the base regime parameters.

In the sequel we denote by $\widehat{s}_j$ the explained value of the seasonality component defined in (B.5), calculated with the estimated coefficients of the regression models (B.8) and (B.9), and by $spot_j^{obs}$ the observed spot price at time t_j in the discrete time framework. We define further by $\mathcal{B}$, $\mathcal{U}$ and $\mathcal{L}$ the set of indices of those observations that are assigned to the base, the upper or the lower spike regime, respectively. Then the filter consists of the following steps:

1. Initialization of x_0^+ and P_0^+. Set $j := 1$ and $\mathcal{B} = \mathcal{U} = \mathcal{L} := \emptyset$.
2. Obtain a prediction x_j^- of the state variable and its variance P_j^- from (B.13) and (B.14). Calculate the regime limits in step j:
$$l_j^U := \widehat{s}_j \cdot \exp\left(\alpha^U_{(j)} \cdot x_j^-\right) \qquad \text{and} \qquad l_j^L := \widehat{s}_j \cdot \exp\left(-\alpha^L_{(j)} \cdot x_j^-\right).$$
3. Observe the spot price $spot_j^{obs}$ and perform one of the following steps:
   (a) If $spot_j^{obs} < l_j^L$,
   set $y_j := l_j^L - spot_j^{obs}$, $\mathcal{L} := \mathcal{L} \cup \{j\}$, $x_j^+ := x_j^-$ and $P_j^+ := P_j^-$.
   (b) If $spot_j^{obs} > l_j^U$,
   set $y_j := spot_j^{obs} - l_j^U$, $\mathcal{U} := \mathcal{U} \cup \{j\}$, $x_j^+ := x_j^-$ and $P_j^+ := P_j^-$.
   (c) If $l_j^L \leq spot_j^{obs} \leq l_j^U$,
      - set $y_j := \ln(spot_j^{obs}/\widehat{s}_j)$, $\mathcal{B} := \mathcal{B} \cup \{j\}$ and calculate
      - the prediction error ϵ_j from (B.15) and its variance R_j from (B.16),
      - the Kalman gain K_j from (B.17) and
      - the update of the state variable estimate x_j^+ and its variance P_j^+ from (B.18) and (B.19).
4. If $j < N$, set $j := j + 1$ and goto step 2. □

After execution of the filter algorithm, the following log-likelihood function can be calculated:

$$
\begin{aligned}
lnL\left(spot_1^{obs},\ldots,spot_N^{obs}|\Phi,\Theta\right) = &-\frac{1}{2}\sum_{j\in\mathcal{B}}\left[\ln(2\pi)+\ln R_j+\epsilon_j^2/R_j\right]\\
&+\sum_{j\in\mathcal{L}}\left[\ln k_{(j)}^- - k_{(j)}^-\ln\lambda_{(j)}^- + (k_{(j)}^- - 1)\cdot\ln y_j - (y_j/\lambda_{(j)}^-)^{k_{(j)}^-}\right]\\
&+\sum_{j\in\mathcal{U}}\left[\ln k_{(j)}^+ - k_{(j)}^+\ln\lambda_{(j)}^+ + (k_{(j)}^+ - 1)\cdot\ln y_j - (y_j/\lambda_{(j)}^+)^{k_{(j)}^+}\right]
\end{aligned}
\tag{B.23}
$$

where

$$\Phi := (\mu,\eta,\kappa_1,\kappa_2,\sigma_1,\sigma_2,x_0^+,P_0^+)',$$

is a vector with the stacked parameters of the base regime model (including initial values for the Kalman filter) and

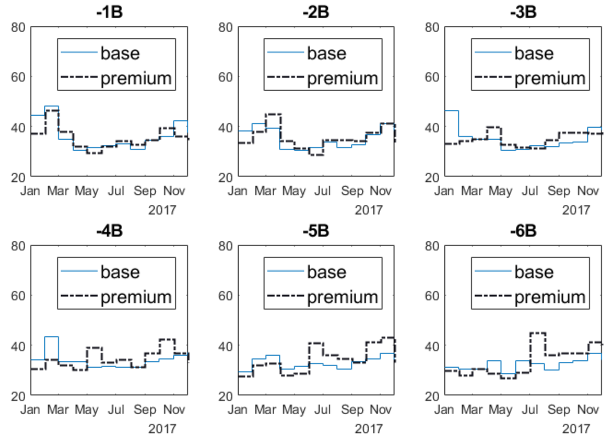
$$\Theta := (\lambda_1^+,\ldots,\lambda_8^+,k_1^+,\ldots,k_8^+,\lambda_1^-,\ldots,\lambda_8^-,k_1^-,\ldots,k_8^-)'.$$

contains the parameters of the spike regimes and regime limits. Recall that we distinguish for the base regime between parameter sets for summer vs. winter and for the spike regimes between summer/winter, workdays/ weekends and daytime/nighttime. The first row in (B.23) results from the normally distributed errors in the base regime, the second and the third row are derived from the density of the Weibull distribution that is assumed for the spikes, e.g., see Johnson et al. 1996 [55].

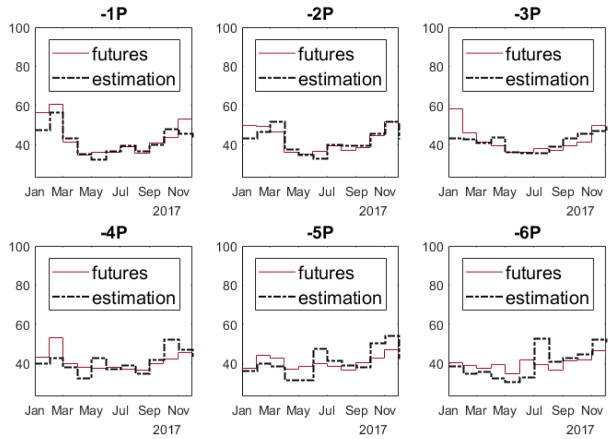
The maximization of the log-likelihood function (B.23) allows the identification of those parameters that fit the model best to the observed data, which requires in each iteration an execution of the above described Kalman filter algorithm to identify the values of the unobservable state. The recursive nature of the filter implies complex dependencies of the involved variables on the parameters. Furthermore, the unique allocation of observations to different regimes, which is an outcome of the filter, leads to a non-smooth objective of the corresponding optimization problem. Therefore, we employ the derivative-free Nelder/Mead method for the maximization of (B.23). However, it is known that the efficiency of this method deteriorates with increasing dimension, e.g., see Han and Neumann 2006 [51], and so we reduce the number of parameters that are identified by the optimization procedure by two measures.

Firstly, the regime limit parameters are linked to the volatility of the spot price log-transform by $a_{(j)}^U = a_{(j)}^L := k\cdot(\varphi_j + b_j\cdot\psi_j)^{\frac{1}{2}}$, where φ_j and ψ_j are defined in (B.22). In this way, the four parameters $\alpha_1^U,\alpha_2^U,\alpha_1^L,\alpha_2^L$ are removed from the estimation problem. Preliminary tests indicated that choosing $k = 3$ as multiple of the volatility provides the best fit to the data. Secondly, since the spike realizations are identified in steps 3a and 3b of the Kalman filter, the Weibull distributions assumed in (2.29) for the spike regimes can be fitted directly to the data by computationally simple procedures like moment matching or regression. We refrain here from a detailed discussion, but refer to Johnson et al. 1996[55], Genschel and Meeker 2010 [42] or Yavuz 2013 [110] for an overview and tests of appropriate methods. Therefore, the parameters in Θ can be obtained immediately after execution of the filter, and only the parameters in Ψ are identified by the Nelder/Mead algorithm, which reduces dimension of the corresponding optimization problem to eight.

B.7 Further figures and tables



(a) Base



(b) Peak

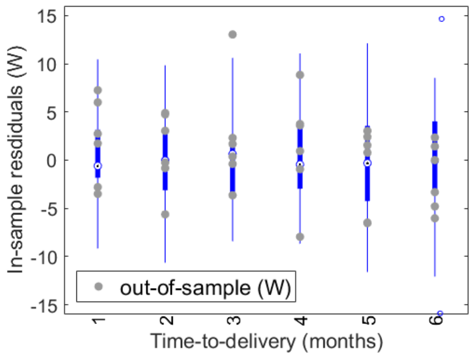
Figure B.1: Out-of-sample comparison between observed monthly futures prices and estimated prices. Futures are separated by time-to-maturity.

Normalized mean square error(NMSE)

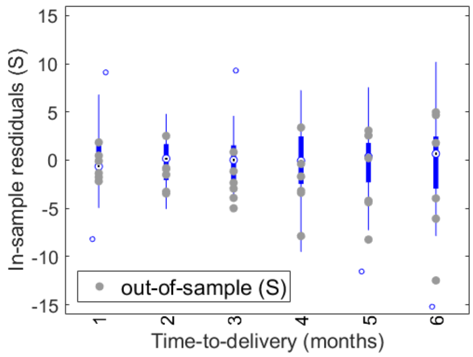
k	1	2	3	4	5	6
Base	0.8429	0.8664	0.8811	0.8642	0.8238	0.7577
Peak	0.8364	0.8328	0.8333	0.8076	0.7505	0.6784

Table B.1: NMSEs for our estimated model (2.31). Values are separated per time-to-delivery. The goodness of fit varies from -Inf (bad fit) to 1 (perfect fit).

From Table B.2 we see that the shifted distortion premium is always significant (variable corresponding to β_1). For base futures, the premium explains around 90% of futures for front months, then around 85% for $k = 3, 4, 5$ and the 75% for $k = 6$. For peak futures, we see that for longer time-to-delivery, the intercept and

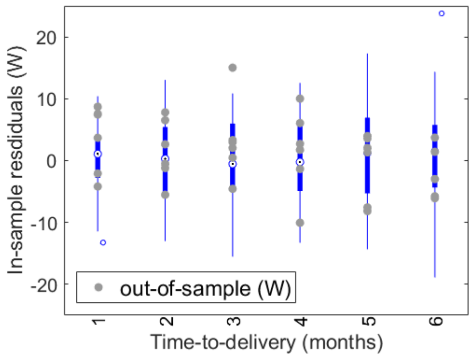


(a) Winter

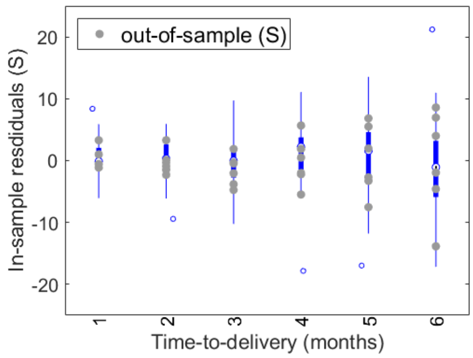


(b) Summer

Figure B.2: Box plot of residuals for the in-sample estimations separated in (a) winter and (b) summer. For each group of base futures, we add the residuals (points) obtained with the out-of-sample estimations.



(a) Winter



(b) Summer

Figure B.3: Box plot of residuals for the in-sample estimations separated in (a) winter and (b) summer. For each group of peak futures, we add the residuals (points) obtained with the out-of-sample estimations.

coefficient for the ambiguity are also significant. We can conclude our model explains better base futures than peak futures. Although for base futures, the intercept and ambiguity premium are non-significant, their inclusion in the model influenced in the shape of the distortions and the value of the shifting parameters. For peak futures up to 3 months ahead, we also have a good fit with our model. Our model explains around 80% the peak futures for these groups. For $k = 4, 5, 6$, recall the ambiguity premium was estimated to be zero. In these cases, we need an intercept (around 10 €/MWh) and to add an ambiguity premium.

(k) Base	Est.	SE	p-Value	adj. R2
β_0	-0.307	8.323	0.971	
(1)β_1	0.934	0.0456	8.487e-34	0.843
β_2	3.505	7.592	0.646	
β_0	1.186	3.992	0.767	
(2) β_1	0.908	0.040	5.090e-37	0.872
β_2	3.064	3.3237	0.359	
β_0	-0.111	4.933	0.982	
(3) β_1	0.893	0.0364	3.159e-39	0.89
β_2	4.380	4.037	0.281	
β_0	-15.138	17.164	0.380	
(4) β_1	0.856	0.036	5.760e-38	0.883
β_2	15.893	12.543	0.209	
β_0	-0.587	5.449	0.914	
(5) β_1	0.813	0.039	1.317e-34	0.86
β_2	6.128	3.617	0.094	
β_0	-5.387	8.658	0.536	
(6) β_1	0.752	0.040	1.596e-31	0.833
β_2	9.221	4.790	0.058	

(a)

(k) Peak	Est.	SE	p-Value	adj. R2
β_0	0.478	6.022	0.937	
(1) β_1	0.834	0.038	1.832e-35	0.864
β_2	5.895	4.197	0.164	
β_0	1.630	4.59	0.723	
(2) β_1	0.820	0.038	4.617e-35	0.865
β_2	5.311	2.957	0.076	
β_0	-2.469	6.250	0.694	
(3) β_1	0.801	0.037	8.125e-36	0.875
β_2	7.93	3.782	0.039	
(4) β_0	11.01	1.788	2.584e-08	0.868
β_1	0.791	0.034	4.982e-38	
(5) β_0	12.163	2.159	2.424e-07	0.818
β_1	0.782	0.040	2.757e-32	
(6) β_0	14.795	2.287	6.676e-09	0.784
β_1	0.738	0.042	2.823e-29	

(b)

Table B.2: Results of fitting a regression model to $F(\tau(k,m),m) \sim \beta_0 + \beta_1\pi_{h;\theta}(\cdot) + \beta_2(\pm\varepsilon||h||_\infty)$. Estimations are separated by time-to-delivery: $k = 1,\dots,6$ months.

APPENDIX C

Further remarks on distributionally robust optimization with multiple time scales

C.1 Parameters for the model

The parameters corresponding to equation (3.7) in Section 3.2.2 are given by:

$$
\begin{aligned}
a_{11}(t) &= \sigma_{\mathrm{t}}^{\mathrm{e,sh}} + \sigma_{\mathrm{t}}^{\mathrm{e,lo}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}}, \qquad a_{12}(t) = \sigma_{\mathrm{t}}^{\mathrm{e,lo}}\, \frac{\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{f}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}}}{\sqrt{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2}}, \\
a_{13}(t) &= \sigma_{\mathrm{t}}^{\mathrm{e,lo}}\, \frac{\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{c}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}} - \frac{(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{f}})(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}} - \varrho_{\mathrm{t}}^{\mathrm{f,c}})}{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2}}{\sqrt{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}})^2 - \frac{\left(\varrho_{\mathrm{t}}^{\mathrm{f,c}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}}\right)^2}{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2}}}, \\
a_{14}(t) &= \sigma_{\mathrm{t}}^{\mathrm{e,lo}} \sqrt{\Bigg\{ 1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}})^2 - \frac{(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{f}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}})^2}{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2}} \\
&\quad - \frac{(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{f}} + \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{c}} - (\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{c}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{lo}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{f,c}} + \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{e}^{\mathrm{lo}}}\, \varrho_{\mathrm{t}}^{\mathrm{f,c}})^2}{((\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2-1)((\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2 + (\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}})^2 + (\varrho_{\mathrm{t}}^{\mathrm{f,c}})^2 - 2\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}}\, \varrho_{\mathrm{t}}^{\mathrm{f,c}} - 1)} \Bigg\}, \\
a_{21}(t) &= \sigma_{\mathrm{t}}^{\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}, \qquad a_{22}(t) = \sigma_{\mathrm{t}}^{\mathrm{f}} \sqrt{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2}, \qquad a_{31}(t) = \sigma_{\mathrm{t}}^{\mathrm{c}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}}, \\
a_{32}(t) &= \sigma_{\mathrm{t}}^{\mathrm{c}}\, \frac{\varrho_{\mathrm{t}}^{\mathrm{f,c}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}}}{\sqrt{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2}}, \qquad a_{3,3} = \sigma_{\mathrm{t}}^{\mathrm{c}} \sqrt{1-(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}})^2 - \frac{(\varrho_{\mathrm{t}}^{\mathrm{f,c}} - \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}}\, \varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{c}})^2}{(\varrho_{\mathrm{t}}^{\mathrm{e}^{\mathrm{sh}},\mathrm{f}})^2 - 1}}\,.
\end{aligned}
$$

The parameters of the multivariate lognormal distribution of the GBM process (3.7), i.e., the expectation vector $\mu(t)$ and the (components of the) variance-covariance matrix $\Sigma(t)$ of the associated multivariate normal distribution are given by:

$$\mu(t) = \begin{pmatrix} \log(F^e_{0,t}) - \frac{1}{2}\int_0^t a_{11}^2(s) + a_{12}^2(s) + a_{13}^2(s) + a_{14}^2(s))\, ds \\ \log(F^f_{0,t}) - \frac{1}{2}\int_0^t a_{21}^2(s) + a_{22}^2(s)\, ds \\ \log(F^c_{0,t}) - \frac{1}{2}\int_0^t a_{31}^2(s) + a_{32}^2(s) + a_{33}^2(s)\, ds \end{pmatrix},$$

$$\begin{aligned}
&\Sigma^{11}(t) = \int_0^t a_{11}^2(s) + a_{12}^2(s) + a_{13}^2(s) + a_{14}^2(s)\, ds\,, \\
&\Sigma^{12}(t) = \Sigma^{21}(t) = \int_0^t a_{11}(s)a_{21}(s) + a_{12}(s)a_{22}(s)\, ds\,, \\
&\Sigma^{13}(t) = \Sigma^{31}(t) = \int_0^t a_{11}(s)a_{31}(s) + a_{12}(s)a_{32}(s) + a_{13}(s)a_{33}(s)\, ds\,, \\
&\Sigma^{23}(t) = \Sigma^{32}(t) = \int_0^t a_{21}(s)a_{31}(s) + a_{22}(s)a_{32}(s)\, ds\,, \\
&\Sigma^{22}(t) = \int_0^t a_{21}^2(s) + a_{22}^2(s)\, ds\,,\ \ \Sigma^{33}(t) = \int_0^t a_{31}^2(s) + a_{32}^2(s) + a_{33}^2(s)\, ds\,.
\end{aligned}$$

C.2 A remark on the nested distance

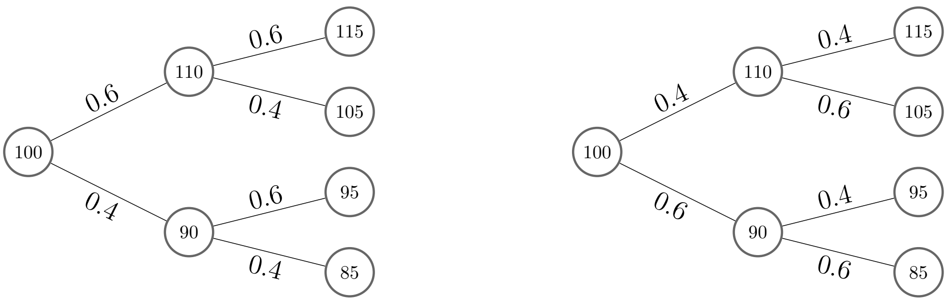


Figure C.1: Two versions of a scenario tree with different transition probabilities. The nested distance is 10. The Wasserstein distance between the first stage subtrees is 4, that between each of the two pairs of second stage subtrees is 2.

Definition C.1. *The nested distance* $\mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}})$ *between two* $\mathbb{R}^m$*-valued filtered stochastic processes* $(\mathbb{P}, \mathcal{F})$ *and* $(\tilde{\mathbb{P}}, \tilde{\mathcal{F}})$ *is defined as the optimal value of the following mass transportation problem, which optimizes over the set of all joint distributions that respect the given conditional marginals:*

$$\begin{aligned}
&\inf_\pi \iint \|\omega - \tilde{\omega}\|\ \pi(d\omega, d\tilde{\omega}) \\
&\text{s.t. } \pi\left(A \times \mathbb{R}^m \middle| \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t\right) = \mathbb{P}\left[A \middle| \mathcal{F}_t\right] \quad A \in \mathcal{F}_T;\ \forall\, 0 \le t \le T \\
&\quad\ \pi\left(\mathbb{R}^m \times B \middle| \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t\right) = \mathbb{P}\left[B \middle| \tilde{\mathcal{F}}_t\right] \quad B \in \tilde{\mathcal{F}}_T;\ \forall\, 0 \le t \le T.
\end{aligned}$$

Figure C.1 below illustrates that the nested distance (between two trees resulting from a variation of the transition probabilities) cannot be bounded by only considering the Wasserstein distance between the subtrees with matching node values. In the given example, the maximum Wasserstein distance in the second stage is 2, the first case Wassersein distance is 4. Thus, $4 + 2 \cdot 2 = 8$ would still be smaller than the nested distance between the two trees, which is 10.

Appendix Abstract

Prices of contracts with risky aspects are typically linked to specific uncertainties and probabilities of adverse scenarios. Insurance companies carry the risk of losses in exchange for a premium, which depends on the loss distribution. Another example where risk is exchanged for a fixed price is swap contracts. Electricity futures can be seen as swaps where the floating component are spot prices and the fixed component is a constant price for delivering electricity over a longer period. The primary goal of this thesis is the incorporation of model ambiguity for pricing these contracts. Moreover, we contemplate the complex structure of energy markets. For this reason, we also explore pricing a real option under model ambiguity.

First of all, we study the theoretical properties of the distortion principle for insurance pricing. We find closed-form solutions for the optimal distortion premium under model ambiguity using Wasserstein distances. In various cases, we also find the distributions that reach the optimal prices. For the distortion principle, we can conclude that the price to pay for ambiguity only depends on the ambiguity radius and the distortion function, but not on the initial distribution. Additionally, we characterize the unboundedness of the robust distortion premium. Besides, we investigate the identification of distortion functions from observed prices. We propose a method to recover them from simulated prices in two cases: the average value-at-risk and power distortion principle.

In the second part of this thesis, we bring together insurance pricing rules and electricity futures pricing rules. Due to the non-storability of electricity, many authors study different rules and empirical results to explain futures prices and the risk premia in this market. We extend the present literature and propose to explain the price formation of these contracts with three different quantities: the distortion premium, a correction factor and an ambiguity premium. These three factors capture main characteristics of futures prices and the risk premia. Among them, we recover a seasonal pattern of the risk premia and explain the changes in risk aversion depending on time-to-delivery. The ambiguity premium is significant and increases with time-to-delivery for base futures. For these calculations, we specify a new regime-switching model for spot prices.

The last part of this thesis studies an appropriate evaluation of a thermal power plant by incorporating model ambiguity. The different uncertainties that affect the expected profits of this real option are electricity prices, fuel prices and CO2 allowances. The power plant takes weekly decisions fixing the production for an

entire week, while the uncertainties may affect the profit within weeks. Firstly, we discretize and quantize the uncertainties in a lattice process. To simulate different prices within weeks, we introduce an interpolation process called bridge process. Secondly, we propose a distance between lattice processes, which is tractable for solving dynamic problems backwards in time. This distance is a Wasserstein distance type with an underlying metric dependent on the state of the power plant. Our empirical results show that the larger the ambiguity radius is, the more conservative the production, and the less the achieved profit is. Although we solve a specific problem, our results can be applied to different multistage decision problems.

Appendix Zusammenfassung

Preise von Verträgen mit riskanten Aspekten sind typischerweise mit spezifischen Unsicherheiten und Wahrscheinlichkeiten von unvorteilhaften Szenarien verbunden. Versicherungsunternehmen tragen Verlustrisiken im Austausch für Prämien, die von den Schadensverteilungen abhängen. Ein weiteres Beispiel, bei dem Risiko gegen einen fixen Preis ausgetauscht wird, sind Swap-Verträge: Strom-Futures können als Swap-Verträge angesehen werden, wobei die variable Komponente Spot-Preise und die fixe Komponente konstante Preise für die Stromlieferung einer längeren Periode darstellen. Das Hauptziel dieser Arbeit ist die Integration von Model-Unsicherheiten in die Preisgestaltung dieser Verträge. Es werden komplexe Strukturen in den unterschiedlichen Bereichen von Energiemärkten untersucht und die Preisgestaltung einer Realoption unter Modell-Ambituität untersucht.

Zunächst werden Distortionsprinzipien bei der Preisgestaltung von Versicherung behandelt. Es werden geschlossene Lösungen für optimale Distortionsprämien unter Modell-Ambiguität mit Wasserstein-Distanzen berechnet. In verschiedenen Fällen werden die Verteilungen, die den optimalen Preis erreichen, errechnet. Bezüglich des Distortionsprinzips kann festgehalten werden, dass der Preis für Ambiguität nur vom Ambiguitätsradius und den Distortionsfunktionen abhängt und nicht von der Basis-Verteilung. Zusätzlich wird die Unbeschränktheit der robusten Distortionsprämie charakterisiert. Außerdem werden Distortionsfunktion auf Grund der beobachteten Preise identifiziert. Dafür wird eine Methode vorgeschlagen, um die Distortionsfunktionen nicht-parametrisch in zwei Fällen zu ermitteln: Average value-at-risk und Potenzfunktionen.

Im zweiten Teil der Arbeit werden Regeln für die Versicherungspreisgestaltung und Regeln für die Strom-Future-Preisgestaltung gemeinsam betrachtet. Dadurch, dass Elektrizität nicht gespeichert werden kann, untersuchen viele Autoren die empirischen Daten, um die Future-Preise und die Risikoprämien in diesem Markt zu erklären. Die vorhandene Literatur wird erweitert und es wird eine Erklärung der Preisbildung dieser Verträge mit drei unterschiedlichen Komponenten vorgeschlagen: eine Distortionsprämie, einen Korrekturfaktor und eine Ambiguitätsprämie. Diese drei Teile stellen einen allgemeinen Mechanismus von Future-Preisen dar. Es wurde nachgewiesen, dass die Future-Preise mit langer Restlaufzeit zunehmen. Zudem wird ein saisonales Muster der Riskioprämien identifiziert und die Veränderung in der Risikoaversion in Abhängigkeit von den Restlaufzeiten erklärt. Die Ambiguitäts-Prämie ist nicht null und nimmt mit Restlaufzeiten der Base-Future-Verträge zu.

Für diese Berechnungen wird ein neues Regime-Switching-Modell für Spot-Preise spezifiziert.

Der letzte Teil dieser Arbeit untersucht eine angemessene Bewertung eines Wärmekraftwerks unter Einbeziehung von Modell-Ambiguität. Die verschiedenen Unsicherheiten, die sich auf die erwarteten Profite dieser Realoption auswirken, sind Strompreise, Treibstoffpreise und CO2-Zertifikate. Für das Wärmekraft-werk werden wöchentlich Entscheidungen getroffen, die die Produktion einer gesamten Woche festlegen, obwohl die Unsicherheiten den Profit innerhalb einer Woche beeinflussen können. Zunächst werden die Unsicherheiten in einem Lattice-Prozess diskretisiert und quantisiert. Um unterschiedliche Preise innerhalb einer Woche zu simulieren, wird ein Interpolations-Prozess, genauer ein Bridge-Prozess, vorgeschlagen. Folgend wird eine Distanz zwischen Lattice-Prozessen vorgeschlagen, die geeignet ist dynamische Probleme rückwärts in der Zeit zu lösen. Diese Distanz ist eine uniforme Wasserstein-Distanz mit einer zugrundeliegenden Metrik in Abhängigkeit von den Zuständen des Wärmekraftwerks. Die emprischen Ergebnissen zeigen dass die Produktion konservativer und der Profit umso geringer ist, je größer der Ambiguitätsradius ist. Es wird zwar ein spezifisches Problem gelöst, jedoch können die Ergebnisse auf viele ähnliche mehrstufige Entscheidungsprobleme angewendet werden.