DISS. ETH NO. 27355

RISK MANAGEMENT IN ENERGY MARKETS:

ANALYTICAL AND MACHINE LEARNING PERSPECTIVES

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
XI KLEISINGER-YU

MSc., Technische Universität Berlin
born 03.02.1988
citizen of Germany

accepted on the recommendation of

Prof. Dr. J. Teichmann	examiner
Prof. Dr. M. Larsson	co-examiner
Prof. Dr. C. Cuchiero	co-examiner

2021

To my daughter Lisa

Abstract

This thesis deals with the pricing and the risk management of energy markets, in particular the mid- to long-term markets of electricity and natural gas. The main contributions are made in proposing modeling frameworks for pricing energy forwards and options, and for the gas storage optimization, and towards the quantification of long-term model risk in energy markets.

Firstly, we propose a multifactor polynomial framework to model and hedge long-term electricity contracts with delivery period. This framework has several advantages: the computation of forwards, risk premia and cross-maturity correlations is fully explicit, and the model can be calibrated to observed electricity forward curves efficiently and accurately. In this framework, we suggest a rolling hedge which only uses liquid forward contracts and is risk-minimizing in the sense of Föllmer and Schweizer [1991]. We calibrate the model to over eight years of German power calendar year forward curves and investigate the quality of the risk-minimizing hedge over various time horizons.

Based on the polynomial framework, we propose a quadratic Gaussian framework to model long-term electricity options that are written on forwards with delivery period. This framework is a subclass of the polynomial framework, which still generalizes the two-factor model. Thus it inherits all the strengths of the polynomial framework. In this model, we develop an exponential-quadratic transform formula, which allows us to compute the characteristic function by solving Riccati equations, and to price options accurately using a Fourier approach. The advantage of this model is that it allows a consistent way of pricing options and the underlying forwards simultaneously without approximations and simplifying assumptions. Moreover, we calibrate the model to various observed options and forwards of the German power exchange.

Furthermore, we propose two machine learning models for the gas storage optimization problem, which uses the deep hedging approach developed by Buehler, Gonon, Teichmann, and Wood [2019]. Both proposed models are the first deep hedging models for this field. The first model

(Model I) uses spot as hedging instrument, and the second model (Model II) extends Model I by using forwards with monthly delivery periods as additional hedging instruments. Moreover, we compare the terminal profit and loss generated by the strategies of our trained models with that using Least Squares Monte Carlo (LSMC), both in in-sample and out-of-sample tests. We find that Model I provides results that are close to LSMC, and Model II significantly outperforms LSMC and is also more volatile.

Last but not least, we provide an extensive review of the model risk quantification literature in which we group the applicable methods into three categories: the pairwise model comparison, the Bayesian model averaging, and the worst-case approach. We present two applications to quantify model risk in modeling and hedging long-term energy options. In the first application, we hedge an illiquid energy call using a misspecified model, and find in numerical tests that the model risk outweighs the tracking error of the hedge and cannot be neglected. For the second application, we use the entropy approach of Glasserman and Xu [2014], which gives explicit model risk with respect to the divergence in entropy terms from a baseline model. In the second application, we calibrate three different models to the real data, where we set one model as the baseline model, and compute the entropy levels of each of the other models, and quantify the model risk with respect to each entropy level. This approach is able to assign a number to the model risk with very little computational efforts.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der Preisgestaltung und dem Risikomanagement in Energiemärkten, insbesondere mit Forwards und Derivaten mit mittel- bis langfristigen Anlagehorizonten in Elektrizität und Erdgas. Der Hauptbeitrag dabei ist der Vorschlag von Modellen zur Preisbestimmung von Energie Forwards und Optionen, sowie für die Erdgasspeicheroptimierung und Ansätze zur Quantifizierung von Modellrisiko in Energiemärkten.

Zuerst schlagen wir ein polynomielles Mehrfaktormodell vor zum Hedgen von langfristigen Elektrizitätsderivaten mit Lieferperioden. Dieses Modell bietet folgende Vorteile: Es berechnet explizit Forwardpreise, Risikoprämien, und die Korrelation zwischen verschiedenen Forwards. Ausserdem lässt sich das Modell durch beobachtete Elektrizitätsforwardkurven präzise und effizient kalibrieren. In diesem Modell schlagen wir einen rollenden Hedge vor, der ausschliesslich liquide Forwardkontrakte nutzt und der risikominimierend ist im Sinne von Föllmer and Schweizer [1991] ist. Wir kalibrieren das Model an einen Datensatz von über acht Jahren deutschen Stromforwardkurven und untersuchen die Qualität des risikominimierenden Hedges für verschiedene Zeithorizonte.

In der Klasse polynomieller Modelle schlagen wir ein quadratisches Gauss-Modell für langfristige Optionen auf Elektrizitätsforwards mit Lieferperioden vor. Dieses Modell kann als eine Unterklasse des polynomielles Modells betrachtet werden, das ein Zweifaktorenmodell verallgemeinert. Dadurch erbt es die Vorteile eines polynomiellen Modells. Zudem erarbeiten wir eine exponentiell-quadratische Transformationsformel, die es erlaubt die charakteristische Funktion durch das Lösen von Riccati-Gleichungen zu berechnen. Somit können wir Optionen effizient mit einem Fourier-Ansatz preisen. Zudem kalibrieren wir das Modell an verschiedenen realen Optionen und Forwardpreise der Deutschen Electriziätsmärkte.

Im weiteren schlagen wir zwei Modelle des maschinellen Lernens für die Erdgasspeicheroptimierung vor, die einen Deep-Hedging-Ansatz von

Buehler, Gonon, Teichmann, and Wood [2019] verwenden. Es sind die ersten Modelle des Deep-Hedging-Ansatzes für dieses Forschungsfeld. Das erste Modell benutzt ausschliesslich den Spot als Hedge und das zweite Modell benutzt zusätzlich Forwards mit monatlichen Lieferperioden. Wir vergleichen den Gewinn/Verlust der Strategien der so trainierten Modelle mit denen eines Monte Carlo Modells der kleinsten Quadrate. Wir sehen, dass das erste Modell dem Monte Carlo Modell nahe kommt und dass das zweite Modell signifikant bessere Renditen liefert bei erhöhter Volatilität.

Schliesslich liefern wir einen umfangreichen Überblick über die Literatur im Bereich Modellrisikoquantifizierung in dem wir Methoden in drei Kategorien einteilen: Den paarweisen Modellvergleich, die Bayes-Modell Durchschnittsbildung und den Worst-Case Ansatz. Wir stellen zwei Anwendungen dafür im Hedgen von langfristigen Energieoptionen vor. In der ersten hedgen wir einen illiquiden Energie-Call mit einem falsch spezifizierten Modell und finden in numerischen Tests, dass das Modellrisiko dem Nachbildungsfehler des Hedges überwiegt und somit nicht vernachlässigbar ist. In der zweiten Anwendung nutzen wir den Entropieansatz von Glasserman and Xu [2014], der explizit das Modellrisiko mit Bezug auf die Divergenz der Entropieterme im Vergleich zum Basismodell angibt. Wir kalibrieren drei verschiedene Modelle an Realdaten, wobei wir eines als Basismodell betrachten. Dann berechnen wir Entropieterme und quantifizieren das Modellrisiko. Dieser Ansatz macht es möglich dem Modellrisiko mit sehr geringem Rechenaufwand eine Zahl zuzuordnen.

Acknowledgement

First and foremost, I would like to express my deepest appreciation to my PhD advisers Martin Larsson and Josef Teichmann for providing me with this great opportunity to work on fascinating research topics. As advisers, they are both full of ideas, deep insights, are extremely good at making connections between different topics with startling creativity, and are highly motivating. Martin taught me how to tackle ad-hoc problems in the industry and to solve them rigorously. Josef adopted me as his PhD student after Martin's departure to CMU, and provided me with very exciting research opportunities. I would like to extend my deepest gratitude to my Master's adviser Martin Keller-Ressel for encouraging me to pursue a PhD and for his recommendation. I wish to thank Axpo Solutions for providing data and funding for my PhD. My deepest appreciation goes to my collaborator Markus Regez from Axpo for all the fruitful discussions and for his expertise in energy markets. I am very grateful to the CRO of Axpo, Lukas Gubler, and to the head of risk management and my collaborator, Vlatka Komaric, for inviting me to talks on model risk and forward modeling on Axpo Risk Day and various other occasions. My sincere gratitude goes to Hanna Wutte and Thomas Krabichler for being wonderful collaborators, and for countless discussions and ideas. I would like to thank Fred Benth for his expertise on energy modeling, and thank Martin Schweizer for his advice on risk-minimizing hedges. I am very grateful to Michel Baes for his deep insights on various topics and am very thankful to Sara Svaluto-Ferro for her expertise on polynomial processes. My special thanks goes to Andrew Allan for proofreading parts of the thesis, and for his advice and for his encouragement. I wish to thank all my office mates, my colleagues and friends in the stochastic finance, insurance mathematics and probability theory group of ETH for all the coffees, aperos, and unforgettable hours inside and outside the office.

Moreover, I want to express my gratitude to all of my family, who constantly supported me in every step of my life and in all my decisions. Finally, I am deeply indebted to my husband Peter Kleisinger, for the love, patience and support he has given me since we first met ten years ago.

Contents

Chapter 1

Introduction

Energy markets is a collective term for all commodity markets which are related to the supply and trade of energy. The major energy commodities traded are crude oil, oil-refined products (cracks such as gas oil, heating oil, jet fuels), natural gas, coal and electricity. Among all energy commodities, electricity and gas stand out due to their distinct nature, and their contract definition, as we will discuss below.

One of the most peculiar properties of electricity is its very limited storability, which makes the classical theory for storable commodities not directly applicable. The main reason is that supply and demand cannot be balanced via storage, and the usual valuation via replication, i.e. the *cash and carry* strategy on buying the spot and storing it for forwards does not hold. Moreover, spot and forward refer to different delivery periods, and are thus essentially different commodities. The lack of storage also causes spikes in prices, which requires advanced methods for modeling. Moreover, together with the limited grid capacity and restrictions on grid connections, it is a regional commodity. Apart from the unstorability, electricity has other specific features such as seasonality, possibility of intra-day negative prices, its unique mechanism in the auction market for the day-ahead trading, and has very liquid short-term markets, and highly illiquid mid-to long-term markets. Another important energy commodity with limited storability is natural gas. It is the less extreme sister of the unstorable electricity, as a small amount of gas can be stored via underground storage facilities or via conversion to liquefied natural gas (LNG), allowing regional and global trading respectively.

However, the proportion of stored gas is small, so that the properties of electricity still apply to gas. As a consequence of the peculiar properties of electricity and natural gas, more tailored solutions are required for their modeling.

Apart from the distinct nature of electricity and gas, another source that contributes to the difficulty in modeling is that the futures and forwards of electricity and gas have a delivery period, that is, a fixed hourly rate delivered throughout the period of delivery. Despite the names, those contracts are in fact swaps, each of which has as reference price an average price, averaged over all spot prices in the delivery period. This significantly increases the complexity, and restricts the choice of models with good mathematical and economical tractabilities for modeling and hedging.

In this thesis, we deal with this interesting and challenging topic of modeling, hedging and risk management of the electricity and gas markets using several approaches. Regarding electricity, we focus on its mid-to long-term markets, as the literature directly addressing these markets is very sparse. Regarding gas, we focus on the machine learning technique for the optimization of a one-year storage schedule for an underground gas storage facility. Throughout the thesis we will refer to electricity and gas as the *energy markets*.

1.1 Outline of the thesis

This thesis is structured as follows: in the remainder of this introductory chapter we provide a brief introduction to the energy markets and a review of quantitative modeling methods, with a particular focus on reduced form models, and a list of notation. In Chapter 2 we review the mathematical tools needed for this thesis, i.e. the polynomial diffusion models. Polynomial diffusion models are very general models which, thanks to the moment formula (see e.g. Cuchiero et al. [2012], Filipović and Larsson [2016]), are very tractable.

Chapter 3 is based on the paper Kleisinger-Yu et al. [2020], published in the SIAM Journal on Financial Mathematics. In this chapter, we propose a multi-factor polynomial framework to model and hedge long-term electricity contracts with delivery period. This framework has several advantages: the computation of forwards, risk premia and correlations between different forwards are fully explicit, and the model can be calibrated to observed electricity forward curves easily and accurately. We suggest a rolling hedge which only uses liquid forward contracts and is risk-minimizing in the sense of Föllmer and Schweizer [1991].

This hedging mechanism allows us to address the non-storability of electricity and the poor liquidity in its long-term markets. Moreover, we calibrate the model to over eight years of German power calendar year forward curves and investigate the quality of the risk-minimizing hedge over various time horizons.

Based on the modeling framework in the previous chapter, we study in Chapter 4 the pricing of options that are written on electricity contracts with delivery period. Here we focus on a subclass of polynomial diffusions, the *quadratic Gaussian models*, as this subclass is a generalization of the two-factor model of Chapter 3, which captures the features of electricity contracts with delivery period accurately. For the quadratic Gaussian framework, we propose a Fourier-style option pricing approach, for which we have developed an exponential-quadratic transform formula to compute the characteristic function by solving Riccati equations. The proof of the theorem is inspired by the proof of extension of the affine transform formula (Filipovic [2009]) The strength of this modeling framework is that it provides a consistent and tractable way of pricing options and the underlying forwards simultaneously. Moreover, we conduct a calibration study where we calibrate our model to volatility curves and their underlying forwards of the German electricity exchange market (EEX) and investigate the quality of fit.

In Chapter 5, we turn our attention to the short-to mid-term markets of gas, and propose two models for the optimization of gas storage which uses the deep hedging approach developed by Buehler et al. [2019]. To the best of the author's knowledge, our models are the first deep learning applications in this field: the first model (Model I) is of the intrinsic valuation type, and solves the gas storage optimization using only day-ahead gas forwards as hedging instruments; the second model (Model II) extends Model I by additionally using the monthly gas forwards with delivery period. Note that in reality, one uses the monthly forwards with delivery periods as hedging instruments, and not an artificial computed daily forward curve. And thus, Model II is useful for the real world application. Deep hedging is machine learning approach. In particular it is a *supervised learning* approach inspired by the concept of reinforcement learning: given a time series of price dynamics as inputs, one can train neural networks to compute the hedging strategy or the storage schedule at any point in time to minimize according to a loss function that incorporates the inputs and outputs (normally set to 0). This approach to solving the gas storage problem allows us to take advantage of the recent advances in computer science, the Keras and Tensorflow modules in Python with very efficient gradient descent methods.

Thus, we conduct numerical tests, both in-sample and out-of-sample, to compare the performance of both models with a benchmark case computed using Least Squares Monte Carlo.

In Chapter 6, we study the quantification of model risk, with a strong focus on methods that are applicable to the industry and in particular to long-term energy markets. We discuss various aspects and definitions of model risk, and provide an extensive review of the model risk quantification literature, in which we group the applicable methods into three categories: the pairwise model comparison (and parameter perturbation), the Bayesian model averaging, and worst-case approach. Moreover, we present two applications to quantify model risk in the modeling and hedging of long-term energy options. In Application I, we hold an illiquid energy call which we hedge using a misspecified model. The discrepancy from the value of hedge and the energy call consists of two errors, namely the model misspecification error (model risk), and the model-based tracking error of the hedge. In our simulation study, in which the Heston model is assumed to describe the real dynamics of electricity, and the Black-Scholes model the misspecified model, we investigate and quantify the model risk and the tracking error. In Application II, we quantify the model risk using the entropy approach proposed by Glasserman and Xu [2014]. The entropy approach is an worst-case approach, which only requires a Monte Carlo estimator of the claim on the baseline model and the divergence from the baseline model in entropy level. In particular, we don't need the knowledge of all models within this entropy level, nor the prices of the claim under each of the models. It is computationally very efficient. Hence we investigate the model risk of the volatility curve of the calendar year 2023 German call options with respect to different entropy levels, which we obtain by calibrating three different models to the data and computing their divergence to a baseline model.

1.2 A brief introduction to energy markets

Since the start of the liberalization of the electricity and gas markets in the nineties, numerous markets for spot and derivative products in various regions have emerged, attracting the interest of various financial investors besides the traditional market player who possesses energy producing facilities. In the following we give a brief introduction to electricity markets.

Electricity has a lot of distinct properties that make it unique. There exist

various commodities that can be used to produce power, e.g. coal, gas, nuclear, wind, solar. The composition of power generation varies from region to region, and depends heavily on the geographical properties and the geopolitical policy of each country or region. However once generated, electricity does not differ in quality, contrary to other energy commodities such as oil, which has different composites: e.g. West Texas Intermediate (WTI) is "sweeter" than Brent due to the sulfur content. Electricity can then be transmitted through the grid to reach end users for both commercial and private purposes. Its usage has various facets, and ranges from enabling production in a factory to lighting and heating of a single household.

The most predominate feature of electricity is its very limited storability, it is thus generally considered a non-storable commodity. It is both very difficult and expensive to store power on a significant scale. Energy producers who own power plants can store electricity in its primary generation commodities such as water reservoirs (for hydro-based electricity production), gas, oil and coal (thermal electricity production). The energy consumers cannot however buy storage, and thus there is no cost of carry relationship between spot and forward as is known in other commodities and asset classes. Because of the lack of storage, Kirchhoff's law applies, that is, the production must equal consumption at all time. Any deviance from equilibrium of supply and demand cannot be balanced through storage, and causes the price to spike, both upwards and downwards. The occurrence of an unexpected shortfall of electricity, for example through the shut-down of a power plant due to sudden breakdown, can immediately cause a price spike, where the price reaches a magnitude of 10-times or even 100-times the equilibrium price. Similarly, whenever there is a supply surplus, the producer will be offered a price for not pumping their electricity generation into the grid, creating a negative price of electricity. As a consequence, the volatility of real-time electricity prices is very high. In general, the volatility of electricity depends on the exact delivery period. The volatility of electricity contracts depends on the remaining time to the start of delivery, i.e. time to maturity. One often observes the so called Samuelson effect, namely the volatility is decreasing with increasing time to maturity.

Another important property of electricity is its seasonality. Because of the lack of storage, the demand side tends to drive the prices. It has a daily pattern, i.e. peak loads and off-peak loads (after work hours delivery, without demand from e.g. factories), a weekly pattern, i.e. weekday loads and weekend loads, and monthly and quarterly patterns, i.e. winter demand and summer demand.

Moreover, electricity is a regional commodity, and it is often not possible to take advantage of cross-regional arbitrage. The reason, besides the unstorability of electricity, is that the grid infrastructure is not always well integrated and the grid capacity is very limited. There is no global market for electricity.

Electricity as a flow commodity is continuously delivering power for local use. This is reflected by its tradable futures and forwards. The terminology seems misleading, as all forwards, futures (and options) of electricity have delivery periods, meaning that the contracts refer to an average-based price averaged over the period of delivery. Therefore, despite being called futures and forwards, these electricity contracts are swaps in nature, where the fixed leg is determined prior to the delivery period (via the market), and the variable leg is then the spot price or real price of electricity delivery. The existence of the swap-like structure adds quickly to the complexity of modeling, making some models (e.g. models of multiplicative exponential affine type) less tractable than others (e.g. models of additive factors) for computation.

Depending on the time to the start of the delivery period, one can distinguish between four market categories:

- The *forward / futures market*, where the power producers sell their production on a forward basis. The traded forward products have delivery periods, which range from single day to weeks, months, quarters and calendar years.
- The *day-ahead market / spot market*, where each of the delivery hours of the following day are auctioned by exchanges today. Typically market participants submit bids in volumes and prices to the exchange till noon local time for the following day.
- The *intra-day market*, which is often organized by exchanges. Here single hours are traded for which the spot market has already been settled.
- The *adjustment / reserve / balancing market* is typically organized by the local TSO (Transmission System Operator), who is reponsible for the grid and its stability of a market place. In this market, the TSO auctions the right to provide flexible supply (or demand) in order to prevent electricity shortage for the consumers and the damage of the transmission grid through oversupply.

A visualization is provided by Figure 1.2.1. Let τ denote the time to maturity. All markets with $\tau \leq 1$ year are considered short-term markets, with $1 \leq \tau \leq 3$ years

mid-term markets, and with $\tau \geq 3$ years long-term markets. Electricity has highly liquid short-term markets and illiquid mid-to long-term markets. In this thesis, we focus on the mid- to long-term forwards / futures markets. That is we study calendar-year contracts with a yearly delivery period, with time to maturities ranging from one year to ten years.

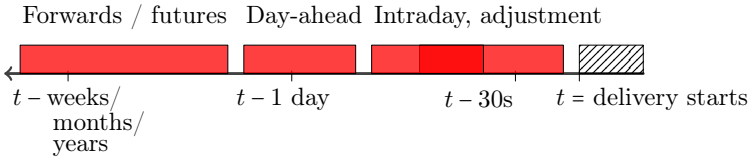


Figure 1.2.1: Four categories of electricity markets with respect to the to the time to the start of delivery period.

In terms of modeling, the price dynamics of electricity shows mean-reversion. It is particularly visible in the very short-term markets with large spikes: one sees that an upward spike is followed shortly after by a downward spike with almost the same magnitude, expressing a mean-reversion effect towards the usual equilibrium level. Similarly yet more subtly, it also applies to markets with longer time to maturity, where the short-term prices fluctuate around and mean-revert towards the long-term level which is determined through economic fundamentals (the outlook of demand and supply of electricity of a region or a country).

Another interesting feature is that the term structure of electricity forward curve requires multi-factor modeling. Empirical studies indicate that a large number of factors is needed. To compare, a model with three factors (e.g. shift, tilt and bend) is usually enough to explain around 95 – 98% of the underlying dynamics of the interest rate term structure; see e.g. Steeley [1990], Litterman and Scheinkman [1991], Dybvig [1988]; a model with three factors can only explain around 70% of the electricity forward term structure; see Koekebakker and Ollmar [2005], Frestad [2008], Benth et al. [2008a].

Natural gas is an important fuel for heating and also for electricity generation. It has limited storability, but contrary to electricity is not completely unstorable for consumers and investors. For example, natural gas can be stored in pipelines, underground storage facilities and converted to LNG. The most important hubs for gas trading are the Henry Hub, located in Louisiana in US, and the National

Balancing Point (NBP) in the UK. Similarly to electricity, Natural gas has price spikes, seasonality, intra-day negative prices, Samuelson effect of volatility and multi-factor nature of term structure of forwards and futures. But the effects are less pronounced than electricity due to some storage capacity. Moreover, its conversion to LNG makes it possible to take advantage of cross-regional or global price arbitrage. The contract definition of natural gas is similar to that of electricity, and thus we omit an extensive discussion of natural gas markets.

1.3 Modeling of energy markets

There are several ways to categorize the modeling of energy markets. We follow the classification of Carmona and Coulon [2014], and split the energy models into the structured models and the reduced form models. Structured models are those models that take some simplified fundamental relationships between price and demand, capacity and/or marginal cost curves, into consideration; see e.g. Carmona and Coulon [2014] for a review of structural models. Reduced form models incorporate stylized properties of spot and futures / forwards and are used mostly for derivative pricing (forward with delivery period, options); see e.g. Benth, Benth, and Koekebakker [2008a] for in-depth treatment. In the following we briefly review the reduced-form modeling of spot and futures / forwards of energy markets, as the model we propose in this thesis is a reduced form model. For the treatment below, we will use forward as the collective term for futures and forwards. Throughout this chapter, we fix a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$.

Much of the energy modeling approach comes from the modeling of interest rate market, and is modified to account for the stylized properties of the energy markets. Most literature concentrates on the modeling of short-term markets, as these liquid markets provide huge amounts of data, and are ideal for time-series analysis.

The modeling of energy markets can also be split into three different tasks: spot price modeling, derivation or modeling of forwards, and pricing of options. Spot modeling S_t plays a central role in the author's view for two reasons. Firstly, the model describing the stochastic dynamics in the spot price is of interest for energy traders. Secondly and most importantly, it is the reference price for the settlement of a forward with delivery period, and such a forward is the main interest of this thesis. Regarding the spot price dynamics, it is standard to

use Ornstein–Uhlenbeck process (OU), as this is a natural way of describing mean-reversion in the price dynamics.

An energy forward is in fact a swap, as it has a delivery period, and the reference price is usually an average price, averaged over all spot prices within the period of delivery. In this sense, an energy forward is already a derivative[1]. Formally, let $F(t, T_1, T_2)$ denote the time-t price of a forward with delivery in $[T_1, T_2)$. For convenience, we use a continuous setting, so that

$$F(t, T_1, T_2) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} f(t, T) dT, \tag{1.3.1}$$

where $f(t, T)$ denotes a forward with an instantaneous delivery at time T. Note that contrary to forwards from other asset classes, $f(t, T)$ is purely a computational forward and is not tradable.

For energy markets, it is difficult but essential to establish a relationship between spot and forward. Below we discuss the forward and spot modeling in two different subsections. In the first subsection, we will discuss in detail three different approaches for forward modeling: the first two approaches build on spots and rely on some spot-forward relationship; the last approach builds the model for the forward directly. In the second subsection we discuss in some detail the spot modeling.

Forward modeling – does a spot-forward relationship exist?

In the following, we briefly discuss three different forward modeling approaches, with a particular focus on their unique arguments for building the spot-forward relationships. For a storable commodity, say oil, the notion *convenience yield* δ is introduced to explain the difference between the spot S_t and the computational forward $f(t, T)$. On the one hand, a storable commodity needs to be transported, stored, insured, additionally to the cost of financing the purchase; therefore, the cost of storage is positive, and in absence of arbitrage, we have a *"traditional contango"*:

$$S_t \leq f(t, T).$$

On the other hand, holding a storable commodity can be highly profitable in times of demand spikes or supply shortages; therefore, in absence of arbitrage,

[1] An option on an energy forward is even more exotic, namely a swaption.

we should have a *"traditional backwardation"* :

$$S_t \geq f(t,T).$$

Formally the convenience yield δ is defined as:

$$\begin{aligned} f(t,T) &= S_t e^{(\mu-\delta)(T-t)} \\ &= S_t e^{\mu(T-t)} \cdot \underbrace{e^{-\delta_1(T-t)}}_{\text{benefit from owning physical}} \cdot \underbrace{e^{c(T-t)}}_{\text{cost of storage}}, \end{aligned}$$

where $\delta := \delta_1 - c$ and μ is the risk-neutral interest rate. Then the relationship can be established via:

$$S_t = \lim_{t \to T} f(t,T), \tag{1.3.2}$$

and (1.3.1).

This concept is popular and has been predominant for a very long time in the classical literature for all energy commodities. The most famous model using the convenience yield is the Gibson–Schwartz model, Gibson and Schwartz [1990]. The log spot price $X_t = \log(S_t)$ evolves according to the following SDE:

$$\begin{aligned} dX_t &= \Big(\mu_t - \delta_t - \frac{1}{2}\sigma^2\Big)dt + \sigma dW_t^X \\ d\delta_t &= \kappa(\theta - \delta_t)dt + \sigma_\delta dW_t^\delta \end{aligned} \tag{1.3.3}$$

with $d\langle W^X, W^\delta\rangle_t = \rho dt$. This famous two-factor model falls into the class of the exponential affine models. The seasonality is incorporated by using time-dependent parameters μ_t. The forward $f(t,T)$ is given by

$$f(t,T) = S_t e^{\int_t^T r_s ds} e^{B(t,T)\delta_t + A(t,T)}$$

where $B(t,T)$ and $A(t,T)$ are solutions of some Riccati-equations. The drawback of this model is that it requires δ_t to be independent of contract maturity T; this is contradictory to the Samuelson effect which implies that the forward volatility is higher than the spot volatility; and the most important drawback is that this theory assumes the use of storage, which contradicts the unstorability of electricity and gas.

An alternative way of explaining the relationship is through the use of a market price of risk $\lambda : \mathbb{R}^d \to \mathbb{R}^d$. We denote the associated Radon–Nikodym density process by

$$M_t^\lambda = \exp\Big(\int_0^t \lambda^\top dW_s - \frac{1}{2}\int_0^t \|\lambda\|^2 ds \Big),$$

and assume that M_t^λ is a true martingale. We can then define $\mathbb{Q}$ on every finite time interval $[0,T]$ via its Radon–Nikodym density $\frac{d\mathbb{P}}{d\mathbb{Q}}|_{\mathcal{F}_T} = M_T^\lambda$. The risk premium $R(t,T)$ then defines the difference between forward and spot, and is given by

$$\begin{aligned} R(t,T) &:= f(t,T) - \mathbb{E}_\mathbb{P}[S_T|\mathcal{F}_t] \\ &= \mathbb{E}_\mathbb{Q}[S_T|\mathcal{F}_t] - \mathbb{E}_\mathbb{P}[S_T|\mathcal{F}_t]. \end{aligned}$$

For $F(t,T_1,T_2)$ the risk premium can be defined simply via integration over $[T_1,T_2)$:

$$R(t,T_1,T_2) := \frac{1}{T_2-T_1}\mathbb{E}_\mathbb{Q}\left[\int_{T_1}^{T_2} S_u\,du \mid \mathcal{F}_t\right] - \frac{1}{T_2-T_1}\mathbb{E}_\mathbb{P}\left[\int_{T_1}^{T_2} S_u\,du \mid \mathcal{F}_t\right].$$

This concept is related to the rational expectation hypothesis in interest rate theory. Using this notion, $R(t,.) < 0$ refers to the *"normal backwardation"*; it describes a situation in which a commodity producer wishes to hedge his revenues by selling forwards, and is therefore willing to pay a premium on the expected spot price; $R(t,.) > 0$ refers to the *"normal contango"*. This concept has also become very popular. A very famous model using the market price of risk is a *short-term/long-term model*, proposed by Schwartz and Smith [2000], Lucia and Schwartz [2002]. Let Λ_t denote the seasonality. In this model, the deseasoned log spot price $X_t = \log S_t - \log \Lambda_t$ is modeled (under $\mathbb{P}$) as:

$$\begin{aligned} X_t &= X_t^{(1)} + X_t^{(2)}, \\ dX_t^{(1)} &= -\kappa_1 X_t^{(1)} dt + \sigma_1 dW_t^{(1)}, \\ dX_t^{(2)} &= \kappa_2 dt + \sigma_2(\rho dW_t^{(1)} + \sqrt{1-\rho^2} dW_t^{(2)}), \end{aligned} \tag{1.3.4}$$

with $d\langle W^{(1)}, W^{(2)}\rangle_t = \rho dt$. The factors $X_t^{(1)}$ and $X_t^{(2)}$ describe the short-term deviation and the long-term equilibrium price levels respectively. Let κ_χ denote the rate at which the short-term deviations are expected to disappear. With a market price of risk λ, one can perform an equivalent measure change to obtain the $\mathbb{Q}$-dynamics of the spot. Moreover, the initial forward price is given by $F(0,T) = \mathbb{E}_\mathbb{Q}[S_T] = e^{-\kappa_\chi T}\chi_0 + \xi_0 + A(T)$, with some deterministic function A. The most notable strength of this method is that even though the convenience yield isn't directly modeled, it can be shown that the model is equivalent to the Gibson–Schwartz model in (1.3.3); see Lucia and Schwartz [2002]. This fact contributes considerably to the popularity of this type of model. In the following years, many extensions of this model were proposed.

An alternative modeling philosophy is to build the forwards directly, instead of building them based on the spot, via the spot-forward relationship, and forward with instantaneous delivery at a single point in time. In concrete, this adopts the Heath–Jarrow–Morton (HJM) approach from the interest rate theory. If needed, spot can be implied by the forward via (1.3.2) as a purely computational instance. An energy model of the HJM type was proposed by Bjerksund, Stensland, and Vagstad [2011]. In their model, the dynamics of the forward curve is represented by:

$$\frac{d\,f(t,T)}{f(t,T)} = \sum_{i=1}^{N} \sigma_i(t,T)dW_i(t)$$

where $d\langle W_i, W_j\rangle_t = 0$ for $i \neq j$. Note that $\sigma_i(t,T) = \sigma_i(T-t)$, $i = 1,\ldots,N$, neglecting the seasonality in volatility terms. Their PCA analysis on the British gas data (the NBP data) indicate that $N = 6$. Forwards can be computed using e.g. an Euler-scheme simulation of the following:

$$f(t,T) = f(0,T)\exp\left(\sum_{i=1}^{N}\left[\int_0^t \sigma_i(T-u)dW_i(u) - \frac{1}{2}\int_0^t \sigma_i^2(T-u)du\right]\right).$$

In this model, the modeling the swap-like forwards as defined in (1.3.1) is a challenging task. And thus, one often uses the simplifying assumption that

$$F(t,T_1,T_2) = f(t,T_1).$$

Spot modeling

In the following we briefly discuss spot modeling, where we focus on modeling the deseasoned and detrended spot curves. According to Benth et al. [2008a], the modeling of energy spots can be divided into geometric and arithmetic models. In geometric models $\log(S_t)$ is modeled as the sum of OU processes; this is a classical approach, where a log-normal distribution is assumed for the energy spot. In arithmetic models, S_t is directly modeled as the sum of OU-processes; these models assume that energy spot is normally distributed.

To formalize, for $i = 1,\ldots,m$, $j = 1,\ldots,n$, we let

$$\begin{aligned} dX_i(t) &= [\mu_i(t) - \alpha_i(t)X_i(t)]dt + \sum_{k=1}^{p}\sigma_{ik}(t)dW_k(t), \\ dY_j(t) &= [\delta_j(t) - \beta_j(t)Y_j(t)]dt + \eta_j(t)dI_j(t), \end{aligned}$$

be stochastic and jump factors respectively, where W_k, $k = 1, \dots, p$, are independent BMs and I_j are independent jump components. Stochastic factors have a good ability to fit the spot underlying the long-term forward prices, and the jump factors are good at capturing possible spikes in short-term prices.

In a geometric model, the spot S_t is defined as:

$$\log S_t = \log \Lambda(t) + \sum_{i=1}^{m} X_i(t) + \sum_{j=1}^{n} Y_j(t),$$

where $\Lambda(t)$ denotes the seasonality component. In an arithmetic model, the spot S_t is defined as:

$$S_t = \Lambda(t) + \sum_{i=1}^{m} X_i(t) + \sum_{j=1}^{n} Y_j(t).$$

The first and simplest geometric model was proposed by Schwartz [1997], where the log spot price evolves according to the single factor:

$$dX_t = \kappa\Big(\mu_t - X_t - \frac{\sigma^2}{2\kappa}\Big)\, dt + \sigma dW_t.$$

In this model, there is no explicit modeling of the spot-forward relationship such as convenience yield. Yet, because of its simplicity, it is frequently used as the benchmark model for oil, gas and electricity. In particular, it is often used as the model for gas prices in the gas storage optimization literature; see e.g. Chen and Forsyth [2008], Boogert and De Jong [2008], Bjerksund et al. [2011].

To account for the spot-forward relationship, the Gibson–Schwartz model (1.3.3) and the Lucia–Schwartz model (1.3.4) were introduced. The latter model became very popular, was extended several times (among others) by Villaplana [2003] and Barlow, Gusev, and Lai [2004].

Arithmetic models can be built easily based on the underlying dynamics of all geometric models mentioned above, i.e. the factors X_t^i should be used for modeling S_t instead of $\log(S_t)$. This can be very useful when modeling and hedging forward with delivery period, that are implied by the spot, as the arithmetic models (working with the normal distribution) have better tractability and calibration efficiency compared to arithmetic models. We omit an extensive discussion on each of the models; see Benth et al. [2008a] for a detailed treatment.

Finally, we want to point out that our model in Chapter 3 is closest to an arithmetic model and it is tractable for modeling and hedging long-term energy forwards with delivery period. In contrast to the arithmetic model, which can

have negative prices and prices are unbounded, our model generates non-negative prices, or prices bounded from below by a number $c \in \mathbb{R}$.

Seasonality

In energy markets, prices are highly dependent on the exact delivery period. Thus, if we compare contracts with the same delivery length but different delivery periods, that is, different subperiods of a year, it is important to first adjust for seasonality before making a reasonable comparison.

For electricity modeling in this thesis, we only consider calendar year baseload forwards with delivery period. All those contracts deliver throughout the year and not only for a specific sub-period of the year. To capture these forwards in mid-to long-term markets, it is not necessary to explicitly model seasonality. For the gas storage optimization treated in this thesis, our focus is on the optimization problem using machine learning and not on the modeling tasks of gas prices. Thus, we don't model seasonality for gas.

Nevertheless, we briefly give a review on common approaches used in seasonality modeling. Seasonality can be either modeled as a deterministic component, denoted Λ_t, or a stochastic component $\Lambda_t(X_t)$ which depends on the underlying stochastic dynamics X_t. The former deterministic approach is the most popular technique used. However, the exact form of the seasonality function always depends on the market. To capture the lower prices in summer and higher prices in winter, a sinusoidal function like the cosine function can be a suitable choice; see e.g. Pilipovic [2007]. Benth and Šaltytė-Benth [2004] proposed the use of a continuous seasonal floor: $\Lambda_t = a_0 + a_1 t + a_2 \sin(2\pi(t - a_3)/M,\ M = 250, 365$. A different ansatz with a focus on the monthly seasonalities throughout the year, suggests to use 12 dummy variables representing the lower seasonal level of each of the 12 months of the year. These dummy variables can be represented via a piece-wise or step function Λ_t, $(\sum_{i=1}^{12} \log(\Lambda_{i/12}) = 0)$, in order to approximate the periodic components and to incorporate them in the implementation stage of the models; see e.g. Jaillet et al. [2004], Knittel and Roberts [2001] for more details.

1.4 List of notation

Throughout this thesis, we fix a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. Let $\mathbb{Q}$ denote an equivalent risk-neutral probability measure that we use for the pricing. For simplicity we assume zero interest rate and thus apply no discounting. We denote by $\mathbb{S}^d$ the set of all symmetric $d \times d$ matrices and $\mathbb{S}_+^d$ the subset consisting of positive semidefinite matrices. Let $\tilde{\alpha}$ denote a multi-index, that is $\tilde{\alpha} := (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_d) \in \mathbb{N}_0^d$ with $|\tilde{\alpha}| = \tilde{\alpha}_1 + \ldots + \tilde{\alpha}_d$. For $x \in \mathbb{R}^d$, let $x^{\tilde{\alpha}}$ be given by

$$x^{\tilde{\alpha}} := x_1^{\tilde{\alpha}_1} \cdot x_2^{\tilde{\alpha}_2} \cdots x_d^{\tilde{\alpha}_d}.$$

Let p denote a polynomial on $\mathbb{R}^d$, i.e.

$$p : \mathbb{R}^d \to \mathbb{R}, \qquad p(x) = \sum_{\alpha \in \mathbb{N}^d} c_\alpha x^\alpha,$$

with only finitely many of $c_\alpha \in \mathbb{R}$ that are non-zero. The degree of p is the total degree of the multivariate polynomial, and is given by:

$$\deg(p) = \max\{|\alpha| : c_\alpha \neq 0\}.$$

If p is the zero polynomial, then $\deg(p) = \infty$. Further we let Pol denote the space containing all polynomials on $\mathbb{R}^d$, and let

$$\mathrm{Pol}_n := \{p \in \mathrm{Pol} : \deg(p) \leq n\}$$

denote the subspace of polynomials with degree at most n.

Chapter 2

A review of polynomial diffusions

2.1 Introduction

Polynomial models are highly popular in financial modeling, see e.g. Filipović et al. [2017], Ackerer and Filipovic [2017], Ackerer et al. [2018], Cuchiero [2018], Filipović and Willems [2018], Ackerer and Filipović [2016], Filipović et al. [2016], Biagini and Zhang [2016], Delbaen and Shirakawa [2002] for references as well as Cuchiero et al. [2012], Filipović and Larsson [2016] for a treatment of the underlying mathematical theory.

In this chapter, we review the basic definition and the main properties of a polynomial diffusion, which provides the mathematical tools needed for Chapter 3 on modeling and hedging electricity forwards with delivery period as well as for Chapter 4 on the option pricing based on those forwards. This chapter is structured in the following way: Section 1 we give definition of a polynomial diffusion. The most important property that makes it popular is the moment formula, which we give in Section 2. As a preparation for our modeling chapters, we also provide a moment formula for polynomials of degree two. Section 3 reviews the conditions for the existence of exponential moments, and for the uniqueness in law. The existence of a polynomial diffusion and its boundary attainment of a subspace of $\mathbb{R}^d$ is then reviewed in Section 4. As jumps are

not needed for the modeling in this thesis, we briefly discuss the properties of a polynomial jump diffusions at the end of the chapter; for more details see e.g. Filipović and Larsson [2020]. The notation of this chapter is very close to that of Filipović and Larsson [2016].

Throughout the thesis, we let $E \subset \mathbb{R}^d$ denote a state space and assume that E has non-empty interior. This will cover all applications of polynomial diffusions in this thesis. Moreover, it allows us to identify polynomials on $\mathbb{R}^d$ with polynomials on E.

2.2 Definition

Let $a : \mathbb{R}^d \to \mathbb{S}^d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$, such that for all i,j:

$$a_{ij} \in \mathrm{Pol}_2 \quad \text{and} \quad b_i \in \mathrm{Pol}_1. \tag{2.2.1}$$

Consider the following SDE:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{2.2.2}$$

where W is a d-dimensional Brownian motion and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is continuous, and $a(x) = \sigma(x)\sigma(x)^\top$. The associated partial differential operator $\mathcal{G}$ is given by

$$\mathcal{G}f(x) = \frac{1}{2}\mathrm{Tr}(a(x)\nabla^2 f(x)) + b(x)^\top \nabla f(x) \tag{2.2.3}$$

for $x \in \mathbb{R}^d$ and any C^2 function f. By Itô's formula, the process

$$f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_u)du \tag{2.2.4}$$

is a local martingale. Note that due to (2.2.1), for any $n \in \mathbb{N}$ and any polynomial $p \in \mathrm{Pol}_n$, $\mathcal{G}p$ is also polynomial of the same degree or lower degree, i.e. $\mathcal{G}p \in \mathrm{Pol}_n$.

Definition 2.2.1 (Polynomial diffusion, Definition 2.1 of Filipović and Larsson [2016])**.** *The operator $\mathcal{G}$ is called* polynomial *if it maps Pol_n to itself for each $n \in \mathbb{N}$. In this case, we call any E-valued solution to (2.2.2) a* polynomial diffusion *on E.*

In fact, any $\mathcal{G}$ that maps Pol_n to itself for any n is of the form (2.2.2) and (2.2.3).

Lemma 2.2.2 (Lemma 2.2 of Filipović and Larsson [2016]). *Let* $\widetilde{\mathcal{G}}f(x) = \frac{1}{2}Tr\left(\widetilde{a}(x)\nabla^2 f(x)\right)+\widetilde{b}(x)^\top \nabla f(x)$ *be a partial differential operator for some maps* $\widetilde{a}:\mathbb{R}^d \to \mathbb{S}^d$ *and* $\widetilde{b}:\mathbb{R}^d \to \mathbb{R}^d$. *Then the following are equivalent:*

(i) $\widetilde{\mathcal{G}}$ *maps* Pol_n *to itself for each* $n \in \mathbb{N}$

(ii) $\widetilde{\mathcal{G}}$ *maps* Pol_n *to itself for each* $n \in \{1,2\}$

(iii) *The components of* $\widetilde{a}$ *and* $\widetilde{b}$ *lie in* Pol_2 *and* Pol_1 *respectively.*

In this case, $\widetilde{a}$ *and* $\widetilde{b}$ *are uniquely determined by the actions of* $\widetilde{\mathcal{G}}$ *on* Pol_2.

Proof. The implication (ii) to (iii) follows by applying $\mathcal{G}$ to monomials of degree one and two. Other implications, (i) to (ii) and (iii) to (i) is obvious. □

2.3 Moment formulas

In this section, we first review the general version of the moment formula of Filipović and Larsson [2016], and then provide a moment formula for polynomials of degree at most two. Both versions are used in Chapter 3 and Chapter 4.

Fix n and let $N = \binom{d+n}{n}$ be the dimension of Pol_n. Let $H : \mathbb{R}^d \to \mathbb{R}^N$ be a function whose components form a basis of Pol_n. As an example, monomials can be used to form a basis. Then for any $p \in \mathrm{Pol}_n$

$$p(x) = H(x)^\top \vec{p}, \tag{2.3.1}$$

$$\mathcal{G}p(x) = H(x)^\top G \vec{p}, \tag{2.3.2}$$

where $\vec{p} \in \mathbb{R}^N$ is the coordinate representation of $p(x)$, and $G \in \mathbb{R}^{N\times N}$ the matrix representation of the generator $\mathcal{G}$.

Theorem 2.3.1 (Moment formula - general version). *Let* $\mathbb{E}[\|X_0\|^{2n}] < \infty$ *and let* p *be a polynomial with coordinate representations* (2.3.1)–(2.3.2). *Further let* X_t *satisfy* (2.2.2). *Then for* $0 \le t \le T$ *we have:*

$$\mathbb{E}_{\mathbb{Q}}[p(X_T)|\ \mathcal{F}_t] = H(X_t)^\top e^{(T-t)G}\vec{p}. \tag{2.3.3}$$

Proof. It is essential to show that the left side is not only a local martingale but a true martingale. See Theorem 3.1. of Filipović and Larsson [2016]. □

Below we give a more explicit version of the moment formula for polynomials of degree two, which is useful for the next two chapters. Here we slightly change the formulation of the SDE (2.2.2) to highlight the mean-reversion effect of the underlying price dynamics of energy markets,i.e. $b(x) := \kappa(\theta - x)$. We also assume deterministic initial values. In detail, we consider

$$dX_t = \kappa(\theta - X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x, \tag{2.3.4}$$

where $\kappa \in \mathbb{R}^{d\times d}$, $\theta \in \mathbb{R}^d$. Moreover, note that the quantity $\mathrm{Tr}(\pi\, a(x))$ with $\pi \in \mathbb{S}^d$ is quadratic in x, and thus of the form

$$\mathrm{Tr}(\pi\, a(x)) = a_0(\pi) + a_1(\pi)^\top x + x^\top a_2(\pi)\, x. \tag{2.3.5}$$

for some $a_0(\pi) \in \mathbb{R}$, $a_1(\pi) \in \mathbb{R}^d$, and $a_2(\pi) \in \mathbb{S}^d$ that depend linearly on π.

Theorem 2.3.2 (Moment formula for polynomials of degree two, Theorem 3.2 in Kleisinger-Yu et al. [2020])**.** *Let $q(x)$ be a polynomial of the form $q(x) = q_0 + \vec{q}^\top x + x^\top Q x$ with $q_0 \in \mathbb{R}$, $\vec{q} \in \mathbb{R}^d$ and $Q \in \mathbb{S}^d$. Further let X_t satisfy (2.3.4). Then for $0 \le t \le T$ we have:*

$$\mathbb{E}_{\mathbb{Q}}[\; q(X_T) \mid \mathcal{F}_t] = \phi(T-t) + \psi(T-t)^\top X_t + X_t^\top \pi(T-t) X_t,$$

where ϕ, ψ, π solve the linear ODE

$$\begin{aligned} \phi' &= \psi^\top \kappa\theta + a_0(\pi), & \phi(0) &= q_0, \\ \psi' &= -\kappa^\top \psi + 2\pi\kappa\theta + a_1(\pi), & \psi(0) &= \vec{q}, \\ \pi' &= -\pi\kappa - \kappa^\top \pi + a_2(\pi), & \pi(0) &= Q. \end{aligned} \tag{2.3.6}$$

Proof. Define

$$M(t, X_t) := \phi(T-t) + \psi(T-t)^\top X_t + X_t^\top \pi(T-t) X_t.$$

Let $\tau = T - t$. Itô's formula along with (2.3.5) and then (2.3.6) gives:

$$\begin{aligned}
dM(t, X_t) &= -\phi'(\tau)dt - \psi'(\tau)^\top X_t dt - X_t^\top \pi'(\tau) X_t dt + \psi(\tau)^\top dX_t \\
&\quad + 2X_t^\top \pi(\tau) dX_t + \frac{1}{2} \cdot 2\mathrm{Tr}(\pi(\tau) d\langle X \rangle_t) \\
&= \Big(-\phi'(\tau) - \psi'(\tau)^\top X_t - X_t^\top \pi'(\tau) X_t + \psi(\tau)^\top \kappa\theta - \psi(\tau)^\top \kappa X_t \\
&\quad + 2\theta^\top \kappa^\top \pi(\tau) X_t - 2X_t^\top \pi(\tau)\kappa X_t + \mathrm{Tr}(\pi(\tau) a(X_t)) \Big) dt \\
&\quad + \widehat{\sigma}(t, X_t) dW_t \\
&= \Big(\big[-\phi'(\tau) + \psi(\tau)^\top \kappa\theta + a_0(\pi(\tau)) \big] \\
&\quad + \big[-\psi'(\tau) - \kappa^\top \psi(\tau) + 2\pi(\tau)\kappa\theta + a_1(\pi(\tau)) \big]^\top X_t \\
&\quad + X_t^\top \big[-\pi'(\tau) - \pi(\tau)\kappa - \kappa^\top \pi(\tau) + a_2(\pi(\tau)) \big] X_t \Big) dt + \widehat{\sigma}(t, X_t) dW_t \\
&= \widehat{\sigma}(t, X_t) dW_t,
\end{aligned}$$

where $\widehat{\sigma}(t, x) := (\psi(\tau) + 2\pi(\tau)x)^\top \sigma(x)$. Thus, $M(t, X_t)$ is a local martingale. Now we let $C \in \mathbb{R}$ be a constant such that $\|a(x)\|_{\mathrm{op}} \le C(1 + \|x\|^2)$. Then with Cauchy-Schwartz inequality,

$$\begin{aligned}
\|\widehat{\sigma}(t, X_t)\|^2 &\le \|\psi(T - t) + 2\pi(T - t)X_t\|^2 \|a(X_t)\|_{\mathrm{op}} \\
&\le \widetilde{C}(1 + \|X_t\|^4),
\end{aligned}$$

for some constant $\widetilde{C} \in \mathbb{R}$. Together with Tonelli's theorem, this bound yields

$$\mathbb{E}\left[\int_0^T \|\widehat{\sigma}(t, X_t)\|^2 dt \right] \le \widetilde{C} \int_0^T \mathbb{E}\left[1 + \|X_t\|^4 \right] dt,$$

which is finite by Theorem 2.3.1. Hence, $M(t, X_t)$ is a square-integrable true martingale. As a result,

$$M(t, X_t) = \mathbb{E}[M(T, X_T)|\mathcal{F}_t] = \mathbb{E}[q_0 + \vec{q} X_T + X_T^\top Q X_T | \mathcal{F}_t] = \mathbb{E}[q(X_T)|\mathcal{F}_t].$$

This is the claimed formula. □

2.4 Uniqueness

The moment formula, Theorem 2.3.1, shows that the mixed moments of a polynomial diffusion are uniquely determined by its generator. Thus the

uniqueness of a polynomial diffusion follows whenever moments exist and determine the distribution. In other words, the question of uniqueness can be transformed to the question of the existence of all finite moments. In the following, we provide a sufficient condition, under which X_t admits finite exponential moments. From this moment condition, the uniqueness in law follows for a polynomial diffusion for (2.2.2) and (2.3.4) respectively. We omit the proofs and refer to Filipović and Larsson [2016] for more details.

Theorem 2.4.1. *If*

$$\mathbb{E}\left[e^{\delta\|X_0\|}\right] < \infty \text{ for some } \delta > 0 \tag{2.4.1}$$

and the diffusion coefficient satisfies the linear growth condition

$$\|a(x)\| \leq C(1 + \|x\|) \text{ for all } x \in E \tag{2.4.2}$$

for some constant C, then for each $t \geq 0$ there exists $\epsilon > 0$ with $\mathbb{E}\left[e^{\epsilon\|X_t\|}\right] < \infty$.

This sufficient condition of finite exponential moments for X_T infer the following result on uniqueness in law.

Theorem 2.4.2. *Let X be an E-valued solution to (2.2.2). If (2.4.1) and (2.4.2) hold, then any E-valued solution to (2.2.2) with the same initial law as X has the same law as X.*

Moreover, if $X_0 = x$ is deterministic, then (2.4.1) is satisfied. Therefore, the only requirement for uniqueness in law is the growth condition (2.4.2). We thus obtain the uniqueness in law for any E-valued solution X_t of (2.3.4), which builds the underlying framework for the electricity modeling in later chapters.

Theorem 2.4.3. *If the linear growth condition (2.4.2) is satisfied, then uniqueness in law holds for any E-valued solution to (2.3.4).*

Proof. This follows immediately from Theorem 2.4.2 with $X_0 = x$ deterministic. □

2.5 Existence and boundary attainment

In this section, we review the most important conditions regarding the existence of an E-value solution and the attainment of boundary of E. Here we omit all proofs, which can be found in Filipović and Larsson [2016].

Since the existence of $\mathbb{R}^d$-valued solutions to (2.2.2) holds due to continuity and linear growth of b and σ, existence of the E-valued solution to (2.2.2) boils down to stochastic invariance of E. Let $E = \{p \geq 0 \,|\, p \in \mathcal{P}\}$ for a finite collection of polynomials $\mathcal{P}$.

The underlying idea is that in order to not leave the state space, for any point at the boundary, its drift should be inward pointing and its diffusive moves should be parallel and not orthogonal to the boundary.

Theorem 2.5.1 (Necessary condition)**.** *Suppose there exists an E-valued solution to (2.2.2) with $X_0 = x$, for any $x \in E$. Then*

(i) *$a\nabla p = 0$ and $\mathcal{G}p \geq 0$ on $E \cap \{p = 0\}$ for each $p \in \mathcal{P}$*

For the sufficient condition, the following assumptions on the geometry of E, as well as the conditions on a, b are needed:

(G1) the ideal generated by $\{p\}$ is real for each $p \in \mathcal{P}$

(A1) $a \in \mathbb{S}_+^d$ on E

(A2) $a\nabla p = 0$ and $\mathcal{G}p > 0$ on $\{p = 0\}$ for each $p \in \mathcal{P}$

In particular, the conditions (G1) and (A2) together imply that $a\nabla p = hp$ for some vector of polynomials h. The following sufficient condition insures the existence of an E-valued solution.

Theorem 2.5.2 (Sufficient condition)**.** *Suppose (G1), (A1) and (A2) hold. Then $\mathcal{G}$ is polynomial on E, and there exists a continuous $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ such that $a = \sigma\sigma^\top$ on E and SDE (2.2.2) admits an E-valued solution X for any initial law of X_0, which spends zero time at the boundary of E:*

$$\int_0^t \mathbb{1}_{\{p(X_s)=0\}} ds = 0 \text{ for all } t \geq 0 \text{ and } p \in \mathcal{P} \tag{2.5.1}$$

The above theorem states that the diffusion X spend zero time at the boundary of the state space, but does not state whether or not the boundary was ever attained. The following theorem provides the necessary and the sufficient conditions for this to occur.

Theorem 2.5.3 (Boundary attainment, necessary and sufficient conditions)**.** *Let X be an E-valued solution to (2.2.2) satisfying (2.5.1). Let $p \in \mathcal{P}$ and h be a vector of polynomials such that $a\nabla p = hp$.*

1. If there exists a neighborhood U of $E \cap \{p = 0\}$ such that

$$2\mathcal{G}p - h^\top \nabla p \geq 0 \text{ on } E \cap U$$

then $p(X_t) > 0$ for all $t > 0$.

2. Let $\bar{x} \in E \cap \{p = 0\}$ and assume $\mathcal{G}p(\bar{x}) \geq 0$ and $2\mathcal{G}p(\bar{x}) - h(\bar{x})^\top \nabla p(\bar{x}) < 0$. Then there exists $\epsilon > 0$ such that if $\|X_0 - \bar{x}\| < \epsilon$ almost surely, then X hits $\{p = 0\}$ with positive probability.

Although in this thesis we only consider diffusion processes, the polynomial property also extends to jump diffusions. The extended generator of an E-valued *polynomial jump diffusion* X_t is of the form

$$\widetilde{\mathcal{G}} f(x) = \frac{1}{2} Tr\left(\widetilde{a}(x)\nabla^2 f(x)\right) + \widetilde{b}(x)^\top \nabla f(x) + \int_{R^d} \Big(f(x+\xi) - f(x) - \xi^\top \nabla f(x) \Big) \nu(x, d\xi),$$

where $\nu(x, d\xi)$ is a Lévy transition kernel; see e.g. Filipović and Larsson [2020] for an extensive treatment of polynomial jump diffusions.

Chapter 3

A multi-factor polynomial framework for long-term electricity forwards with delivery period

3.1 Introduction

Electricity differs from other energy commodities due to specific features such as limited storability, possibility of intra-day and day-ahead negative prices, its unique mechanism of the auction market, high liquidity of short- to medium-term trading and illiquidity of its long-term trading. Much of the academic literature is dedicated to short- to medium-term modeling of electricity spot and futures prices, as its highly frequent and huge data amount makes it ideal for empirical studies of time series analysis. However, the literature addressing the modeling of long-term electricity forwards and the corresponding hedging problems is scarce.

In this chapter, we propose a mathematically tractable multi-factor polynomial diffusion framework to model long-term forwards, which captures long-term properties such as mean reversion well. In this framework the computation of forwards and cross-maturity correlations is fully explicit. Fitting

the model to long time series of single market electricity data works easily and accurately. Furthermore, we set up a rolling hedge mechanism that only uses liquid forward contracts. This allows us to address the non-storability of electricity and poor liquidity in its long-term markets. Within the setup the hedging strategy we suggest minimizes the conditional variance of the cost processes at any time, and thus is risk-minimizing in the sense of Föllmer and Schweizer. A simulation study using the estimated model shows that the risk-minimizing rolling hedge significantly reduces, yet does not fully eliminate, the variance and skew of the long-term exposures.

The proposed modeling framework has various applications in forward modeling. It can be used to smoothly extrapolate the curve to the non-liquid horizon while calibrating it to the liquid horizon; it can also be used to smooth the forward surface implied by the market once calibrated and to filter out market noise; moreover, it can be used to model the prices within the real data horizon between two quotation dates. Furthermore, the model can be extended to model multiple electricity markets and other energy markets simultaneously. It can thus serve as an alternative model for risk management purposes, and for conducting simulations. We do however not pursue such multi-market extensions in this thesis.

Compared to other electricity modeling classes such as affine processes (mostly used as geometric models), this modeling framework has the advantage of being general but still very tractable, so that pricing formulas of spots, forwards (with instantaneous delivery) and forwards with delivery period have closed-form solutions. Moreover, it is possible to explicitly compute locally risk-minimizing hedging strategies in this framework which uses a rolling mechanism.

Our framework is introduced to model long-term markets and yearly forward contracts, which are the most liquidly traded long-term contracts. The primary focus is to capture dynamics over very long time horizons, including contracts with maturities far beyond the liquidity of long-term futures traded on the exchange. We calibrate the model to over-the-counter forwards with maturities of up to ten years from the quotation date. However, our framework can easily be extended to capture features such as spikes, seasonality and negative prices for spots and forwards with shorter time-to-maturity (day-ahead, week-ahead, month-ahead, quarter-ahead) and with shorter time frames (daily, monthly, quarterly). Incorporating such features does not change the polynomial structure, so that pricing and hedging remains tractable.

Polynomial models have been used to solve a number of problems in finance.

With the exception of Ware [2019], the polynomial processes have not been used for electricity modeling. Our polynomial framework makes assumption on properties of spot and forward and not on supply-demand relation, and thus falls into the category of classical reduced-form model (see Carmona and Coulon [2014] for details on reduced-form model versus structural approach). It is closest to the arithmetic models of Benth et al. [2007b,a, 2008a], and extends them by making the spot price not a linear combination but a squared combination of underlying polynomial processes. In doing so we extend the class of stochastic process on the one hand, and guarantee non-negative spot prices on the other hand.

The local risk-minimization hedging criterion of Föllmer & Schweizer 1991 is one the two main quadratic hedging approaches; see e.g. Föllmer and Schweizer [1991], Heath et al. [1999], Schweizer [1999, 1990] for references of the general theory of local risk-minimization, Follmer and Sondermann [1986] for the mean-variance hedge, and Heath et al. [2001] for a comparison of the two approaches. In a recent paper on hedging, a locally risk-minimizing hedge was given for the arithmetic model of Benth et al. under illiquidity; see Christodoulou et al. [2018]. Our work differs from theirs, as we consider a rolling hedge which only uses liquid forward contracts and give explicit expression for the locally risk-minimizing hedging strategy for our modeling framework.

This chapter is structured in the following way: In Section 2, we define the underlying polynomial framework, model the spot price as a quadratic function of it, and provide two main specifications. In Section 3, we define electricity forwards with and without delivery period. We give pricing formulas for forwards, as well as explicit expressions for covariances and correlations between different forwards. In Section 4, to incorporate time series observations of forward prices, we specify a market price of risk function, which determines the forward price dynamics under the real-world measure $\mathbb{P}$, and define the forward risk premium. In Section 5, we introduce a rolling hedge mechanism with liquidity constraints for hedging a long-term electricity commitment. Further, we give a rolling-hedge that is locally risk-minimizing in the sense of Föllmer and Schweizer. In Section 6, we perform model estimation of a specification of the polynomial framework to a time series of real observations of power forwards using a quadratic Kalman filter. Further we simulate forward curves and investigate the quality of the risk-minimizing hedge over various time horizon.

This chapter is based on Kleisinger-Yu, Komaric, Larsson, and Regez [2020].

3.2 The model

In this section we define the underlying polynomial framework. Firstly, we model the spot price S_t as a quadratic function of an underlying d-dimensional state variable X_t which evolves according to a polynomial diffusion. More precisely, we let

$$S_t = p_S(X_t) \tag{3.2.1}$$

$$dX_t = \kappa(\theta - X_t)dt + \sigma(X_t)dW_t \tag{3.2.2}$$

where $p_S(x) = c + x^\top Q x$ with $c \in \mathbb{R}_+$ and $Q \in \mathbb{S}^d_+$, $\kappa \in \mathbb{R}^{d\times d}$, $\theta \in \mathbb{R}^d$, W a d-dimensional Brownian motion under $\mathbb{Q}$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is continuous. We assume that the components of the diffusion matrix $a(x) := \sigma(x)\sigma(x)^\top$ are polynomials of degree at most two. This ensures that X_t is a polynomial diffusion, see Lemma 2.2 in Filipović and Larsson [2016].

The above formulation allows in particular to capture mean reversion, an important feature of electricity price dynamics. Empirically, this has been backed up by e.g. Koekebakker and Ollmar [2005]. They examined Nordic electricity forwards from 1995–2001 and observed that the short-term price varies around the long-term price, indicating mean reversion. Several economic arguments also support the mean-reverting property; see e.g. Escribano et al. [2011].

We will now focus on the following two specifications.

Specification 3.2.1 (Two-factor model)**.** *Let κ_Z, $\kappa_Y \in \mathbb{R}$, $\sigma_Z, \sigma_Y > 0$, and $\rho \in (-1, 1)$. The process $X_t := (Z_t, Y_t)^\top$ evolves according to the SDE*

$$\begin{aligned} dZ_t &= -\kappa_Z Z_t dt + \sigma_Z dW_t^{(1)} \\ dY_t &= \kappa_Y (Z_t - Y_t)dt + \rho\sigma_Y dW_t^{(1)} + \sigma_Y\sqrt{1-\rho^2}dW_t^{(2)} \end{aligned} \tag{3.2.3}$$

with $Z_0, Y_0 \in \mathbb{R}$ and $W_t = (W_t^{(1)}, W_t^{(2)})^\top$ a standard two-dimensional Brownian motion. Here Y_t mean-reverts at rate κ_Y towards the correlated process Z_t. And thus, Y_t and Z_t can be seen as factor processes that drive the short-end and long-end dynamics of spot prices respectively. This model is consistent with the empirical findings by Koekebakker and Ollmar [2005] regarding mean reversion. Let α, β, $c \in \mathbb{R}_+$ and let the spot price be given by

$$S_t := c + \alpha Y_t^2 + \beta Z_t^2.$$

This guarantees nonnegative spot price, as $S_t \geq c \geq 0$. This specification is of the form (3.2.1)–(3.2.2) with

(3.2.4)
$$Q = \begin{pmatrix} \beta & 0 \\ 0 & \alpha \end{pmatrix}, \ \kappa = \begin{pmatrix} \kappa_Z & 0 \\ -\kappa_Y & \kappa_Y \end{pmatrix}, \ \theta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \sigma(x) = \sigma(z,y) = \begin{pmatrix} \sigma_Z & 0 \\ \rho\sigma_Y & \sigma_Y\sqrt{1-\rho^2} \end{pmatrix}.$$

Specification 3.2.2 (Three-factor model)**.** *We now present a specification which extends the two-factor model by modeling correlation between the underlying processes stochastically via a Jacobi process. Conditions under which the model exists and is unique are given below. Let κ_Z, $\kappa_Y \in \mathbb{R}$, $\kappa_R, \sigma_Z, \sigma_Y, \sigma_R > 0$, and $\theta_R \in (-1,1)$. The process $X_t := (Z_t, Y_t, R_t)^\top$ evolves according to the SDE*

$$\begin{aligned} dZ_t &= -\kappa_Z Z_t dt + \sigma_Z dW_t^{(1)} \\ dY_t &= \kappa_Y(Z_t - Y_t)dt + R_t\sigma_Y dW_t^{(1)} + \sigma_Y\sqrt{1-R_t^2}dW_t^{(2)} \\ dR_t &= \kappa_R(\theta_R - R_t)dt + \sigma_R\sqrt{1-R_t^2}dW_t^{(3)} \end{aligned} \tag{3.2.5}$$

with $Z_0, Y_0 \in \mathbb{R}$, $R_0 \in (-1,1)$, and $W_t = (W_t^{(1)}, W_t^{(2)}, W_t^{(3)})^\top$ a standard three-dimensional Brownian motion. Let α, β, $c \in \mathbb{R}_+$ and let the spot price be given by

$$S_t := c + \alpha Y_t^2 + \beta Z_t^2.$$

This specification is of the form (3.2.1)–(3.2.2) with

$$\begin{aligned} Q = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 0 \end{pmatrix}, \kappa = \begin{pmatrix} \kappa_Z & 0 & 0 \\ -\kappa_Y & \kappa_Y & 0 \\ 0 & 0 & \kappa_R \end{pmatrix}, \theta = \begin{pmatrix} 0 \\ 0 \\ \theta_R \end{pmatrix}, \\ \sigma(x) = \sigma(z,y,r) = \begin{pmatrix} \sigma_Z & 0 & 0 \\ r\sigma_Y & \sigma_Y\sqrt{1-r^2} & 0 \\ 0 & 0 & \sigma_R\sqrt{1-r^2} \end{pmatrix}. \end{aligned} \tag{3.2.6}$$

Remark 3.2.3. *Although Specification 3.2.2 is not used in our empirical analysis, we include it as an illustration of the flexibility of the polynomial framework.*

A possible use of Specification 3.2.2 is to model multi-energy commodities simultaneously. Here is a simple illustration of this: let one factor (Z_t) drive the short-term price of one market, and let the other factor (Y_t) drive the short-term price of the other market. Since energy markets evolve dynamically and prices are generally non-stationary over time (Krečar et al. [2019]), it is useful to have stochastic correlation between (the short ends of) different markets, modeled by

a factor (R_t). The setup could be complemented with a fourth factor driving common long-term prices.

Alternatively, two markets can also be modeled as follows: two factors with the dynamics of Y_t, (Y_t^i, $i = 1,2$), can be used to model short-term prices of each market; one factor (Z_t) drives the common long-end prices. In order to account for the changing relationship between short-term and long-term prices, another two factors with the dynamics of R_t, (R_t^i, $i = 1,2$), can be added to model the stochastic correlation between the short-term and long-term prices in each market.

Proposition 3.2.4. *Recall that $\kappa_R > 0$, $\theta_R \in (-1,1)$, and assume moreover that*

$$\kappa_R(1+\theta_R) \geq \sigma_R^2, \tag{3.2.7}$$

$$\kappa_R(1-\theta_R) \geq \sigma_R^2. \tag{3.2.8}$$

Then for any initial condition with $Z_0 \in \mathbb{R}$, $Y_0 \in \mathbb{R}$ and $R_0 \in (-1,1)$, there exists a unique strong solution $X_t = (Z_t, Y_t, R_t)^\top$ of the SDE (3.2.5). Furthermore, this solution satisfies $R_t \in (-1,1)$ for all $t \geq 0$.

Proof. In the following we show the existence and uniqueness of a strong solution R_t as well as its boundary non-attainment. Once this is shown, we can explicitly find $\widetilde{X}_t := (Y_t, Z_t)$ in terms of R_t. Indeed, Itô's formula yields

$$d\left(e^{\tilde{\kappa}t}\widetilde{X}_t\right) = e^{\tilde{\kappa}t}\tilde{\kappa}\theta dt + e^{\tilde{\kappa}t}\tilde{\sigma}(R_t)d\widetilde{W}_t,$$

where $\tilde{\kappa} = \begin{pmatrix} \kappa_Z & 0 \\ -\kappa_Y & \kappa_Y \end{pmatrix}$ and $\tilde{\sigma}(r) = \begin{pmatrix} \sigma_Z & 0 \\ r\sigma_Y & \sigma_Y\sqrt{1-r^2} \end{pmatrix}$, which implies that

$$\widetilde{X}_t = e^{-\tilde{\kappa}t}\widetilde{X}_0 + \int_0^t e^{-\tilde{\kappa}(t-s)}\tilde{\kappa}\theta ds + \int_0^t e^{-\tilde{\kappa}(t-s)}\tilde{\sigma}(R_t)d\widetilde{W}_s.$$

We now prove existence of a weak solution of the SDE for R_t. Let $\varphi(r)$ be a continuous function that is equal to one for $r \in [-1,1]$ and is equal to zero for $|r| > 2$, for example

$$\varphi(r) = \begin{cases} 1 & |r| \leq 1 \\ 2 - |r| & 1 < |r| \leq 2 \\ 0 & |r| > 2. \end{cases}$$

We let $\tilde{b}(r) := b(r)\varphi(r)$ with $b(r) := \kappa_R(\theta_R - r)$ and $\tilde{\sigma}(r) := \sigma_R\sqrt{(1-r^2)_+}$. Then $\tilde{b}(r)$ and $\tilde{\sigma}(r)$ are continuous and bounded, and hence an $\mathbb{R}$-valued weak solution

R_t exists for the SDE $dR_t = \tilde{b}(R_t)dt + \tilde{\sigma}(R_t)dW_t^{(3)}$; see Theorem 4.22 of Section 5.4D in Karatzas and Shreve [1998]. We next show that R_t stays in $(-1,1)$ using a version of "McKean's argument". Let $p(r) := 1 - r^2$ and note that $p(R_0) > 0$. Further define the stopping times $\tau_n := \inf\{t : p(R_t) \le \frac{1}{n}\}$ and $\tau := \lim_{n\to\infty} \tau_n$. Observe that (3.2.7)–(3.2.8) imply that $\kappa_R(1 - \theta_R R_t) - \sigma_R^2 \ge 0$ for all $t < \tau$. Combined with Itô's formula, this yields

$$\begin{aligned} d\log p(R_t) &= \left(-(2\kappa_R - \sigma_R^2) + 2\frac{\kappa_R(1-\theta_R R_t) - \sigma_R^2}{1 - R_t^2}\right)dt - \frac{2\sigma_R R_t}{\sqrt{1-R_t^2}}dW_t^{(3)} \\ &\ge -(2\kappa_R - \sigma_R^2)dt - \frac{2\sigma_R R_t}{\sqrt{1-R_t^2}}dW_t^{(3)}, \end{aligned}$$

for $t < \tau$. Consider the process

$$M_t := \int_0^t \frac{2\sigma_R R_s}{\sqrt{1-R_s^2}}dW_s^{(3)}, \quad t < \tau.$$

Then M_t is a local martingale on the stochastic interval $[0,\tau)$. By definition, this means that for all $n \in \mathbb{N}$, $M_{t\wedge\tau_n}$ is a local martingale. We now show that $\tau = \infty$ a.s. Suppose for contradiction that $\mathbb{P}(\tau < \infty) > 0$. Then there exists a large $T < \infty$ such that $\mathbb{P}(\tau < T) > 0$. Note that

$$M_t \ge -(2\kappa_R - \sigma_R^2)t + \log p(R_0) - \log p(R_t) \ge -(2\kappa_R - \sigma_R^2)T + \log p(R_0) \tag{3.2.9}$$

for all $t < T \wedge \tau$. Thus $M_{t\wedge T}$ is uniformly bounded from below, and hence a local supermartingale on the stochastic interval $[0,\tau)$. The supermartingale convergence theorem for processes on stochastic interval $[0,\tau)$ now gives that $\lim_{t\to\tau} M_{t\wedge T}$ exists in $\mathbb{R}$ almost surely; see e.g. the proof of Theorem 5.7 in Filipović and Larsson [2016]. Hence, in view of (3.2.9), $-\log p(R_t)$ is pathwise bounded above on $[0, T \wedge \tau)$, which in turn means that $\tau > T$ a.s. This contradiction shows that $R_t \in (-1,1)$ for all $t \ge 0$.

Now let $\sigma(r) = \sqrt{1-r^2}$. Then $\tilde{b}(R_t) = b(R_t)$ and $\tilde{\sigma}(R_t) = \sigma(R_t)$ on $(-1,1)$, and therefore, R_t is an $(-1,1)$-valued weak solution of the SDE $dR_t = b(R_t)dt + \sigma(R_t)dW_t^{(3)}$. For the existence and uniqueness of strong solutions, we note that $b(.)$ is Lipschitz continuous, and $\sigma(.)$ is Hölder continuous of order 1/2. Hence, pathwise uniqueness holds for this SDE; see Theorem 3.5(ii) in Revuz and Yor [2013]. As a result, any $(-1,1)$-valued solution is a strong solution by the Yamada–Watanabe theorem; see e.g. Theorem 1.7 in Revuz and Yor [2013]. □

Although our main focus in this chapter is on pricing and hedging of long-term contracts, let us indicate how the framework can be adjusted to incorporate features that are important over shorter time horizons.

Negative prices

In short-term electricity markets (real-time or day-ahead markets), prices frequently become negative; see e.g. Carmona and Coulon [2014] for PJM, Genoese et al. [2010] for German EEX. As electricity is non-storable, any disturbance of demand or of supply can cause negative prices.[1] The polynomial model can be extended to allow for negative prices for short-term modeling by simply taking $c < 0$. This way the spot price is bounded from below by c, $S_t \geq c$, which can be negative. This small modification does not change the polynomial structure, and thus, all computations and properties for forwards and hedges remain the same.

For long-term markets this feature is less relevant, as long-term prices are generally insensitive to temporary shocks. Indeed, the data of German Calendar year baseload forwards (over 8 years) does not contain negative prices.

Seasonality

In electricity markets, prices highly depend on the exact delivery period, e.g. offpeak vs. peak hours, winter months vs. summer months, or specific quarters. Thus, if we compare contracts with same delivery length but different delivery periods, that is, different subperiods of a year, it is important to first adjust for seasonality before making reasonable comparison. It is possible to incorporate seasonality by making p_S not only a state-dependent, but also time-dependent mapping. More specifically, we can let $p_S(t, x) := c(t) + x^\top Q(t)x$, where c and Q have temporal components. This leads to a time-inhomogeneous version of the polynomial property, which remains tractable.

Note that all yearly baseload contracts deliver throughout the year and not only for a specific subperiod of the year. To capture these forwards in long-term markets, it is not necessary to explicitly model seasonality.

[1] To be more precise, negative prices can be caused by e.g. error predictions of the load, high temperature volatilities, network transmission and congestion issues (causing oversupply in one region and undersupply in another), and overdemand through prediction error from generation via renewable energy (wind and PV).

Spikes/Jumps

In short-term markets, one often observes extreme price changes in spot prices, known as spikes. These result from unanticipated shocks in demand, and exist only temporarily. In other words, prices don't stay at the new level, but revert rapidly back to the previous level. Because of their temporary nature, it is reasonable to argue that the spikes have a negligible effect on long-term prices, and therefore, should not be included in the framework for modeling long-term electricity forwards.

However, our model can be extended to account for spikes if needed, say to model short-term spot prices, or joint short- and long-term markets. One possible way of doing so is to multiply the spot price by a mean-reverting jump process that jumps and then very quickly mean-reverts towards its standard level of 1. A simple example is given by:

$$\begin{aligned} S_T &= p_S(X_t)J_t, \\ dJ_t &= \theta_J(1 - J_t)dt + \int \sigma_J(X_t, v)N(dv, dt), \end{aligned}$$

where $N(dv, dt)$ is a Poisson random measure, θ_J a large mean-reversion parameter, which forces the process to revert quickly to the previous level after a jump. Another possibility is to incorporate spikes by an additive component, e.g.

$$\begin{aligned} S_T &= p_S(X_t) + J_t, \\ dJ_t &= -\theta_J J_t dt + \int \sigma_J(X_t, v)N(dv, dt). \end{aligned}$$

Either way, the extensions do not change the behavior of long-dated forward but only the short-term forward and spot, because all the jumps mean-revert very quickly and so do not have an effect on long term prices. Provided $\sigma_J(x, v)$ is chosen appropriately, many of the properties of polynomial diffusions (such as the moment formula) still apply; see [Filipović and Larsson, 2020, Section 5] for more details.

3.3 The term structure of forward prices

In this section we define electricity forwards, present their pricing formulas and give expressions for covariances and correlations between different forwards.

The price at time-t of an electricity forward with instantaneous delivery at time $T \geq t$ is given by

$$f(t,T,X_t) := \mathbb{E}_{\mathbb{Q}}\left[S_T \middle| \mathcal{F}_t\right]. \tag{3.3.1}$$

In practice, electricity is not delivered instantaneously, but gradually over a period of time. This leads us to the following definition: the time-t price of an electricity forward with delivery period $[T_1, T_2)$, $t \leq T_1 < T_2$, is given by

$$F(t,T_1,T_2,X_t) := \frac{1}{T_2 - T_1}\mathbb{E}_{\mathbb{Q}}\left[\int_{T_1}^{T_2} S_u \, du \middle| \mathcal{F}_t\right]. \tag{3.3.2}$$

Note that a forward contract (financial or physical) can have settlement that takes place either before or after the delivery period. Discounting is not needed in the pricing, as the difference in cashflow can be evened out by the purchase of a bond of that time period. $F(t,T_1,T_2,X_t)$ is often also referred to as swap price, as the delivery of underlying power happens over a period of time and thus the price is the averaged price over that period.

It is intuitive that a forward with delivery period is the summation of all forwards (with instantaneous delivery) that deliver at single time points within the delivery period; moreover, a forward with delivery period which collapses into one single time point should be priced the same as a forward with instantaneous delivery. The following proposition confirms this relationship between forwards with and without delivery period.

Proposition 3.3.1. *For $t \leq T_1 \leq T_2$, we have:*

$$F(t,T_1,T_2,X_t) = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} f(t,u,X_t)du.$$

Moreover,

$$\lim_{T_2 \to T_1} F(t,T_1,T_2,X_t) = f(t,T_1,X_t).$$

Proof. In view of (3.3.1) and (3.3.2), the first identity follows from the conditional version of Tonelli's theorem since S_t is nonnegative. The second identity then follows from the fundamental theorem of calculus, using that $f(t,T,X_t)$ is continuous in T, see Proposition 3.3.2 below. □

The following result gives closed-form expression for the forward prices.

Proposition 3.3.2 (Pricing formula for forwards)**.** *Let $\vec{p}_S$ be the coordinate representation of $p_S(x)$. The time-t price of $f(t,T,X_t)$ for $t \le T$ is*

$$f(t,T,X_t) = H(X_t)^\top e^{(T-t)G}\vec{p}_S,$$

and the time-t price of $F(t,T_1,T_2,X_t)$ for $t \le T_1 \le T_2$ is

$$F(t,T_1,T_2,X_t) = \frac{1}{T_2-T_1} H(X_t)^\top e^{(T_1-t)G}\int_0^{T_2-T_1} e^{uG}du\,\vec{p}_S.$$

Proof. This follows from Theorem 2.3.1 and rearranging terms. □

Note that G is a non-invertible matrix. Still, $\int_0^\tau e^{uG}du$ is explicit; see Appendix 3.A for the explicit computation.

Specification 3.2.1 Recall the Specification 3.2.1 in Section 3.2. We consider the basis given by

$$H(x) = (1, z, y, z^2, yz, y^2)^\top,\ x = (z,y)^\top. \tag{3.3.3}$$

Then S_t can be uniquely represented as:

$$S_t = H(X_t)^\top \vec{p}_S \quad \text{with} \quad \vec{p}_S = (c,0,0,\beta,0,\alpha)^\top. \tag{3.3.4}$$

For any C^2 function f and $x = (z,y)^\top \in \mathbb{R}^2$, the generator $\mathcal{G}$ is :

$$\mathcal{G}f(x) = \begin{pmatrix} -\kappa_Z z \\ \kappa_Y z - \kappa_Y y \end{pmatrix}^\top \nabla f(x) + \frac{1}{2}\mathrm{Tr}\left(\begin{pmatrix} \sigma_Z^2 & \rho\sigma_Y\sigma_Z \\ \rho\sigma_Y\sigma_Z & \sigma_Y^2 \end{pmatrix}\nabla^2 f(x)\right).$$

Applying $\mathcal{G}$ to each element of $H(X_t)$ gives its matrix representation,

$$G = \begin{pmatrix} 0 & 0 & 0 & \sigma_Z^2 & \rho\sigma_Y\sigma_Z & \sigma_Y^2 \\ 0 & -\kappa_Z & \kappa_Y & 0 & 0 & 0 \\ 0 & 0 & -\kappa_Y & 0 & 0 & 0 \\ 0 & 0 & 0 & -2\kappa_Z & \kappa_Y & 0 \\ 0 & 0 & 0 & 0 & -\kappa_Z-\kappa_Y & 2\kappa_Y \\ 0 & 0 & 0 & 0 & 0 & -2\kappa_Y \end{pmatrix}. \tag{3.3.5}$$

Specification 3.2.2 Recall Specification 3.2.2 in Section 3.2. Here the general $\mathcal{G}$ actually preserves a proper subspace of Pol_2, namely the one spanned by the components of

$$H(x) = (1, z, y, r, z^2, yz, y^2)^\top, \ x = (z, y, r)^\top. \tag{3.3.6}$$

Therefore it is not necessary to include the remaining basis functions in the definition of H. Then S_t can be uniquely represented as

$$S_t = H(X_t)^\top \vec{p}_S \ \text{ with } \ \vec{p}_S = (c,\ 0,\ 0,\ 0,\ \beta,\ 0,\ \alpha)^\top. \tag{3.3.7}$$

For any C^2 function f and $x = (z, y, r)^\top \in \mathbb{R}^3$, the generator $\mathcal{G}$ is

$$\mathcal{G}f(x) = \begin{pmatrix} -\kappa_Z z \\ -\kappa_Y y + \kappa_Y z \\ \kappa_R(\theta_R - r) \end{pmatrix}^\top \nabla f(x) + \frac{1}{2}\mathrm{Tr}\left(\begin{pmatrix} \sigma_Z^2 & \sigma_Y\sigma_Z r & 0 \\ \sigma_Y\sigma_Z r & \sigma_Y^2 & 0 \\ 0 & 0 & \sigma_R^2(1-r^2) \end{pmatrix}\nabla^2 f(x)\right).$$

Applying $\mathcal{G}$ to each element of $H(X_t)$ gives

$$G = \begin{pmatrix} 0 & 0 & 0 & \kappa_R\theta_R & \sigma_Z^2 & 0 & \sigma_Y^2 \\ 0 & -\kappa_Z & \kappa_Y & 0 & 0 & 0 & 0 \\ 0 & 0 & -\kappa_Y & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\kappa_R & 0 & \sigma_Y\sigma_Z & 0 \\ 0 & 0 & 0 & 0 & -2\kappa_Z & \kappa_Y & 0 \\ 0 & 0 & 0 & 0 & 0 & -\kappa_Z - \kappa_Y & 2\kappa_Y \\ 0 & 0 & 0 & 0 & 0 & 0 & -2\kappa_Y \end{pmatrix}. \tag{3.3.8}$$

For later use, we briefly discuss the instantaneous quadratic covariation and correlations between different forwards and give explicit forms for both specifications. The instantaneous covariation between two forwards with instantaneous delivery at T_1 and T_2 is, at time $t \le T_1 \wedge T_2$,

$$\frac{d}{dt}\langle f(t, T_1, X_t), f(t, T_2, X_t)\rangle = \vec{p}_S^\top e^{(T_2-t)G^\top}\Sigma(X_t)e^{(T_1-t)G}\vec{p}_S, \tag{3.3.9}$$

where

$$\Sigma(X_t)dt = d\langle H(X), H(X)\rangle_t. \tag{3.3.10}$$

We define the corresponding instantaneous correlation as

$$\begin{aligned} &\mathrm{Corr}[f(t, T_1, X_t), f(t, T_2, X_t)] \\ &= \frac{\vec{p}_S^\top e^{(T_2-t)G^\top}\Sigma(X_t)e^{(T_1-t)G}\vec{p}_S}{\sqrt{\vec{p}_S^\top e^{(T_1-t)G^\top}\Sigma(X_t)e^{(T_1-t)G}\vec{p}_S\ \vec{p}_S^\top e^{(T_2-t)G^\top}\Sigma(X_t)e^{(T_2-t)G}\vec{p}_S}} \end{aligned} \tag{3.3.11}$$

with $\Sigma(X_t)$ from (3.3.10). The matrices $\Sigma(X_t)$ for Specification 3.2.1 and Specification 3.2.2 are given in Appendix 3.B. Similarly, for $t \le T_1 < T_2$ and $t \le T_3 < T_4$, the time-t instantaneous covariation of forwards with delivery periods $[T_1, T_2)$ and $[T_3, T_4)$ is

$$\frac{d}{dt}\langle F(t,T_1,T_2,X_t), F(t,T_3,T_4,X_t)\rangle = \vec{w}_{34}^{\top} e^{(T_3-t)G^{\top}} \Sigma(X_t) e^{(T_1-t)G} \vec{w}_{12}, \tag{3.3.12}$$

and the time-t instantaneous correlation is:

$$\begin{aligned} &\operatorname{Corr}[\ F(t,T_1,T_2,X_t), F(t,T_3,T_4,X_t)\ ] \\ &= \frac{\vec{w}_{34}^{\top} e^{(T_3-t)G^{\top}} \Sigma(X_t) e^{(T_1-t)G} \vec{w}_{12}}{\sqrt{\vec{w}_{12}^{\top} e^{(T_1-t)G^{\top}} \Sigma(X_t) e^{(T_1-t)G} \vec{w}_{12}\ \vec{w}_{34}^{\top} e^{(T_3-t)G^{\top}} \Sigma(X_t) e^{(T_3-t)G} \vec{w}_{34}}} \end{aligned} \tag{3.3.13}$$

with $\Sigma(X_t)$ from (3.3.10) and

$$\vec{w}_{ij} = \int_{T_i}^{T_j} e^{uG} du\, \vec{p}_S. \tag{3.3.14}$$

Remark 3.3.3 (Option pricing)**.** *Let $p(X_T)$ be the payoff function of an option based on a forward or a spot. For example, for a European call on a forward with delivery period $[T_1, T_2)$, strike K, and maturing T, we have $p(X_T) = (F(T,T_1,T_2,X_T) - K)^+$. Modulo discounting, the time-t price of such an option is the $\mathcal{F}_t$-conditional expectation of $p(X_T)$ under $\mathbb{Q}$. If p is a polynomial function, we can obtain explicit pricing for the option by Theorem 2.3.1 (if the option is based on a spot) or Proposition 3.3.2 (if the option is based on a forward). If p is not a polynomial, an approximation scheme is required. For example, one can use the polynomial expansion method described in [Filipović and Larsson, 2019, Section 7].*

3.4 Market price of risk specification

In order to incorporate time series observations of real-world forward curves, we must specify the forward dynamics under the real-world probability measure $\mathbb{P}$. Thus, in this section, we specify a market price of risk function $\lambda : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\lambda(x) = \sigma(x)^{-1}(\gamma + \Lambda x)$$

for some $\gamma \in \mathbb{R}^d$ and $\Lambda \in \mathbb{S}^{d\times d}$, and denote the associated Radon–Nikodym density process by

$$M_t^{\lambda} = \exp\Big(\int_0^t \lambda(X_s)^{\top} dW_s - \frac{1}{2}\int_0^t \|\lambda(X_s)\|^2 ds \Big). \tag{3.4.1}$$

We choose γ and Λ such that M_t^λ is a true martingale. We can then define $\mathbb{P}$ on every finite time interval $[0,T]$ via its Radon–Nikodym density $\frac{d\mathbb{P}}{d\mathbb{Q}}|_{\mathcal{F}_T} = M_T^\lambda$. Then, by Girsanov's theorem, the $\mathbb{P}$-dynamics of X_t becomes

$$dX_t = [(\kappa\theta + \gamma) - (\kappa - \Lambda)X_t]dt + \sigma(X_t)\, dW_t^{\mathbb{P}} \tag{3.4.2}$$

with $dW_t^{\mathbb{P}} := dW_t - \lambda(X_t)dt$. Note that the speed of mean reversion is now adjusted to $\kappa - \Lambda$ from κ.

Consider now Specification 3.2.1. In this case M_t^λ is a true martingale for any choice for γ and Λ, as the following result shows.

Proposition 3.4.1. *Let X_t evolve according to (4.2.5). Then M_t^λ from (3.4.1) is a martingale.*

Proof. Define $\widetilde{X}_t := (Z_t, Y_t, Z_t^2, Y_t Z_t, Y_t^2, \int_0^t (\sigma^{-1}\gamma + \sigma^{-1}\Lambda X_t)^\top dW_t)^\top$. Note that $\widetilde{X}_t$ has drift and diffusion that are affine in $\widetilde{X}_t$; see computations in Section 3.3, Appendix 3.B and Section 3.6.2. Thus, by Kallsen & Muhle-Karbe (Corollary 3.9 in Kallsen and Muhle-Karbe [2010]), M_t^λ is a true martingale. □

To be explicit, let $\Lambda = \mathrm{diag}(\lambda_Z, \lambda_Y)$ and $\gamma = (\gamma_Z, \gamma_Y)^\top$. Then the $\mathbb{P}$-dynamics of X_t is given by:

$$dX_t = \left[\begin{pmatrix}\gamma_Z\\ \gamma_Y\end{pmatrix} - \begin{pmatrix}\kappa_Z - \lambda_Z & 0\\ -\kappa_Y & \kappa_Y - \lambda_Y\end{pmatrix} X_t\right] dt + \begin{pmatrix}\sigma_Z & 0\\ \rho\sigma_Y & \sigma_Y\sqrt{1-\rho^2}\end{pmatrix} dW_t^{\mathbb{P}}. \tag{3.4.3}$$

This can also be written as $dX_t = \kappa'(\theta' - X_t)dt + \sigma(X_t)dW_t^{\mathbb{P}}$ with

$$\kappa' = \begin{pmatrix}\kappa_Z - \lambda_Z & 0\\ -\kappa_Y & \kappa_Y - \lambda_Y\end{pmatrix}, \qquad \theta' = \begin{pmatrix}\frac{\gamma_Z}{\kappa_Z - \lambda_Z}\\ \frac{\gamma_Y}{\kappa_Y - \lambda_Y} + \frac{\kappa_Y \gamma_Z}{(\kappa_Z - \lambda_Z)(\kappa_Y - \lambda_Y)}\end{pmatrix},$$

and $\sigma(x)$ from (3.2.4).

In the case of Specification 3.2.2 it is a more delicate problem to determine those market price of risk parameters for which M_t^λ is a true martingale. Since we will not use Specification 3.2.2 in our empirical analysis, we do not consider this issue here.

Forward risk premium

We define the forward risk premium as the difference of the forward and the predicted spot price. The time-t forward risk premium of a forward with

instantaneous delivery at $T \geq t$ is thus given by

$$R(t,T,X_t) := \mathbb{E}_{\mathbb{Q}}[S_T \mid \mathcal{F}_t] - \mathbb{E}_{\mathbb{P}}[S_T \mid \mathcal{F}_t],$$

and the time-t forward risk premium of a forward with delivery period $[T_1,T_2)$, $t \leq T_1 < T_2$, is given by

$$R(t,T_1,T_2,X_t) := \frac{1}{T_2-T_1}\mathbb{E}_{\mathbb{Q}}\left[\int_{T_1}^{T_2} S_u\,du \mid \mathcal{F}_t\right] - \frac{1}{T_2-T_1}\mathbb{E}_{\mathbb{P}}\left[\int_{T_1}^{T_2} S_u\,du \mid \mathcal{F}_t\right].$$

The notion above is consistent with the ex-ante notion of forward risk premium used by e.g. Benth, Cartea, and Kiesel [2008b], Benth and Meyer-Brandis [2009], Benth, Kiesel, and Nazarova [2012], Benth and Schmeck [2014], Benth, Piccirilli, and Vargiolu [2019], Krečar, Benth, and Gubina [2019]. Both the $\mathbb{Q}$- and $\mathbb{P}$-conditional expectations can be computed using the pricing formula in Proposition 3.3.2. We obtain the following explicit expressions for forward risk premia:

$$R(t,T,X_t) = H(X_t)^\top\left[e^{(T-t)G} - e^{(T-t)G^\lambda}\right]\vec{p}_S,$$

and

$$R(t,T_1,T_2,X_t) = \frac{1}{T_2-T_1}H(X_t)^\top\left[e^{(T_1-t)G}\int_0^{T_2-T_1} e^{uG}du - e^{(T_1-t)G^\lambda}\int_0^{T_2-T_1} e^{uG^\lambda}du\right]\vec{p}_S,$$

where G^λ denotes the matrix representation of the generator $\mathcal{G}$ under $\mathbb{P}$. For example, for Specification 3.2.1 under $\mathbb{P}$, X_t evolves according to (3.4.3), and G^λ is given by

$$G^\lambda = \begin{pmatrix} 0 & \gamma_Z & \gamma_Y & \sigma_Z^2 & \rho\sigma_Y\sigma_Z & \sigma_Y^2 \\ 0 & \lambda_Z-\kappa_Z & \kappa_Y & 2\gamma_Z & \gamma_Y & 0 \\ 0 & 0 & \lambda_Y-\kappa_Y & 0 & \gamma_Z & 2\gamma_Y \\ 0 & 0 & 0 & 2(\lambda_Z-\kappa_Z) & \kappa_Y & 0 \\ 0 & 0 & 0 & 0 & (\lambda_Z+\lambda_Y)-(\kappa_Z+\kappa_Y) & 2\kappa_Y \\ 0 & 0 & 0 & 0 & 0 & 2(\lambda_Y-\kappa_Y) \end{pmatrix}$$

The forward risk premium arises from the market price of risk $\lambda(X_t)$ and the associated measure change via the Girsanov's theorem, designed so that the polynomial structure is preserved. This produces stochastic and time varying forward risk premia. The risk premia do not have a definite sign, and can alternate between being positive and negative.[2] There is an extensive literature

[2]Empirical studies of electricity forward risk premia show mixed findings; see e.g. Bunn and Chen [2013] for a literature survey, and Valitov [2019], Viehmann [2011] for discussions of the risk premium in the short-term German market in particular.

on market price of risk specifications, forward risk premia, and measure changes for electricity modeling; see e.g Benth, Cartea, and Kiesel [2008b], Weron [2008], Benth and Meyer-Brandis [2009], Benth and Schmeck [2014], Krečar, Benth, and Gubina [2019], Benth, Piccirilli, and Vargiolu [2019].

3.5 Hedging

In this section we first describe a rolling hedge setup with constraints which addresses the illiquidity and non-storability issues when hedging a long-term electricity contract. Rolling hedges for commodities form a well-known hedging scheme; see for example Glasserman [2001], Neuberger [1999]. We then briefly review the locally risk-minimizing hedge of Föllmer and Schweizer, and give a rolling hedge for our modeling framework that is risk-minimizing.

3.5.1 A rolling hedge setup

Suppose we have committed to deliver power from year $\widetilde{T}$ to year $\widetilde{T}+1$ for a large $\widetilde{T} \in \mathbb{N}$ (e.g. $\widetilde{T} = 10$ years) and our objective is to hedge this long-term electricity commitment. In our framework the time-t valuation of the commitment is

$$\widetilde{F}_t := F(t, \widetilde{T}, \widetilde{T}+1, X_t) = \mathbb{E}_{\mathbb{Q}}\left[\int_{\widetilde{T}}^{\widetilde{T}+1} S_u du \middle| \mathcal{F}_t\right]$$

Note that $\widetilde{F}_t$ is a $\mathbb{Q}$-martingale and the pricing formula (Proposition 3.3.2) gives explicit pricing at any $t \in [0, \widetilde{T}]$. In an interest rate context, the analogous hedging problem is rather easy: just buy bonds and hold them as the payout in 10 years is known in advance. For electricity the problem is more difficult for a number of reasons:

- Long-term forwards are not liquidly traded (otherwise *buy and hold* the financial contracts as in the interest rate context);
- Electricity cannot be stored without significant costs (otherwise *cash and carry* as for other storable commodities: simply buy the amount needed in $[\widetilde{T}, \widetilde{T}+1]$ and hold).
- Only short-term / near-dated contracts with same delivery length is available. But its underlying commodity (electricity) is not the same as the one underlying a long-term contract because power is not storable.

> Some empirical evidence suggests that short-term prices carry limited information about what spot prices will be far into the future (see Handika et al. [2012]).

One possible strategy in this case is a *rolling hedge*, where we take a long position in near-term contracts as a hedge, and roll the hedge going forward. The underlying assumption of this strategy is that near-dated yearly contracts are highly correlated with far-dated yearly contracts, and become more so as the maturity date approaches[3]. This assumption is supported by the data; see Figure 3.C.1 in Appendix 3.C. A visualzation of this rolling mechanism is provided in Figure 3.5.1.

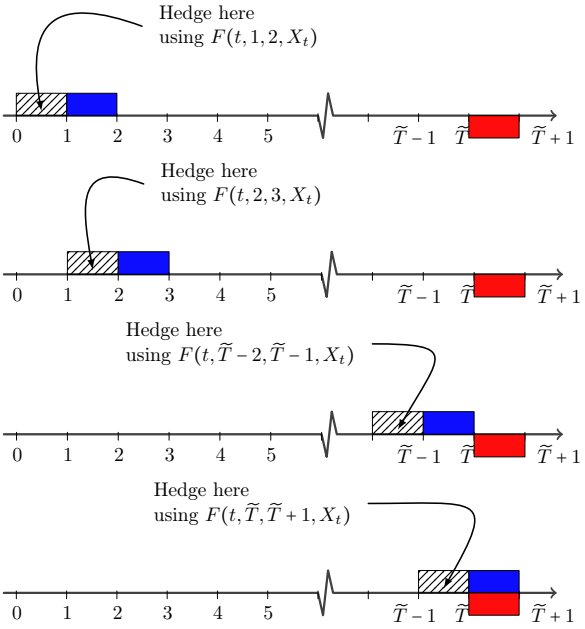


Figure 3.5.1: The mechanism of rolling hedges.

[3]Note that this statement does not contradict the common perception that the short- and long-term data are not very correlated, e.g. Koekebakker and Ollmar [2005]. The first nearby calendar year forward is often considered a medium-term or even a long-term contract.

To formalize this, let us first define the price process P_t containing all calendar-year forwards with a one-year delivery period (short: cal forward):

$$P_t = \begin{pmatrix} P_t^1 \\ P_t^2 \\ \vdots \\ P_t^{N-1} \\ P_t^N \end{pmatrix} = \begin{pmatrix} F(t,1,2,X_t) \\ F(t,2,3,X_t) \\ \vdots \\ F(t,N-1,N,X_t) \\ F(t,N,N+1,X_t) \end{pmatrix} \tag{3.5.1}$$

where $N = \widetilde{T}$ and $P_t^N = \widetilde{F}_t$. Note that each P_t^k is a $\mathbb{Q}$-martingale by its definition (3.3.2). By Proposition 3.3.2, P_t^k can be expressed as

$$P_t^k = H(X_t)^\top e^{(k-t)\,G} \vec{w}_{01}, \tag{3.5.2}$$

where $\vec{w}_{01}$ is defined in (3.3.14).

An admissible[4] hedging strategy is an $\mathbb{R}^{N+1}$-valued process $\varphi_t = (\eta_t, \xi_t)^\top = (\eta_t, \xi_t^1, \ldots, \xi_t^N)^\top$, where η_t is adapted (representing bank account) and ξ_t is predictable (representing amount of tradable assets or hedge ratio), and satisfies

$$\xi_t^i = 0 \qquad \forall t \notin [k-1, k),\ k = 1, \ldots, N. \tag{3.5.3}$$

The constraint (3.5.3) reflects the liquidity issue and trading rule of those markets:

- only the first-nearby forwards are liquid;
- a contract that has started to deliver can no longer be traded.

The value process (or the P & L) at time $t \in [k-1, k)$ for $k \in \{1, ..., N\}$ is

$$V_t(\varphi) = \eta_t + \xi_t^\top P_t = \eta_t + \xi_t^k P_t^k = \eta_t + \xi_t^k F(t, k, k+1, X_t).$$

The cumulative cost of the hedge up to time t is:

$$C_t(\varphi) := V_t(\varphi) - G_t(\varphi),$$

where G_t denotes the cumulative gain of the hedge up to time t:

$$\begin{aligned} G_t(\varphi) &= \int_0^t \xi_s^\top dP_s = \sum_{i=1}^{k-1} \int_{i-1}^{i} \xi_s^i dP_s^i + \int_{k-1}^{t} \xi_s^k dP_s^k \\ &= \sum_{i=1}^{k-1} \int_{i-1}^{i} \xi_s^i dF(s, i, i+1, X_s) + \int_{k-1}^{t} \xi_s^k dF(s, k, k+1, X_s) \end{aligned} \tag{3.5.4}$$

[4] Note that for any polynomial processes $p(X_t)$ all moments of $P_t := \mathbb{E}_{\mathbb{Q}}[p(X_T)|\mathcal{F}_t]$ exist. Therefore, integration with respect to any moments of P is well-defined. And thus, $\xi \in L^2(P)$, i.e. $\mathbb{E}_{\mathbb{Q}}[\int_0^T \xi_s^\top d\langle P\rangle_s \xi_s] < \infty$, and $\varphi := (\eta, \xi)^\top$ is admissible.

for $t \in [k-1, k)$.

Note that the market is incomplete under the restriction (3.5.3), since there are two different Brownian motions, but only one risky asset to invest in at any given time. In an incomplete market a claim generally cannot be fully replicated at maturity by a self-financing hedging strategy. Depending on the restriction on cash account η, one can either use a strategy that is self-financing but does not perfectly replicate the claim at maturity, or use a strategy that fully replicates the claim at maturity but needs additional investment throughout the hedge, i.e. is not self-financing. In the first case, we have residual risk and in the latter case additional cash infusion is needed. Either way, risk cannot be fully eliminated and can only be minimized. In the following we briefly review the concept of risk-minimizing strategy in the sense of Föllmer and Schweizer, and then give a rolling hedge that is locally risk-minimizing.

3.5.2 A locally risk minimizing hedging criterion

The risk-minimization criterion proposed and developed by Föllmer and Schweizer (see e.g. Heath et al. [2001], Heath et al. [1999], Schweizer [1999], Schweizer [1990], Föllmer and Schweizer [1991] for details), is to minimize the conditional variance $R_t(\varphi)$ of the cost process $C_t(\varphi)$,

$$R_t(\varphi) := \mathbb{E}_{\mathbb{Q}}\left[(C_T(\varphi) - C_t(\varphi))^2 | \mathcal{F}_t\right],$$

among all not necessarily self-financing strategies φ that perfectly replicate $\widetilde{F}$ at maturity:

$$V_{\widetilde{T}}(\varphi) = \widetilde{F} \quad \mathbb{Q}\text{-a.s.} \tag{3.5.5}$$

In our setup, (3.5.5) is equivalent to $\eta_{T_N} = 0$ and $\xi^N_{T_N} = 1$.

A strategy φ^* is called *risk-minimizing* if for any φ that satisfies (3.5.5) we have $R_t(\varphi^*) \le R_t(\varphi)$, $\mathbb{Q}$-a.s. for every $t \in [0, \widetilde{T}]$; see Schweizer (page 545 in Schweizer [1990]). One can show that any risk-minimizing strategy is mean self-financing, i.e. $C_t(\varphi)$ is a $\mathbb{Q}$-martingale. Föllmer and Schweizer showed that the existence of such a strategy φ is guaranteed if the price process P_t is a $\mathbb{Q}$-local martingale. Moreover, in the martingale case, finding such a strategy is equivalent to finding the Galtchouk–Kunita–Watanabe (GKW) decomposition of $\widetilde{F}$, namely

$$\widetilde{F} = \mathbb{E}[\widetilde{F}] + \int_0^{\widetilde{T}} \widetilde{\xi}_s^{\intercal} dP_s + \widetilde{L}_{\widetilde{T}}, \tag{3.5.6}$$

where $\widetilde{\xi}$ is an admissible, predictable process and $\widetilde{L}$ is a square-integrable $\mathbb{Q}$-martingale strongly orthogonal to P with $\widetilde{L}_0 = 0$. The risk-minimizing hedging strategy φ^{rm} is then given by

$$\varphi_t^{rm} = (\eta_t^{rm}, \xi_t^{rm})^\top = \left(V_t(\varphi^{rm}) - \xi_t^{rm^\top} P_t, \widetilde{\xi}_t\right)^\top,$$

where the value process is $V_t(\varphi^{rm}) = \mathbb{E}[\widetilde{F}|\mathcal{F}_t] = \widetilde{F}_t = \widetilde{F}_0 + \int_0^t \xi_s^{\widetilde{F}^\top} dP_s + \widetilde{L}_t$ and the cost process is $C_t(\varphi^{rm}) = \widetilde{F}_0 + \widetilde{L}_t$. Obviously this risk-minimizing strategy satisfies $V_{\widetilde{T}}(\varphi^{rm}) = \widetilde{F}_{\widetilde{T}}$, and the associated risk process $R_t(\varphi^{rm})$ is minimal (zero) at $t = \widetilde{T}$.

3.5.3 A risk-minimizing rolling hedge

Recall that the price process P_t is a $\mathbb{Q}$-martingale. Then the time-t valuation of the long-term electricity commitment $\widetilde{F}_{\widetilde{T}}$ has a GKW-decomposition as in (3.5.6). We now compute the process $\widetilde{\xi}$ in this decomposition. This will give us the hedging strategy. Using (3.5.6), (3.5.3) and (3.5.4), we obtain for any $t \in [k-1, k)$, $k \in \mathbb{N}$:

$$\langle P^k, \widetilde{F}\rangle_t - \langle P^k, \widetilde{F}\rangle_{k-1} = \int_{k-1}^t d\langle P^k, \int_{k-1}^{\bullet} \widetilde{\xi}_s^k dP_s^k\rangle_u + \int_{k-1}^t d\langle P^k, \widetilde{L}\rangle_u = \int_{k-1}^t \widetilde{\xi}_s^k d\langle P^k, P^k\rangle_s,$$

where $\langle P^k, \widetilde{L}\rangle_t = 0$ as $\widetilde{L}$ is orthogonal to P, and $\langle P^k, \widetilde{F}\rangle_{k-1} = 0$ as $\widetilde{F}_{k-1}$ is constant and known at $t \geq k-1$. Rearranging and using (3.3.10) and (3.3.14) we get the k-th component of $\widetilde{\xi}_t$ for $t \in [k-1, k)$:

$$\begin{aligned}
\widetilde{\xi}_t^k &= \frac{d\langle P^k, \widetilde{F}\rangle_t}{d\langle P^k, P^k\rangle_t} \\
&= \frac{\vec{w}_{01}^\top e^{(\widetilde{T}-t)G^\top} d\langle H(X), H(X)\rangle_t \, e^{(k-t)G}\vec{w}_{01}}{\vec{w}_{01}^\top e^{(k-t)G^\top} d\langle H(X), H(X)\rangle_t \, e^{(k-t)G}\vec{w}_{01}} \\
&= \frac{\vec{w}_{01}^\top e^{(\widetilde{T}-t)G^\top} \Sigma(X_t) \, e^{(k-t)G}\vec{w}_{01}}{\vec{w}_{01}^\top e^{(k-t)G^\top} \Sigma(X_t) \, e^{(k-t)G}\vec{w}_{01}}.
\end{aligned}$$

Therefore, the risk-minimizing hedging strategy of the tradable assets is given by

$$\xi_t^{rm} = \left(\xi_t^{rm,1}, \ldots, \xi_t^{rm,N}\right)^\top, \tag{3.5.7}$$

where

$$\xi_t^{rm,k} = \begin{cases} \widetilde{\xi}_t^k, & \text{for } t \in [k-1,k); \\ 0, & \text{otherwise.} \end{cases} \tag{3.5.8}$$

And thus, for $t \in [k-1,k)$, the cash account η_t^{rm} is then given by

$$\eta_t^{rm} = V_t(\varphi^{rm}) - {\xi_t^{rm}}^\top P_t = \widetilde{F}_t - \widetilde{\xi}_t^k P_t^k,$$

and the associated cost process is

$$C_t(\varphi^{rm}) = \widetilde{F}_t - \int_0^t {\xi^{rm}}^\top dP_s.$$

Remark 3.5.1. *The risk minimizing strategy also minimizes the quadratic covariation between the claim and the value of hedge without the cash account. Indeed, formally one has*

$$\min_\xi d\langle \widetilde{F} - \xi^k P^k \rangle_t = \min_\xi \left(d\langle \widetilde{F} \rangle_t - 2\xi_t^k d\langle \widetilde{F}, P^k \rangle_t + (\xi^k)^2 d\langle P^k \rangle_t \right)$$

This expression is minimized by $\xi_t^k = \frac{d\langle P^k, \widetilde{F} \rangle_t}{d\langle P^k, P^k \rangle_t}$ *as in* (3.5.8).

3.6 Empirical analysis

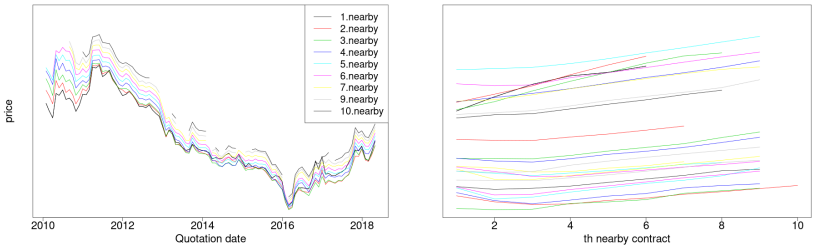
In this section we demonstrate the use of our polynomial framework for modeling and hedging long-term electricity forwards and analyzing their performance. Based on a time series of real observations of power forwards provided by Axpo Solutions AG, we estimate parameters of a model specification. Further we simulate forward curves and investigate the quality of risk-minimizing hedges over various time horizons.

3.6.1 The data

Electricity long-term contracts lack liquidity and are not available on exchange.[5] In fact, long-term forwards with delivery periods are only offered by a small group of market participants over the counter (OTC), mostly by energy producing and trading companies.

[5] People usually refer to contracts with more than 2-3 years time to maturity/start of delivery as *long-term* contracts.

The data we use are provided by Axpo Solutions AG, and come originally from Totem Markit service, which surveys prices of various electricity contracts from each member firm and in turn provides market consensus prices. More concretely, the data are German calender-year baseload (Cal) forwards that are quoted monthly from January 2010 to April 2018.[6] On each quotation date, we have at most 10 quoted contracts, i.e. first to tenth nearby Cal forwards. For each quoted contract on each quotation date, we have consensus price and the price spread between the highest quoted price and the lowest quoted price. A visualization of consensus prices is given in Figure 3.6.1.



(a) rolling forwards (b) a selection of forward curves

Figure 3.6.1: German Calendar-year Baseload forward from January 2010 to April 2018. Y-axes are removed for data protection. Figure (a) shows the dynamics of each nearby Cal forward contract with respect to quotation date. We see that not every contract is available on every quotation date. In Figure (b), each curve is the forward curve of a quotation date, i.e. each curve shows the prices of the first to at most tenth nearby Cal forwards of that date. For the sake of a clearer view, we take a selection of forward curves. These curves (of chosen quotation dates) are stacked and time-lagged into a day. We note two shapes of forward curves: a straight contango curve and a curve which is flat with slight backwardation at the front and contango at the back end of the curve.

3.6.2 Model estimation

In order to capture the dynamics of the forward curves with our model, a non-linear filter is needed for model estimation, as the forward prices are quadratic in the Gaussian underlying factor process X_t. Recall that the fundamental assumption of Kalman filter is that the measurement space is linear and Gaussian in the state space. Thus, in order to work with a Kalman filter, we can either

[6] Note that German Cal Base forwards are the most liquidly traded contracts among all illiquid long-term contracts.

linearize the quadratic relationship between state and measurement. This leads to a so-called extended Kalman filter. Alternatively, we can augment the state to incorporate the linear and quadratic terms of X_t, so that the measurements become linear in the augmented state.

Inspired by the work of Monfort et al. [2015], we will use a time-dependent version of the latter approach to estimate a discrete version of Specification 3.2.1 based on the data from Section 3.6.1. The estimation will be under $\mathbb{P}$, which means that we also need to estimate the market price of risk parameters.

Note that we do not have direct access to the underlying state process X_t through the available data. Indeed, at each quotation date t_k, we only see the prevailing price F_k^j of the j-th nearby forward contract, with $j = 1, \dots, 10$.[7] We view F_k^j as a noisy observation of the model price. More precisely, we assume that

$$F_k^j = F(t_k, T_j, T_j + 1) + N_k^j \eta_k^j,$$

where $F(t_k, T_j, T_j + 1)$ is the model price computed using Proposition 3.3.2, η_k^j are iid standard Gaussian noise, modulated by some parameters $N_k^j > 0$. The role of N_k^j is to encode the trustworthiness of the price of the j-th nearby contract on quotation date t_k. A large value means that the price is considered noisy and uncertain, and a small value that the price is considered accurate. The N_k^j are chosen based on the spreads δ_k^j between the highest and lowest quoted price for the j-th nearby contract on date t_k. Specifically, we use

$$(N_k^j)^2 = \frac{1}{3} \times \delta_k^j + \frac{1}{3} \times \delta^j + \frac{1}{3} \times \delta,$$

where δ^j denotes the time series average of the spreads δ_k^j for a fixed maturity j, and δ denotes the overall average of all the spreads δ_k^j. The use of iid noise corresponds to assuming that our model captures all systematic effects. This is a standard assumption to reduce the complexity of the estimation.

A quadratic Kalman filter for Specification 3.2.1 We will now overload notation in the following manner: we write X_k for the state X_{t_k} at quotation date t_k, and similarly for other quantities that depend on time.

Since model prices at date t_k are quadratic in the state X_k, we have the expression

$$F_k^j = a_k^j + B_k^j X_k + X_k^\top C_k^j X_k + N_k^j \eta_k^j$$

[7] Actually, we see even less, since price data is often missing for longer maturities.

for some $a_k^j \in \mathbb{R}$, $B_k^j \in \mathbb{R}^2$ and $C_k^j \in \mathbb{S}^2$ that can be deduced from the pricing formula in Proposition 3.3.2. In view of (3.3.3), and following Monfort et al. [2015], we observe that F_k^j is affine in the augmented state vector

$$\widetilde{X}_k = (Z_k,\ Y_k,\ Z_k^2,\ Y_k Z_k,\ Y_k^2)^\top.$$

Specifically, the vector of prices, $F_k = (F_k^1, \dots, F_k^{10})^\top$ is given by

$$F_k = a_k + \widetilde{B}_k \widetilde{X}_k + N_k \eta_k,$$

where $a_k := (a_k^1, \dots, a_k^{10})^\top$ and $\widetilde{B}_k := (\widetilde{B}_k^1, \dots, \widetilde{B}_k^{10})^\top$ can be computed as follows: for each maturity $j = 1, \dots, 10$, we have

$$\begin{pmatrix} a_k^j \\ \widetilde{B}_k^j \end{pmatrix} := e^{(T_j - t)G} \int_0^1 e^{uG} du\, \vec{p}_S,$$

with $\vec{p}_S$ from (3.3.4) and G from (3.3.5). Moreover, we have defined $N_k := \operatorname{diag}\left(N_k^1, \dots, N_k^{10}\right)$ and $\eta_k := (\eta_k^1, \dots, \eta_k^{10})^\top$. Next, the discretized (non-augmented) state dynamics is given by

$$X_k = b + D X_{k-1} + K \varepsilon_k$$

where ε_k are independent bi-variate standard Gaussians and

$$b = \begin{pmatrix} \gamma_Z \Delta t \\ \gamma_Y \Delta t \end{pmatrix}, D = \begin{pmatrix} 1 - (\kappa_Z - \lambda_Z)\Delta t & 0 \\ \kappa_Y \Delta t & 1 - (\kappa_Y - \lambda_Y)\Delta t \end{pmatrix}, K = \begin{pmatrix} \sigma_Z \sqrt{\Delta t} & 0 \\ \rho \sigma_Y \sqrt{\Delta t} & \sigma_Y \sqrt{(1-\rho^2)\Delta t} \end{pmatrix}.$$

Here we use the market price of risk parameters $\Lambda = \operatorname{diag}(\lambda_Z, \lambda_Y)$ and $\gamma = (\gamma_Z, \gamma_Y)^\top$ from Section 3.4. The discretized dynamics of the augmented state $\widetilde{X}_k$ is

$$\widetilde{X}_k = \widetilde{b}(X_{k-1}) + \widetilde{D} \widetilde{X}_{k-1} + \widetilde{K}(X_{k-1}) \varepsilon_k,$$

where the involved quantities are conveniently expressed using the standard vector stacking operator $Vec()$, Kronecker product $\otimes$, selection matrix H_d, and duplication matrix G_d. The resulting expressions are:

$$\widetilde{b}(X_{k-1}) = \begin{pmatrix} b \\ H_2 Vec(bb^\top + \Sigma) \end{pmatrix}, \quad \widetilde{D} = \begin{pmatrix} D & 0 \\ H_2(b \otimes D + D \otimes b) G_2 & H_2(D \otimes D) G_2 \end{pmatrix},$$

$$\Gamma_{k-1} = I_2 \otimes (b + D X_{k-1}) + (b + D X_{k-1}) \otimes I_2,$$

$$\widetilde{\Sigma}(X_{k-1}) = \begin{pmatrix} \Sigma & \Sigma \Gamma_{k-1}^\top H_2^\top \\ H_2 \Gamma_{k-1} \Sigma & H_2 \Gamma_{k-1} \Sigma \Gamma_{k-1}^\top H_2^\top + H_2(I_4 + \Lambda_2)(\Sigma \otimes \Sigma) H_2^\top \end{pmatrix},$$

where $\Sigma := KK^\top$ and I_d is the identity matrix of size d, and Λ_m is the standard commutation matrix of size $m^2 \times m^2$. We then let $\widetilde{K}(X_{k-1})$ be the Cholesky factor of $\widetilde{\Sigma}(X_{k-1})$, i.e., $\widetilde{K}(X_{k-1})\widetilde{K}(X_{k-1})^\top = \widetilde{\Sigma}(X_{k-1})$. We finally define $\mathcal{F}_{k-1} := \sigma(F_{k-1}, F_{k-2}, ..., F_1)$. The filtering algorithm is then described in Algorithm 1, where we use the notation

$$\widetilde{X}_{k|k-1} := \mathbb{E}[\widetilde{X}_k|\mathcal{F}_{k-1}], \qquad \widetilde{V}_{k|k-1} := \mathbb{V}[\widetilde{X}_k|\mathcal{F}_{k-1}],$$
$$F^j_{k|k-1} := \mathbb{E}[F^j_k|\mathcal{F}_{k-1}], \qquad M^j_{k|k-1} := \mathbb{V}[F^j_k|\mathcal{F}_{k-1}].$$

Algorithm 1 Quadratic Kalman filtering algorithm

Anchoring:
$\widetilde{X}_{1|1} = \widetilde{x}_0 = (x_0^\top, H_2 Vec(x_0 x_0^\top))^\top = (z_0, y_0, z_0^2, y_0 z_0, y_0^2)^\top,$
$\widetilde{V}_{1|1} = \widetilde{\Sigma}(x_0).$
State prediction:
$\widetilde{X}_{k|k-1} = \widetilde{b}(X_{k-1|k-1}) + \widetilde{D}\widetilde{X}_{k-1|k-1},$
$\widetilde{V}_{k|k-1} = \widetilde{D}\widetilde{V}_{k-1|k-1}\widetilde{D}^\top + \widetilde{\Sigma}(X_{k-1|k-1}).$
Measurement prediction:
$F_{k|k-1} = a_k + \widetilde{B}_k\widetilde{X}_{k|k-1}.$
$M_{k|k-1} = \widetilde{B}_k\widetilde{V}_{k|k-1}\widetilde{B}_k^\top + N_k N_k^\top.$
$\mathcal{C}_k = (F_k^{\text{real}} - F_{k|k-1})$ gives the prediction error.
Update:
$\mathcal{K}_k = \widetilde{V}_{k|k-1}\widetilde{B}_k^\top M^{-1}_{k|k-1}$ gives the gain matrix,
$\widetilde{X}_{k|k} = \widetilde{X}_{k|k-1} + \mathcal{K}_k\mathcal{C}_k,$
$\widetilde{V}_{k|k} = \widetilde{V}_{k|k-1} - \mathcal{K}_k M_{k|k-1}\mathcal{K}_k^\top = (\mathbb{1} - \mathcal{K}_k\widetilde{B}_k)\widetilde{V}_{k|k-1},$
$F^j_{k|k} = a_k + \widetilde{B}_k\widetilde{X}_{k|k}.$

Optimization with the quadratic Kalman filter for Specification 3.2.1 For the model estimation with the quadratic filter, we use both the Least-Squares (LS) and the Maximum Likelihood (ML) criteria. We start with LS, as it is robust and converges fast. Once a stable result is obtained, we apply ML to obtain further improvement. Moreover, we impose $1 \geq \kappa_Y \geq \kappa_Z \geq 0$ on the parameters, in line with the interpretation that Y_t and Z_t drive the short and the long end of the forward curve respectively and thus mean-revert at different

speed. The filtered underlying process $X_t = (Z_t, Y_t)^\top$ is given in Figure 3.6.2. The estimated parameters are shown in Table 3.1.

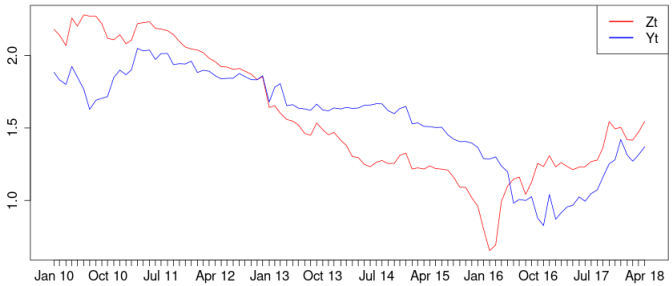


Figure 3.6.2: The filtered underlying dynamics $X_t = (Z_t, Y_t)^\top$ of Specification 3.2.1.

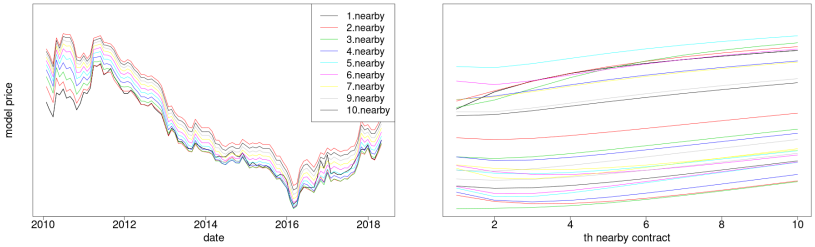
c	0.239614
α	10.250035
β	0.176807
κ_Z	0.010022
κ_Y	0.400207
σ_Z	0.406479
σ_Y	0.889130
ρ	0.112439
λ_Z	0.089990
λ_Y	0.111842
γ_Z	0.086791
γ_Y	0.127365
z_0	2.358048
y_0	2.007557

Table 3.1: Estimated parameters of Specification 3.2.1.

In the implementation we use the R package DEoptim, which is an optimizer based on a differential evolution algorithm; see Storn and Price [1997], Price et al. [2006] for details of the algorithm and `https://cran.r-project.org/web/packages/DEoptim/index.html`, Ardia et al. [2011a], Ardia et al. [2016], Mullen

et al. [2011] Ardia et al. [2011b] for use of the package.

Figure 3.6.3 gives a visualization of the model estimation using Specification 3.2.1. We quantify the goodness of fit in terms of relative errors, both cross-sectionally at each quotation date (Figure 3.6.4a), and across time for each nearby forward contract (Figure 3.6.4b). The overall relative error, i.e. the average relative error across all contracts and quotation dates, is as low as 0.661%, indicating a very good model fit.

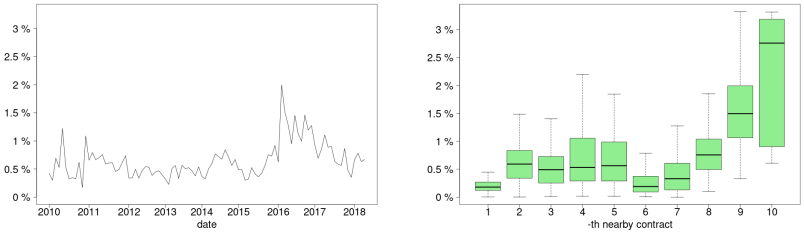


(a) model estimated rolling forwards (b) a selection of estimated forward curves

Figure 3.6.3: Forward curves from Specification 3.2.1 using estimated parameters: in (a) each nearby forward is shown as a time series; in (b) each curve is a forward curve at a particular quotation date (same date selection as in Figure 3.6.1b). Y-axes are removed for data protection. Comparing these figures with the real observations (Figure 3.6.1), we find that the model captures the shapes and dynamics of the time series observation of electricity forward curves well.

In Figure 3.6.4a we notice a single spike of the time series of averaged errors reaching almost 2% (on a quotation date in February 2016). This is due to a single dramatic price drop of a forward curve on that date that is moderately captured by our model as it is continuous and gives smooth prices.

Looking at the estimation of the time series of each nearby forward (Figure 3.6.4b), we find that the front end fit (i.e. the first nearby to the sixth nearby forward contract) works very well while the fit deteriorates for longer maturities. This occurs by construction, as the prices of contracts with very long time-to-maturity are less reliable than those on the front end of the forward curve. In the filter this is captured by the data variance N_t^j, which is influenced by the time series of price spread of each forward; in general N_t^j tends to be higher for longer time-to-maturity (i.e. larger j).



(a) averaged relative errors with respect to quotation date

(b) first to third quantile of time series of relative errors with respect to rolling contract

Figure 3.6.4: Relative errors of model estimation. The overall relative error (averaged over all contracts and all quotation dates) is 0.661%. In (a) the averaged relative error of forward curve on each quotation date is shown. The spike in February 2016 is caused by a large downward drop of the observed forward curve, leading the model to deviate 2% on average on that date. In (b) the distribution of relative pricing errors for each nearby contract over time is given in boxplot: each whisker gives the range from mininum value to maximum value of the time series of relative errors for that contract (outliers are removed). Each green box marks the 25th to 75th percentile of the time series. The thick black line marks the median relative error. In addition to (b), the time averaged relative errors and standard deviations for each contract are given in the table below. We see that the first to sixth nearby contracts are well estimated by the model, while the seventh to tenth nearby contracts have much larger estimation errors. This occurs by construction. The real data on the back end of the forward curve are very rare and thus have a huge price uncertainty; in particular the tenth nearby contract was only available on four quotation dates on over nine years of monthly quotation data. The uncertainty of real data is captured by the parameters N_t^j for each j-th nearby rolling contract in the quadratic filter.

nearbycontract	1	2	3	4	5
av.rel.error	0.2162%	0.6211%	0.5666%	0.7362%	0.6990%
std(*rel.error*)	0.1741%	0.4036%	0.4401%	0.6223%	0.5560%
nearbycontract	6	7	8	9	10
av.relerror	0.3355%	0.4509%	0.8530%	1.6583%	2.1549%
std(*rel.error*)	0.4397%	0.4330%	0.5986%	0.9784%	1.2975%

We also performed model estimation under $\mathbb{Q}$. This is equivalent to assuming $\mathbb{P} = \mathbb{Q}$, meaning that the market price of risk is zero ($\lambda(X_t) = 0$). This produces different parameters than those in Table 3.1, but the fit remains remarkably good.

3.6.3 Simulation and hedging analysis

In the following, we simulate forward surfaces, run locally risk-minimizing hedging strategies on those, and analyze their performance with respect to different hedging horizons.

Simulation of forward surfaces With a given set of parameters, we generate samples of entire forward surfaces over a fixed time horizon $\widetilde{T}$. This can be done efficiently by first simulating the $\mathbb{P}$-dynamics of the underlying process $X_t = (Y_t, Z_t)^\top$ until year $\widetilde{T}$ using a simple Euler scheme (with, say, N discretization steps). We can then compute the forward price for the 1-st through L-th nearby contract at each point $t \le \widetilde{T}$ on the time grid by applying the pricing formula, Proposition 3.3.2. The complexity of simulating M evolutions of forward curves is of the order $\mathcal{O}((M \times N)^L)$. A brief pseudo code is given in Algorithm 2.

Simulation study of hedging performance We aim to evaluate hedging performance by comparing the unhedged exposures with exposures when we use the locally risk-minimizing rolling hedges from Section 3.5 on different hedging horizons. For this, we consider different claims $F(t,T) := F(t,T,T+1,X_t)$ with $T = 2,\dots,10$ years. Next, we simulate $M = 5000$ forward curve evolutions using the estimated parameters from Table 3.1. For the Euler discretization we use 120 time points per year. For the hedging we use a monthly rebalancing frequency. Finally, we compare the percentage exposure if left unhedged, i.e.

$$\frac{F(T,T) - F(0,T)}{F(0,T)},$$

with the percentage exposure if hedged, i.e.

$$\frac{F(T,T) - F(0,T) - \int_0^T \xi_t^{rm\top} dP_s}{F(0,T)},$$

with ξ_s^{rm} from (3.5.7)–(3.5.8) and P_s from (3.5.1)–(3.5.2). A visual comparison of those exposures (hedged versus unhedged) with respect to different hedging horizons is given in Figure 3.6.5. We see that the distribution of the exposure widens with increasing hedging horizon, and that the sample standard deviation and skewness go up; see the table below Figure 3.6.5. The exposure is significantly higher if left unhedged. Moreover, in all cases, the locally risk-minimizing rolling

Algorithm 2 Simulate forward surfaces under $\mathbb{P}$ (with market price of risk)

Input: $\varepsilon_j^Y, \varepsilon_j^Z \overset{iid}{\sim} \mathcal{N}(0,1), j = 1, ..., N, \widetilde{T}, M, N, L$ and all model parameters (see e.g. Table 3.1).
Output: M simulated forward surfaces over $\widetilde{T}$ years.

$\Delta t = T/N$
$Y_0 = y_0$
$Z_0 = z_0$
$H(X_0) = (1, z_0, y_0, z_0^2, y_0 z_0, y_0^2)^\top$
for $l = 1, ..., L$ **do**
  $F_0^l = H(X_0)\, e^{lG}\, \vec{w}_{0,1}$
end for
for all M simulations **do**
  for $j = 1, ..., N$ **do**
    $Z_j = \gamma_Z \Delta t + (1 - (\kappa_Z - \lambda_Z)\Delta t)\, Z_{j-1} + \sigma_Z \sqrt{\Delta t}\, \varepsilon_j^Z$
    $Y_j = \gamma_Y \Delta t + \kappa_Y \Delta t Z_{j-1} + (1 - (\kappa_Y - \lambda_Y)\Delta t) Y_{j-1} + \sigma_Y \sqrt{\Delta t}(\rho \varepsilon_j^Z + \sqrt{1-\rho^2}\varepsilon_j^Y)$
    $H(X_j) = (1, Z_j, Y_j, Z_j^2, Y_j Z_j, Y_j^2)^\top$
    for $l = 1, ..., L$ **do**
      $F_j^l = H(X_j)\, e^{(l-(j\Delta t \mod 1))G}\, \vec{w}_{0,1}$
    end for
  end for
end for

hedge significantly reduces, but does not eliminate, the variance and skew of long-term exposures.

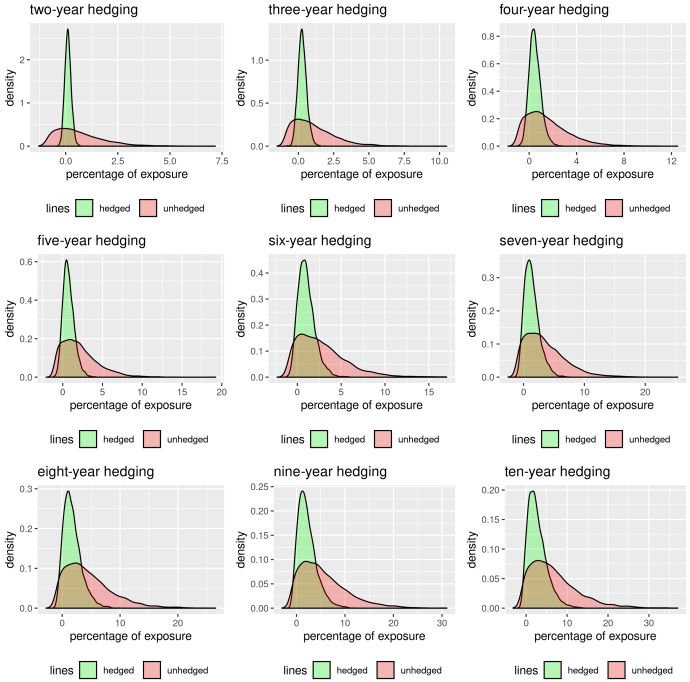


Figure 3.6.5: Density of hedged exposure (green) versus that of unhedged exposure (red) with respect to different hedging horizons. Forwards are simulated using the $\mathbb{P}$-dynamics and market price of risk. In each figure, a different forward is simulated such that the time to maturity corresponds the hedging horizon: i.e. in top left figure, we simulate a forward that matures and starts delivery in two years and compute the exposure at maturity; we then compute a risk-minimizing hedge (with two years hedging horizon), the hedged exposure, and obtain the comparison. Standard deviations and skewnesses are reported in the table below.

hedging horizon	hedged		unhedged	
	std	skew	std	skew
2 years	0.1532	0.2728	1.1278	1.1724
3 years	0.3099	0.3658	1.4700	1.2107
4 years	0.4959	0.5477	1.8143	1.1738
5 years	0.7125	0.6992	2.2762	1.2201
6 years	0.9583	0.8474	2.8011	1.2439
7 years	1.2266	0.9061	3.3729	1.2361
8 years	1.5406	1.0017	4.0898	1.1926
9 years	1.8991	1.0625	4.8472	1.1660
10 years	2.2982	1.0777	5.7729	1.2224

3.A Explicit computation of $\int_0^t e^{Gs} ds$

The G-matrices arising in both specifications have a zero first column, and are therefore not invertible. This is in general the case when 1 is part of the basis $H(x)$, as $\mathcal{G}1 = 0$. Moreover, if we remove the first row and column of G, the submatrix G' is invertible and upper-triangular. In the following we show a straightforward way to compute $\int_0^t e^{Gs} ds$ for such G, which helps to reduce the computational effort of evaluating the pricing formula.

Proposition 3.A.1. *Let A be an upper triangular matrix of the form*

$$A = \begin{pmatrix} 0 & b^\top \\ \vec{0} & C \end{pmatrix}$$

for some vector b and upper triangular invertible matrix C. Then

$$e^{At} = \begin{pmatrix} 1 & b^\top C^{-1}(e^{Ct} - I) \\ \vec{0} & e^{Ct} \end{pmatrix} \quad \text{and} \quad \int_0^t e^{As} ds = \begin{pmatrix} t & b^\top (C^{-1})^2 (e^{Ct} - I) - t b^\top C^{-1} \\ \vec{0} & C^{-1}(e^{Ct} - I) \end{pmatrix}$$

Proof. Let $F(t)$ denote the claimed expression for e^{At}. One easily checks that $F'(t) = AF(t)$ and that $F(0)$ is the identity. This implies that $F(t) = e^{At}$. The expression for $\int_0^t e^{As} ds$ is easily obtained by integrating each block of $F(t)$. □

3.B Specifications of $\Sigma(X_t)$

Instantaneous covariations and correlations in Specification 3.2.1
Equations (3.3.9), (3.3.12), (3.3.11), (3.3.13) hold with H from (3.3.3), $\vec{p}_S$ from (3.3.4) and $\Sigma(X_t)$ as below:

$$\Sigma_{X_t} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_Z^2 & \rho\sigma_Y\sigma_Z & 2\sigma_Z^2 Z_t & \sigma_Z^2 Y_t + \rho\sigma_Y\sigma_Z Zt & 2\rho\sigma_Y\sigma_Z Y_t \\ 0 & \rho\sigma_Y\sigma_Z & \sigma_Y^2 & 2\rho\sigma_Y\sigma_Z Z_t & \sigma_Y^2 Z_t + \rho\sigma_Y\sigma_Z Y_t & 2\sigma_Y^2 Y_t \\ 0 & 2\sigma_Z^2 Z_t & 2\rho\sigma_Y\sigma_Z Z_t & 4\sigma_Z^2 Z_t^2 & 2\sigma_Z^2 Y_t Z_t + 2\rho\sigma_Y\sigma_Z Z_t^2 & 4\rho\sigma_Y\sigma_Z Y_t Z_t \\ 0 & \sigma_Z^2 Y_t + \rho\sigma_Y\sigma_Z Zt & \sigma_Y^2 Z_t + \rho\sigma_Y\sigma_Z Y_t & 2\sigma_Z^2 Y_t Z_t + 2\rho\sigma_Y\sigma_Z Z_t^2 & \sigma_Z^2 Y_t^2 + \sigma_Y^2 Z_t^2 + 2\rho\sigma_Y\sigma_Z Y_t Z_t & 2\rho\sigma_Y\sigma_Z Y_t^2 + 2\sigma_Y^2 Y_t Z_t \\ 0 & 2\rho\sigma_Y\sigma_Z Y_t & 2\sigma_Y^2 Y_t & 4\rho\sigma_Y\sigma_Z Y_t Z_t & 2\rho\sigma_Y\sigma_Z Y_t^2 + 2\sigma_Y^2 Y_t Z_t & 4\sigma_Y^2 Y_t^2 \end{pmatrix}$$

Instantaneous covariations and correlations in Specification 3.2.2
Equations (3.3.9), (3.3.12), (3.3.11), (3.3.13) hold with H from (3.3.6), $\vec{p}_S$ from (3.3.7) and $\Sigma(X_t)$ as below:

$$\Sigma_{X_t} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_Z^2 & \sigma_Y \sigma_Z R_t & 0 & 2\sigma_Z^2 Z_t & \sigma_Z^2 Y_t + \sigma_Y \sigma_Z R_t Zt & 2\sigma_Y \sigma_Z R_t Y_t \\ 0 & \sigma_Y \sigma_Z R_t & \sigma_Y^2 & 0 & 2\sigma_Y \sigma_Z R_t Z_t & \sigma_Y^2 Z_t + \sigma_Y \sigma_Z R_t Y_t & 2\sigma_Y^2 Y_t \\ 0 & 0 & 0 & \sigma_R^2(1-R_t^2) & 0 & 0 & 0 \\ 0 & 2\sigma_Z^2 Z_t & 2\sigma_Y \sigma_Z R_t Z_t & 0 & 4\sigma_Z^2 Z_t^2 & 2\sigma_Z^2 Y_t Z_t + 2\sigma_Y \sigma_Z R_t Z_t^2 & 4\sigma_Y \sigma_Z R_t Y_t Z_t \\ 0 & \sigma_Z^2 Y_t + \sigma_Y \sigma_Z R_t Z_t & \sigma_Y^2 Z_t + \sigma_Y \sigma_Z R_t Y_t & 0 & 2\sigma_Z^2 Y_t Z_t + 2\sigma_Y \sigma_Z R_t Z_t^2 & \sigma_Z^2 Y_t^2 + \sigma_Y^2 Z_t^2 + 2\sigma_Y \sigma_Z R_t Y_t Z_t & 2\sigma_Y \sigma_Z R_t Y_t^2 + 2\sigma_Y^2 Y_t Z_t \\ 0 & 2\sigma_Y \sigma_Z R_t Y_t & 2\sigma_Y^2 Y_t & 0 & 4\sigma_Y \sigma_Z R_t Y_t Z_t & 2\sigma_Y \sigma_Z R_t Y_t^2 + 2\sigma_Y^2 Y_t Z_t & 4\sigma_Y^2 Y_t^2 \end{pmatrix}$$

3.C Correlation of forwards implied by the data

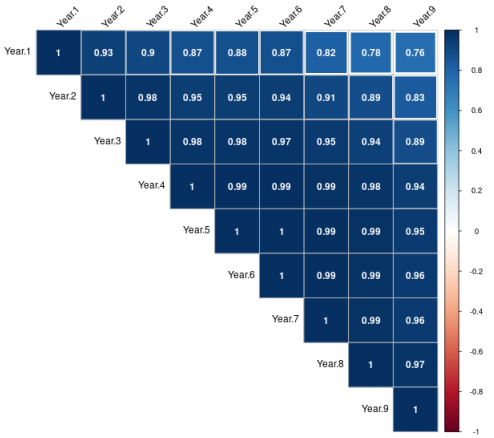


Figure 3.C.1: Correlation between different nearby Calender year contracts implied by the data.

Chapter 4

Quadratic Gaussian models and option pricing of electricity forwards with delivery period

4.1 Introduction

In the last chapter, we have shown that the polynomial framework is very tractable, and is able to capture the underlying dynamics of electricity forwards with delivery period easily and accurately. It is natural then to investigate the option pricing of those forwards. Within the polynomial framework, it is possible to use the polynomial expansion method to price options, see Remark 3.3.3. However, this is rather a theoretical result, and does not fulfill the needs of an applicable approach for the option pricing. In this chapter, we focus on a subclass of the polynomial framework, the *quadratic Gaussian model*, for which long-term electricity options that are written on forwards with delivery period can be computed and calibrated easily. For the quadratic Gaussian model, we develop an exponential-quadratic transform formula, which allows us to compute the characteristic function by solving Riccati equations. And thus, we can price

options efficiently using a Fourier approach. Our option pricing presents a model-consistent way of pricing forward and option prices, and it is able to capture characteristics such as the Samuelson's effect to certain extent. Moreover, as the quadratic Gaussian model generalizes the two-factor model, Specification 3.2.1, it can capture properties of electricity forwards very well.

The biggest challenge in modeling electricity forwards and options comes from the fact that electricity is non-storable and the contracts have delivery periods. This substantially increases the complexity of the modeling, and at the same time, it reduces the choice of models with good mathematical tractability. Various models have been proposed to model electricity forwards and options; see in particular Benth and Schmeck [2014], Benth et al. [2008a], Benth and Krühner [2015], Bjerksund et al. [2010], Burger et al. [2004], Carmona and Durrleman [2003], Kiesel et al. [2009], Kluge [2006], where features such as delivery periods are addressed, and implementation and calibration of options written on forwards with delivery period are discussed. A stream of the literature suggests to price energy options via Monte Carlo simulation. The majority of literature makes simplifying assumptions on the distribution of the underlying forwards for the sake of tractable closed-form formula for the option pricing. In detail, they assume log-normal distributed forwards with delivery period, and compute approximation of option prices via the Black-Scholes-76 formula; see in particular Kiesel et al. [2009], Bjerksund et al. [2010]. The strength of our approach is that we propose a consistent way of pricing forwards and options without approximating assumptions. Moreover, we calibrate simultaneously forwards with delivery periods and European-style options that are written on these forwards. Furthermore, our approach can be extended to price and calibrate spread options, which are popular types of energy options.

This chapter is structured in the following way: In Section 2, we define the underlying quadratic Gaussian model and review the two-factor model as a specification of it. We model the spot price as a quadratic function of the underlying quadratic Gaussian model, and show that spot has a unique representation. In Section 3, we define electricity forwards with delivery period and options on those forwards. We give pricing formulas for forwards and for options. The pricing formula of forward uses the moment formula of degree two, Theorem 2.3.2. The pricing formula of options, Theorem 4.3.4, is a Fourier-based approach and relies on the main result of this chapter, the *exponential-quadratic transform formula* in Theorem 4.4.1, which we give in Section 4. The exponential-quadratic transform formula is of interest in its own, as it allows us to compute

the characteristic function by solving a system of Riccati equations. Moreover, this formula holds for all $x \in \mathbb{R}^d$ if one of the two side is well-defined. We also discuss the analytic solutions of the Riccati equations. In Section 5, we give the proof of the exponential-quadratic transform formula. The proof is inspired by the proof of affine transform formula, Theorem 10.3 of Filipovic [2009]. In Section 6, we conduct a calibration study using the market data of European Energy Exchange (EEX). Here we calibrate the two-factor model to several traded German electricity volatility curves that are written on calendar year forwards, and investigate the goodness of fit of the model.

This chapter is based on joint work with Damir Filipović and Martin Larsson. We would like to thank Markus Regez for providing data and for discussion.

4.2 The model

In this section we define the underlying *quadratic Gaussian model*. We model the spot price S_t as a quadratic function of an underlying d-dimensional state variable X_t which is a $\mathbb{R}^d$-valued Ornstein–Uhlenbeck process. More precisely, we let

$$S_t = p_S(X_t) \tag{4.2.1}$$

$$dX_t = \kappa(\theta - X_t)dt + \sigma dW_t \tag{4.2.2}$$

where $p_S(x) = c + x^\top Q x$ with $c \in \mathbb{R}_+$ and $Q \in \mathbb{S}^d_+$, $\kappa \in \mathbb{R}^{d\times d}$, $\theta \in \mathbb{R}^d$, $\sigma \in \mathbb{R}^{d\times d}$ so that $a = \sigma\sigma^\top \in \mathbb{S}^d_+$ and W a d-dimensional Brownian motion. Note that $p_S(X_t)$ is a positive quadratic polynomial, bounded from below by c.

In the model definition above, we note that the model is not uniquely determined by the choice of parameters. The following theorem addresses the problem of parameter redundancy and gives a model formulation which yields a unique parameter representation of a quadratic Gaussian model.

Theorem 4.2.1. *Any d-dimensional quadratic Gaussian model that yields S_t bounded from below can be represented as (4.2.1)–(4.2.2) with*

$$p_S(x) = c + x_1^2 + x_2^2 + \ldots + x_k^2, \qquad k \le d,\ c \in \mathbb{R} \tag{4.2.3}$$

Proof. Firstly, we claim that $p(x)$ can be expressed as

$$p(x) = c + (x - \theta)^\top Q(x - \theta). \tag{4.2.4}$$

for some positive constant $c, \theta \in \mathbb{R}$ and $Q \in \mathbb{S}_+^d$. Note that any (positive) quadratic polynomial q on $\mathbb{R}^d$ can be written $p(x) = \tilde{c} + \xi^\top x + x^\top Q x$ for some $(\tilde{c}, \xi, Q) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$. By completing the square, we see that for any $\theta \in \mathbb{R}^d$,

$$p(x) = c + (x-\theta)^\top Q(x-\theta) + (\xi - 2Q\theta)^\top x$$

with $c = \tilde{c} + \theta^\top Q\theta$. We would like choose θ so that the last term vanishes. To see that this is possible, we pick any $\widehat{\xi} \in \mathrm{Ker}(Q)$ and any $t \in \mathbb{R}$. Then $p(t\widehat{\xi}) = c + t\,\xi^\top \widehat{\xi}$, which implies that $\xi^\top \widehat{\xi} = 0$ since q was assumed positive. Hence ξ is orthogonal to $\mathrm{Ker}(Q) = \mathrm{Im}(Q)^\perp$, showing that a vector θ of the desired form exists. It remains to prove that $Q \in \mathbb{S}_+^d$. If it fails, there is an eigenvector v with eigenvalue $\lambda < 0$. In this case $p(tv + \theta) = c + t^2\lambda\|v\|^2$, which is unbounded below. This contradicts the positivity assumption and proves that a representation as in (4.2.4) exists.

Since $Q \in \mathbb{S}^d$ is symmetric, it is orthogonally diagonalizable, i.e. $Q = U\Gamma U^\top$ for some orthogonal matrix U, and $\Gamma = \mathrm{diag}(\gamma_1, \dots, \gamma_d)$ with eigenvalues $\gamma_i \geq 0$; see [Horn and Johnson, 2012, Theorem 2.5.6]. It is clear that the affine transformation $x \mapsto U^\top(x - \theta)$ leaves the form (4.2.2) of the state process invariant. We can thus assume that $p(x) = c + x^\top \Gamma\, x$. Finally, transforming the factor process according to $x \mapsto \Gamma^{1/2} x$ then gives the claim in (4.2.3). □

Below we review the two-factor model of the polynomial framework, Specification 3.2.1, as it is also a specification of the quadratic Gaussian model.

Specification 4.2.2 (Two-factor model)**.** *Let κ_Z, $\kappa_Y \in \mathbb{R}$, $\sigma_Z, \sigma_Y > 0$, and $\rho \in (-1, 1)$. The process $X_t := (Z_t, Y_t)^\top$ evolves according to the SDE*

$$\begin{aligned} dZ_t &= -\kappa_Z Z_t dt + \sigma_Z dW_t^{(1)} \\ dY_t &= \kappa_Y (Z_t - Y_t)dt + \rho\sigma_Y dW_t^{(1)} + \sigma_Y\sqrt{1-\rho^2}\, dW_t^{(2)} \end{aligned} \tag{4.2.5}$$

with $Z_0, Y_0 \in \mathbb{R}$ and $W_t = (W_t^{(1)}, W_t^{(2)})^\top$ a standard two-dimensional Brownian motion. Let α, β, $c \in \mathbb{R}_+$ and let the spot price be given by

$$S_t := c + X_t^\top \begin{pmatrix} \beta & 0 \\ 0 & \alpha \end{pmatrix} X_t = c + \alpha Y_t^2 + \beta Z_t^2.$$

This specification is of the form (4.2.2)–(4.2.3) with $\widehat{X}_t = (\widehat{Z}_t, \widehat{Y}_t)^\top = (\sqrt{\beta} Z_t, \sqrt{\alpha} Y_t)^\top$ and

$$\kappa = \begin{pmatrix} \kappa_Z & 0 \\ -\frac{\sqrt{\alpha\beta}}{\beta}\kappa_Y & \kappa_Y \end{pmatrix}, \quad \theta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \sigma = \begin{pmatrix} \sigma_Z\sqrt{\beta} & 0 \\ \rho\sigma_Y\sqrt{\alpha} & \sigma_Y\sqrt{\alpha(1-\rho^2)} \end{pmatrix}. \tag{4.2.6}$$

X_t is a polynomial diffusion, and S_t is a polynomial function of X_t of degree two. In other words, the quadratic Gaussian models form a subclass of a polynomial framework. As a consequence, the moment formulas and other important properties of polynomial diffusions apply. In particular, the moment formula of degree two, Theorem 2.3.2, is tailored for the quadratic Gaussian model, where the coefficients with respect to each element of the monomial basis is given as the solutions of a system of linear ODEs.

4.3 Pricing of forwards and options

In this section we first define electricity forwards with and without delivery period, and define European call options on them. We then present pricing formulas for electricity forwards and options. While the forward pricing can be obtained immediately from the moment formula, the option pricing relies on the Fourier transform method and the characteristic function of S_t. In the next section we will provide the main mathematical result for Fourier pricing, an exponential-quadratic transform formula, which links the computation of the characteristic function of S_t to the unique solution of a Riccati-system.

The price at time-t of an electricity forward with instantaneous delivery at time $\widetilde{T} \geq t$ is given by

$$f(t, \widetilde{T}, X_t) := \mathbb{E}_{\mathbb{Q}}\left[S_{\widetilde{T}} \middle| \mathcal{F}_t\right]. \tag{4.3.1}$$

The price of a European call option at time t, written on this forward, where the option is exercised at time T $(t \leq T \leq \widetilde{T})$ and at strike K, is given by

$$C^f(t, X_T; K, \widetilde{T}) := \mathbb{E}_{\mathbb{Q}}\left[p^f_{\text{call}}(X_T; K, \widetilde{T})^+ \middle| \mathcal{F}_t\right],$$

where $p^f_{\text{call}}(X_T; K, \widetilde{T}) := f(T, \widetilde{T}, X_T) - K$.

In practice, electricity is not delivered instantaneously, but gradually over a period of time. This leads us to the following definition: the time-t price of an electricity forward with delivery period $[T_1, T_2)$, $t \leq T_1 < T_2$, is given by

$$F(t, T_1, T_2, X_t) := \frac{1}{T_2 - T_1}\mathbb{E}_{\mathbb{Q}}\left[\int_{T_1}^{T_2} S_u \, du \middle| \mathcal{F}_t\right]. \tag{4.3.2}$$

The time-t price of an European call option on this forward, where the option is exercised at time T $(t \leq T \leq T_1 < T_2)$ and at strike K, is given by

$$C^F(t, X_T; K, T_1, T_2) := \mathbb{E}_{\mathbb{Q}}\left[p^F_{\text{call}}(X_T; K, T_1, T_2)^+ \middle| \mathcal{F}_t\right],$$

where $p^F_{\text{call}}(X_T; K, T_1, T_2) := F(T, T_1, T_2, X_T) - K$.

Proposition 4.3.1 (Pricing formula for forwards)**.** *The time-t price of $f(t, \widetilde{T}, X_t)$ for $t \le \widetilde{T}$ is given by*

$$f(t, \widetilde{T}, X_t) = c + A(\widetilde{T} - t) + B(\widetilde{T} - t)^\top X_t + X_t^\top C(\widetilde{T} - t) X_t,$$

and the time-t price of $F(t, T_1, T_2, X_t)$ for $t \le T_1 < T_2$ is

$$\begin{aligned} F(t, T_1, T_2, X_t) =& c + \frac{1}{T_2 - T_1}\Big[\int_{T_1}^{T_2} A(s-t)ds + \int_{T_1}^{T_2} (B(s-t)ds)^\top X_t \\ &+ X_t^\top \int_{T_1}^{T_2} C(s-t)ds X_t\Big]. \end{aligned}$$

where $(A(t), B(t), C(t))$ solves the linear ODE (2.3.6) with the initial conditions:

$$A(0) = 0, \qquad B(0) = \vec{0}, \qquad C(0) = Q. \tag{4.3.3}$$

Proof. This follows from Theorem 2.3.2. □

Remark 4.3.2. *It is also possible to compute the forward with the general version moment formula, Theorem 2.3.1, where spot and forwards are given as matrix representations w.r.t a basis, and pricing relies on the explicit computation of integral of matrix exponential where the matrix is not invertible; for details see the pricing formula Proposition 3.3.2 ([Kleisinger-Yu et al., 2020, Proposition 4.2]).*

As an immediate consequence, the time-T price of $p^f_{\text{call}}(X_T; K, \widetilde{T})$ for $T \le \widetilde{T}$ is given by

$$p^f_{\text{call}}(X_T; K, \widetilde{T}) = f(T, \widetilde{T}, X_T) - K = \tilde{u} + \tilde{v}^\top X_T + X_T^\top \tilde{w} X_T,$$

with

$$\tilde{u} := c - K + A(\widetilde{T} - T), \qquad \tilde{v} := B(\widetilde{T} - T), \qquad \tilde{w} := C(\widetilde{T} - T),$$

and the time-T price of $p^F_{\text{call}}(X_T; K, T_1, T_2)$ for $T \le T_1 < T_2$ is

$$p^F_{\text{call}}(X_T; K, T_1, T_2) = F(T, T_1, T_2, X_T) - K = \tilde{u} + \tilde{v}^\top X_T + X_T^\top \tilde{w} X_T,$$

with

$$\begin{aligned}\tilde{u} &:= c - K + \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} A(s-T)ds,\\ \tilde{v} &:= \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} B(s-T)ds,\\ \tilde{w} &:= \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} C(s-T)ds,\end{aligned} \tag{4.3.4}$$

where $A(t), B(t)$ and $C(t)$ solve the linear ODE (2.3.6) with the initial conditions (4.3.3).

In case of the two-factor model, Specification 4.2.2, the linear ODE (2.3.6) simplifies to the following:

$$\begin{aligned} A' &= \mathrm{Tr}(a\pi), & A(0) &= 0,\\ B' &= -\kappa^\top B, & B(0) &= \vec{0},\\ C' &= -C\kappa - \kappa^\top C, & C(0) &= Q. \end{aligned}$$

Here we note that $B(t) = B(0)e^{-\kappa^\top t} = 0$ for any t. Thus, $\tilde{v} = 0$ which simplifies (4.3.4), and

$$p^F_{\mathrm{call}}(X_T; K, T_1, T_2) = F(T, T_1, T_2, X_T) - K = \tilde{u} + X_T^\top \tilde{w} X_T, \tag{4.3.5}$$

The computations of $\mathbb{E}_{\mathbb{Q}}\left[p^f_{\mathrm{call}}(X_T)^+ \middle| \mathcal{F}_t\right]$ and $\mathbb{E}_{\mathbb{Q}}\left[p^F_{\mathrm{call}}(X_T)^+ \middle| \mathcal{F}_t\right]$ appear to be a challenging problem in general, as it is difficult to identify the region on which the inner functions remain positive. In the following we use an approach based on Fourier transform methods where we derive an exponential-quadratic transform formula to compute the characteristic function by solving Riccati equations. For the ease of notion, we will omit the superscripts f and F, the temporal parameters as well as the strikes in the following.

Proposition 4.3.3. *For any $\zeta > 0$, the following identity*

$$\tilde{s}^+ = \frac{1}{2\pi}\int_{\mathbb{R}} e^{(\zeta + i\lambda)\tilde{s}} \frac{1}{(\zeta + i\lambda)^2} d\lambda$$

holds.

Proof. Let us assume w.l.o.g. $\tilde{s} = s - k$ for some $s, k \in \mathbb{R}$. For $z \in \mathbb{C}$ with $\mathrm{Im}(z) < 0$, the general Fourier transformation of $\tilde{s}^+$ is given by

$$\begin{aligned}\int_{-\infty}^{\infty} e^{izk}\tilde{s}^+\,dk &= \int_{-\infty}^{s} e^{izk}(s-k)\,dk \\ &= \int_{-\infty}^{s} e^{izk}s\,\mathrm{d}k - \int_{-\infty}^{s} e^{izk}k\,dk \\ &= \frac{e^{izk}}{iz}s\Big|_{k=-\infty}^{s} - \frac{e^{izk}}{iz}k\Big|_{k=-\infty}^{s} + \int_{-\infty}^{s}\frac{e^{ikz}}{iz}\,dk \\ &= \frac{e^{izk}}{(iz)^2}\Big|_{k=-\infty}^{s} \\ &= -\frac{e^{izs}}{z^2}.\end{aligned}$$

The assumption $\mathrm{Im}(z) < 0$ is necessary to prevent the integrals to explode. Then for $z \in \mathbb{C}$ with $\mathrm{Im}(z) < 0$, the generalized Fourier inverse is given by

$$\begin{aligned}\tilde{s}^+ &= \mathcal{F}^{-1}\left[-\frac{e^{izs}}{z^2}\right] \\ &= -\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izk}\frac{e^{izs}}{z^2}\,dz \\ &= -\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iz\tilde{s}}}{z^2}\,dz,\end{aligned}$$

where $dz = d[\mathrm{Re}(z)]$. Now set $\mathrm{Re}(z) := \lambda$ and $\mathrm{Im}(z) := -\zeta$. Then for $\zeta > 0$,

$$\begin{aligned}\tilde{s}^+ &= -\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i(-i\zeta+\lambda)\tilde{s}}\frac{1}{((-i\zeta+\lambda))^2}\,\mathrm{d}\lambda \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{(\zeta+i\lambda)\tilde{s}}\frac{1}{(\zeta+i\lambda)^2}\,\mathrm{d}\lambda.\end{aligned}$$

This is the claimed formula. □

Theorem 4.3.4. *Let us define*

$$\widehat{q}_{\mathrm{call}}(z) := \mathbb{E}\left[\exp\left(zp_{\mathrm{call}}(X_T)\right) \mid \mathcal{F}_t\right],$$

for every $z \in \mathbb{C}$ for which the expectation is well-defined. Pick any $\zeta > 0$ such that $\widehat{q}_{\mathrm{call}}(\zeta) < \infty$. Then the time-t call price is given by

$$C(t, X_T) = \frac{1}{\pi}\int_0^{\infty} \mathrm{Re}\left[\frac{\widehat{q}_{\mathrm{call}}(\zeta + \mathrm{i}\lambda)}{(\zeta + \mathrm{i}\lambda)^2}\right] d\lambda.$$

Proof. Let $q(ds)$ denote the distribution of the random variable $p_{\text{call}}(X_T)$ conditionally on $\mathcal{F}_t$, and define its characteristic function[1]:

$$\widehat{q}_{\text{call}}(z) = \int_{\mathbb{R}} e^{zs} q(ds).$$

for every $z \in \mathbb{C}$ such that the right side is well-defined and finite. Pick $\zeta > 0$ such that $\int_{\mathbb{R}} \mathrm{e}^{\zeta s} q(ds) < \infty$. Then,

$$\begin{aligned} \int_{\mathbb{R}^2} \left| \mathrm{e}^{(\zeta+\mathrm{i}\lambda)s} \frac{1}{(\zeta+\mathrm{i}\lambda)^2} \right| d\lambda \otimes q(ds) &= \int_{\mathbb{R}^2} \frac{\mathrm{e}^{\zeta s}}{\zeta^2+\lambda^2} d\lambda \otimes q(ds) \\ &= \int_{\mathbb{R}} \mathrm{e}^{\zeta s} q(ds) \int_{\mathbb{R}} \frac{1}{\zeta^2+\lambda^2} d\lambda < \infty, \end{aligned} \tag{4.3.6}$$

where the second equality follows from Tonelli's theorem. Together with Fubini's theorem and Proposition 4.3.3, we obtain

$$\begin{aligned} \mathbb{E}\left[p_{\text{call}}(X_T)^+ \mid \mathcal{F}_t\right] &= \int_{\mathbb{R}} s^+ q(ds) \\ &= \int_{\mathbb{R}} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{(\zeta+\mathrm{i}\lambda)s} \frac{1}{(\zeta+\mathrm{i}\lambda)^2} d\lambda \right) q(ds) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\widehat{q}_{\text{call}}(\zeta+\mathrm{i}\lambda)}{(\zeta+\mathrm{i}\lambda)^2} d\lambda \\ &= \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{\widehat{q}_{\text{call}}(\zeta+\mathrm{i}\lambda)}{(\zeta+\mathrm{i}\lambda)^2} \right] d\lambda, \end{aligned}$$

where the last equality uses that the left, and hence right, side is real, together with the observation that the real part of $(\zeta+\mathrm{i}\lambda)^{-2}\widehat{q}_{\text{call}}(\zeta+\mathrm{i}\lambda)$ is an even function of λ. □

4.4 Exponential-quadratic transform formula

The option valuation is now reduced to the computation of the characteristic function $\widehat{q}_{\text{call}}(z)$. In this section, we discuss how the computation of the characteristic function can be transformed to the solutions of a system of Riccati equations. The exponential-quadratic transform formula, Theorem 4.4.1. It is the main theorem of this chapter, and it links the exponential of the electricity

[1] Note that characteristic function is normally defined on purely imaginary number $i\mathbb{R}$. Here we define it only on all complex number z.

forward or spot to the solution of a system of Riccati equations. It is valid if either the exponential of the forward is finite (in L^1) or the maximal lifetime of the Riccati equations are not reached. Moreover, we discuss the explicit local solution of the Rccati system.

Theorem 4.4.1. *For any $u \in \mathbb{C}$, $v \in \mathbb{C}^d$, $w \in \mathbb{S}^d + \mathrm{i}\mathbb{S}^d$, consider the following system of ODE with initial condition:*

$$\begin{aligned} \phi'(t) &= \psi(t)^\top \kappa\theta + \frac{1}{2}\psi(t)^\top a\psi(t) + \mathrm{Tr}(a\pi(t)), & \phi(0) &= u, \\ \psi'(t) &= -\kappa^\top \psi(t) + 2\pi(t)a\psi(t) + 2\pi(t)\kappa\theta, & \psi(0) &= v, \\ \pi'(t) &= -\pi(t)\kappa - \kappa^\top \pi(t) + 2\pi(t)a\pi(t), & \pi(0) &= w. \end{aligned} \tag{4.4.1}$$

There exists a maximal lifetime $T^*(u,v,w) = T^*(w)$, such that a unique solution exists on $[0, T^*(w))$.[2] *Further, let X_t satisfy (4.2.2), then for any $T \geq 0$, the following conditions are equivalent:*

(i) *The finiteness condition*

$$\mathbb{E}_x\left[\left|\mathrm{e}^{u+v^\top X_T + X_T^\top w X_T}\right|\right] < \infty$$

holds for all $x \in \mathbb{R}^d$.

(ii) *The inequality $T < T^*(\mathrm{Re}(w))$ holds.*

If either condition is satisfied, the following equality

$$\mathbb{E}_x\left[\mathrm{e}^{u+v^\top X_T + X_T^\top w X_T}\right] = \mathrm{e}^{\phi(T)+\psi(T)^\top x + x^\top \pi(T) x} \tag{4.4.2}$$

holds for all $x \in \mathbb{R}^d$.

The proof of the theorem is given in Section 4.5. It closely follows the arguments in [Filipovic, 2009, Chapter 10], but is simplified slightly by the observation that the Riccati system (4.4.1) admits an explicit (local) solution. Since this will also be useful for designing efficient methods to solve the system numerically, we discuss in the following the explicit solution.

[2] Standard ODE theory (see e.g. [Filipovic, 2009, Lemma 10.1]) implies that each initial condition (u,v,w) has an associated maximal lifetime, prior to which a unique solution exists. Since the expression for ψ' is linear in ψ, and since the expression for ϕ' does not involve ϕ at all, it follows that solutions ϕ and ψ exist whenever a solution π exists. The lifetime thus only depends on w, thus $T^*(u,v,w) = T^*(w)$. The maximality of $T^*(w)$ means that $\lim_{t \uparrow T^*(w)} \|\pi(t)\| = \infty$ if $T^*(w) < \infty$.

Explicit solution of the Riccati sytem

Following Reid [1972], one considers the ODE

$$\begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix}\begin{pmatrix} U' \\ V' \end{pmatrix} = \begin{pmatrix} 0 & -\kappa^\top \\ -\kappa & 2a \end{pmatrix}\begin{pmatrix} U \\ V \end{pmatrix},$$

with $U(0) = \mathrm{Id}$ and $V(0) = w$. It is easy to verify that $\pi(t) := V(t)U(t)^{-1}$ satisfies the third equation in (4.4.1) for all t near which $U(t)$ is invertible. Moreover, the above system is linear with constant coefficients, so one readily computes:

$$U(t) = \mathrm{e}^{\kappa t}\left(\mathrm{Id} - 2\int_0^t \mathrm{e}^{-\kappa s} a\, \mathrm{e}^{-\kappa^\top s} ds\, w\right), \qquad V(t) = \mathrm{e}^{-\kappa^\top t} w. \tag{4.4.3}$$

Consequently, in case w is invertible, one obtains the formula

$$\pi(t) = \mathrm{e}^{-\kappa^\top t}\left(w^{-1} - 2\int_0^t \mathrm{e}^{-\kappa s} a\, \mathrm{e}^{-\kappa^\top s} ds\right)^{-1} \mathrm{e}^{-\kappa t}. \tag{4.4.4}$$

Next, with the Ansatz $\psi(t) = \pi(t)f(t)$, the problem of finding ψ is reduced to solving the linear equation $-\kappa f(t) + f'(t) = 2\kappa\theta$, $f(0) = w^{-1}v$. Doing this, one obtains

$$\psi(t) = \pi(t)\mathrm{e}^{\kappa t}\left(w^{-1}v + 2\int_0^t \mathrm{e}^{-\kappa s} ds\, \kappa\theta\right). \tag{4.4.5}$$

To get ϕ one simply integrates the first equation in (4.4.1). Of course, (4.4.4) and (4.4.5) are only valid prior to the first t for which $U(t)$ becomes singular, which is always strictly positive.

The explicit expressions (4.4.4) and (4.4.5) clarify how the solution depends on the initial conditions, and how it depends on the process parameters. As we will see in the context of model calibration to a time series of call option values, it is crucial to be able to compute the solution at a fixed time T, for a fixed set of process parameters, but for a variety of initial conditions. In this case components of the solution can be precomputed. Specifically, the integrals in (4.4.4) and (4.4.5) do not depend on the initial conditions. Moreover, if $-\kappa$ is nonsingular we have

$$\int_0^t \mathrm{e}^{-\kappa s} ds = -\kappa^{-1}(\mathrm{e}^{-\kappa t} - \mathrm{Id}).$$

Furthermore, differentiation yields

$$\frac{d}{dt}\left(\mathrm{e}^{-\kappa t} a\, \mathrm{e}^{-\kappa^\top t}\right) = -\kappa \mathrm{e}^{-\kappa t} a\, \mathrm{e}^{-\kappa^\top t} - \mathrm{e}^{-\kappa t} a\, \mathrm{e}^{-\kappa^\top t}\kappa^\top,$$

and by integrating both sides from 0 to t we find that $I(t) := \int_0^t \mathrm{e}^{-\kappa s} a \, \mathrm{e}^{-\kappa^\top s} ds$ solves the linear matrix equation

$$-\kappa I(t) - I(t)\kappa^\top = \mathrm{e}^{-\kappa t} a \, \mathrm{e}^{-\kappa^\top t} - a. \tag{4.4.6}$$

By computing these quantities for $t = T$ and using (4.4.4)–(4.4.5), $\psi(T)$ and $\pi(T)$ can found directly. To calculate $\phi(T)$, numerical integration is needed, and this necessitates calculating $\psi(t)$ and $\pi(t)$ on a grid of t-values in $[0, T]$. However, by using quadrature methods the number of grid points can be taken very small (on the order of five or ten) while retaining good accuracy. It is then typically more efficient to compute $\psi(t)$ and $\pi(t)$ using the above explicit expressions, than to solve the full Riccati system using a general purpose ODE solver.

Let us finally remark that if (u, v, w) are not real, then (i) (and hence (ii)) of Theorem 4.4.1 may fail, while at the same time there is a global solution to the Riccati system. As an example, consider the case $d = 1$, $\theta = 0$, $-\kappa = 0$, $a = \frac{1}{2}$, so that $X_t = x + (1/\sqrt{2})W_t$. Letting $u = 0$ and $v = 0$, the Riccati system has the explicit solution

$$\pi(t) = \frac{1}{w^{-1} - t}, \qquad \psi(t) = 0, \qquad \phi(t) = -\frac{1}{2}\ln(1 - wt), \qquad t < T^*(w),$$

where $T^*(w)$ is the first time $t = w^{-1}$. If w has a nonzero imaginary part this never happens, and so a global solution exists. On the other hand, the expectation $\mathbb{E}_x[|\mathrm{e}^{wX_t}|]$ is only finite for $t < 1/\mathrm{Re}(w)$. For these values of t, therefore, the equality

$$\mathbb{E}_x[\mathrm{e}^{wX_t}] = \frac{1}{\sqrt{1 - wt}} \exp\left(\frac{x^2}{w^{-1} - t}\right)$$

holds. For $t \geq 1/\mathrm{Re}(w)$ and w not real, the left side becomes ill-defined, whereas the right side remains well-defined and finite.

4.5 Proof of Theorem 4.4.1

It is clear that there is no loss of generality to assume $u = 0$, so we do this from now on. The proof relies on a number of lemmas. The first one establishes the result under additional assumptions on the initial conditions.

Lemma 4.5.1. *In the setting of Theorem 4.4.1, assume that w has negative definite real part. Then the conditions (i) and (ii) both hold, as does the transform formula (4.4.2).*

Proof. By Lemma 4.5.2 below, w has a symmetric inverse with negative definite real part, so the explicit formula (4.4.4) is valid before the first time the expression in parenthesis becomes singular. However, this never happens. Indeed, since a is positive semidefinite, the expression in parenthesis has a negative definite real part, so by Lemma 4.5.2 it is invertible with an inverse whose real part is again negative definite. Therefore π given by (4.4.4) is a global solution, always with a negative definite real part. Condition (ii) of Theorem 4.4.1 follows. Furthermore, for any $c \in \mathbb{R}^d$ and negative definite $Q \in \mathbb{S}^d$ we have

$$\max_{y\in\mathbb{R}^d}\left(c^\top y + y^\top Q y\right) = -\frac{1}{4}c^\top Q^{-1}c \le \frac{1}{4}\|c\|^2\|Q^{-1}\|.$$

Therefore, for any $y \in \mathbb{R}^d$,

$$\begin{aligned}\left|\mathrm{e}^{\phi(\tau)+\psi(\tau)^\top y+y^\top\pi(\tau)y}\right| &= \mathrm{e}^{\mathrm{Re}(\phi(\tau))+\mathrm{Re}(\psi(\tau))^\top y+y^\top\mathrm{Re}(\pi(\tau))y}\\ &\le \mathrm{e}^{\mathrm{Re}(\phi(\tau))+\frac{1}{4}\|\mathrm{Re}(\psi(\tau))\|^2\|\mathrm{Re}(\pi(\tau))^{-1}\|}.\end{aligned}$$

By continuity of (ϕ,ψ,π) and invertibility of π on the compact interval $[0,T]$, the right side is bounded above by some constant that does not depend on τ. This implies that condition (i) of Theorem 4.4.1 holds. It also implies that the following complex-valued process M is uniformly bounded on $[0,T]$:

$$M_t = \mathrm{e}^{\phi(T-t)+\psi(T-t)^\top X_t + X_t^\top\pi(T-t)X_t}, \qquad 0\le t\le T.$$

Since (ϕ,ψ,π) satisfies the Riccati system, applying Itô's formula separately to the real and imaginary parts shows that M is a local martingale. Moreover,since M is uniformly bounded, it is in fact a true martingale. The equality $\mathbb{E}_x[M_T] = M_0$ is then precisely (4.4.2), and the lemma is proved. □

Lemma 4.5.2. *Let $w_1, w_2 \in \mathbb{S}^d$, with w_1 negative definite. Then $w_1 + \mathrm{i}w_2$ is invertible with symmetric inverse. Moreover, the real part of the inverse is negative definite.*

Proof. In general, a matrix $w_1 + \mathrm{i}w_2$, with w_1 and w_2 real and w_1 invertible, is itself invertible if and only if the Schur complement $A = w_1 + w_2 w_1^{-1} w_2$ is nonsingular. In this case $(w_1+\mathrm{i}w_2)^{-1} = A^{-1} + \mathrm{i}w_1^{-1}w_2A^{-1}$. Under the assumptions of the lemma, it is clear that A is symmetric and invertible with negative definite inverse. That $w_1^{-1}w_2A^{-1}$ is symmetric is easily verified. □

We now introduce some notation. Let $\phi(t,v,w)$, $\psi(t,v,w)$, $\pi(t,w)$ be the value at time $t < T^*(w)$ of the solution to (4.4.1) with $u = 0$. We also introduce the following notation for the maximal set of initial conditions for which a solution (ψ,π) on $[0,t]$ can be found:

$$D_{\mathbb{R}}(t) = \{(v,w) \in \mathbb{R}^d \times \mathbb{S}^d : t < T^*(w)\}$$
$$D_{\mathbb{C}}(t) = \{(v,w) \in \mathbb{C}^d \times (\mathbb{S}^d + \mathrm{i}\mathbb{S}^d) : t < T^*(w)\}.$$

(While T^* only depends on w, it will be convenient to keep the v component in the above definitions.) For any $x \in \mathbb{R}^d$, $t \geq 0$, $(v,w) \in D_{\mathbb{C}}(t)$, we also define

$$F(t,x,v,w) = \mathrm{e}^{\phi(t;0,v,w)+\psi(t;v,w)^\top x + x^\top \pi(t;w)x}.$$

Before giving some basic properties of these objects, we state the following comparison result for ODEs. It follows from a general theorem due to Volkmann [Volkmann, 1973, Satz 2]. We let $\preceq$ denote the partial order on $\mathbb{S}^d$ induced by the cone $\mathbb{S}^d_+$.

Lemma 4.5.3. *Let $w_1, w_2 \in \mathbb{S}^d$ with $w_1 \preceq w_2$, and $a_1, a_2 \in \mathbb{S}^d_+$ with $a_1 \preceq a_2$. If $(f_i(t) : 0 \leq t < T)$, $i = 1,2$, solves*

$$f_i'(t) = f_i(t) - \kappa + -\kappa^\top f_i(t) + 2f_i(t)a_i f_i(t), \qquad f_i(0) = w_i,$$

then $f_1(t) \preceq f_2(t)$ for all $t \in [0,T)$. In particular, with $a_1 = 0$, $a_2 = a$, $w_1 = w_2 = w \in \mathbb{S}^d$, we obtain

$$\mathrm{e}^{-\kappa^\top t} w \mathrm{e}^{-\kappa t} \preceq \pi(t,w), \qquad \tau \in [0, T^*(w)).$$

Proof. This is an immediate application of the specialization of Volkmann's theorem given in [Cuchiero et al., 2011, Theorem 4.8]. □

Lemma 4.5.4. *For all $t \geq 0$ and $x \in \mathbb{R}^d$, the set $D_{\mathbb{C}}(t)$ is open in $\mathbb{C}^d \times (\mathbb{S}^d + \mathrm{i}\mathbb{S}^d)$, and the map $(v,w) \mapsto F(t,x,v,w)$ is analytic in $D_{\mathbb{C}}(t)$. The set $D_{\mathbb{R}}(t)$ is open in $\mathbb{R}^d \times \mathbb{S}^d$ and star-shaped around zero.*

Proof. The openness of $D_{\mathbb{C}}(t)$ (and hence of $D_{\mathbb{R}}(t)$) and analyticity of $F(t,x,\cdot,\cdot)$ in this set follow directly from [Filipovic, 2009, Lemma 10.1] upon identifying $\mathbb{C}^d \times (\mathbb{S}^d + \mathrm{i}\mathbb{S}^d)$ with $\mathbb{C}^{d+d(d+1)/2}$. To prove that $D_{\mathbb{R}}(t)$ is star-shaped around zero, we pick any $(v,w) \in D_{\mathbb{R}}(t)$ and any $\theta \in (0,1]$, and show that $(\theta v, \theta w) \in D_{\mathbb{R}}(t)$. Using the ODE for π, one checks that $\widetilde{\pi}(\tau) = \theta\pi(\tau,w)$ satisfies

$$\widetilde{\pi}'(\tau) = \widetilde{\pi}(\tau) - \kappa + -\kappa^\top \widetilde{\pi}(\tau) + 2\widetilde{\pi}(\tau)\left(\theta^{-1}a\right)\widetilde{\pi}(\tau), \qquad \widetilde{\pi}(0) = \theta w,$$

for $\tau \in [0,t]$. Letting $\underline{w}$ be a negative definite matrix such that $\underline{w} \preceq \theta w$, and noting that $a \preceq \theta^{-1}a$, the comparison lemma (Lemma 4.5.3) yields

$$\pi(\tau, \underline{w}) \preceq \pi(\tau, \theta w) \preceq \widetilde{\pi}(\tau), \qquad \tau < T^*(\underline{w}) \wedge T^*(\theta w) \wedge t.$$

But $T^*(\underline{w}) = \infty$ by Lemma 4.5.1, so we deduce that $T^*(\theta w) > t$. Hence $(\theta v, \theta w) \in D_{\mathbb{R}}(t)$, as claimed. □

For $x \in \mathbb{R}^d$ and $t \geq 0$, define the sets

$$V(t,x) = \left\{(v,w) \in \mathbb{R}^d \times \mathbb{S}^d : \mathbb{E}_x\left[\mathrm{e}^{v^\top X_t + X_t^\top w X_t}\right] < \infty\right\},$$

$$V(t) = \bigcap_{x \in \mathbb{R}^d} V(t,x).$$

Moreover, consider the function

$$G(t,x,v,w) = \mathbb{E}_x\left[\mathrm{e}^{v^\top X_t + X_t^\top w X_t}\right],$$

which is well-defined and finite for all $x \in \mathbb{R}^d$, $t \geq 0$, $(v,w) \in \mathcal{S}(V(t,x))$, where here and in the sequel we write

$$\mathcal{S}(D) = D + \mathrm{i}(\mathbb{R}^d \times \mathbb{S}^d)$$

for the strip in $\mathbb{C}^d \times (\mathbb{S}^d + \mathrm{i}\mathbb{S}^d)$ generated by a subset $D \subset \mathbb{R}^d \times \mathbb{S}^d$. Note that $G(t,x,v,w)$ is well-defined for all real (v,w) if we allow it to take the value infinity. With this convention the monotone convergence theorem implies that $G(t,x,v,\cdot)$, with $v \in \mathbb{R}^d$, is left-continuous and nondecreasing on $\mathbb{S}^d$ in the following sense: If $w_n \to w$, $w_n \preceq w_{n+1}$, then $G(t,x,v,w_n) \uparrow G(t,x,v,w)$. Finally, note that the sets $V(t,x)$ are convex since $G(t,x,\cdot,\cdot)$ is a convex function.

Lemma 4.5.5. *For all $t \geq 0$ and all $x \in \mathbb{R}^d$ we have $D_{\mathbb{R}}(t) \subset V(t,x)$. Moreover, the equality*

$$F(t,x,v,w) = G(t,x,v,w)$$

holds for all $x \in \mathbb{R}^d$, $t \geq 0$, and all $(v,w) \in D_{\mathbb{C}}(t) \cap \mathcal{S}(D_{\mathbb{R}}(t))$.

Proof. By Lemma 4.5.4, $D_{\mathbb{R}}(t)$ is star-shaped around zero. The claim thus follows from [Filipovic, 2009, Lemma 10.9], with $U' = D_{\mathbb{C}}(t)$ and $h(v,w) = F(t,x,v,w)$, once we establish $\mathrm{i}(\mathbb{R}^d \times \mathbb{S}^d) \subset D_{\mathbb{C}}(t)$ and $F(t,x,\mathrm{i}v,\mathrm{i}w) = G(t,x,\mathrm{i}v,\mathrm{i}w)$ for all

$(v, w) \in \mathbb{R}^d \times \mathbb{S}^d$. First note that for any $\lambda \in \mathbb{R}$, $w_2 \in \mathbb{S}^d$ and $w_1 \in \mathbb{S}^d$ positive semidefinite, the following equality holds:

$$\begin{aligned}\det(\lambda \mathrm{Id} - w_1 w_2) &= \det(w_1^{1/2}[\lambda w_1^{-1/2} - w_1^{1/2} w_2]) \\ &= \det(\lambda w_1^{-1/2} - w_1^{1/2} w_2) \det(w_1^{1/2}) \\ &= \det(\lambda - w_1^{-1/2} w_2 w_1^{-1/2})\end{aligned}$$

In other words, the eigenvalues of $w_1 w_2$ and $\sqrt{w_1} w_2 \sqrt{w_1}$ coincide. Using this it is easy to show that $U(\tau)$ appearing in (4.4.3), with w replaced by $\mathrm{i}w$, is invertible for all $\tau \geq 0$. Hence $\mathrm{i}(\mathbb{R}^d \times \mathbb{S}^d) \subset D_{\mathbb{C}}(t)$. Continuity of F, Lemma 4.5.1, and dominated convergence now give

$$\begin{aligned}F(t, x, \mathrm{i}v, \mathrm{i}w) &= \lim_n F(t, x, \mathrm{i}v, \mathrm{i}w - n^{-1}\mathrm{Id}) \\ &= \lim_n \mathbb{E}_x \left[\mathrm{e}^{\mathrm{i}v^\top X_t + X_t^\top (\mathrm{i}w - n^{-1}\mathrm{Id}) X_t} \right] \\ &= G(t, x, \mathrm{i}v, \mathrm{i}w).\end{aligned}$$

□

Lemma 4.5.6. *For each $n \in \mathbb{N}$, suppose $(v, w_n) \in D_{\mathbb{R}}(T)$, $w_n \preceq w_{n+1}$, $\lim_n \|\pi(T, v, w_n)\| = \infty$, and assume the eigenvalues of $\pi(T, v, w_n)$ are bounded from below uniformly in n. Then for some $\widehat{x} \in \mathbb{R}^d$,*

$$\lim_n F(T, \widehat{x}, v, w_n) = \infty.$$

Proof. Lemma 4.5.5 implies that for all $x \in \mathbb{R}^d$,

$$F(T, x, v, w_n) = G(T, x, v, w_n) \leq G(T, x, v, w_{n+1}) = F(T, x, v, w_{n+1}),$$

so the limit in the statement of the lemma exists. Moreover, setting $x = 0$ in this inequality shows $\phi(t, v, w_n) \leq \phi(t, v, w_{n+1})$. Hence it suffices to find $\widehat{x}$ so that

$$\limsup_n \left(\psi_n^\top \widehat{x} + \widehat{x}^\top \pi_n \widehat{x} \right) = \infty, \tag{4.5.1}$$

where we defined $\psi_n = \psi(T, v, w_n)$ and $\pi_n = \pi(T, v, w_n)$. Suppose first $\|\psi_n\|$ is unbounded, and pick a subsequence (n_k) along which $\|\psi_{n_k}\| \to \infty$ and $\psi_{n_k}/\|\psi_{n_k}\|$ converges. Let $\widehat{x}$ be the limit. Since π_n has eigenvalues bounded from below, there is a constant $c > 0$ such that

$$\lim_k \frac{1}{\|\psi_{n_k}\|} \left(\psi_{n_k}^\top \widehat{x} + \widehat{x}^\top \pi_{n_k} \widehat{x} \right) \geq \lim_k \frac{1}{\|\psi_{n_k}\|} \left(\psi_{n_k}^\top \widehat{x} - c \right) = 1.$$

Now (4.5.1) follows. Suppose instead $\|\psi_n\| \le c$ for some $c > 0$. Pick a subsequence (n_k) along which $\pi_{n_k}/\|\pi_{n_k}\|$ converges to a limit π. Since $\|\pi_{n_k}\| \to \infty$ and π_{n_k} has eigenvalues bounded from below, we can find $\widehat{x}$ so that $\widehat{x}^\top \pi \widehat{x} > 0$. Then

$$\lim_k \frac{1}{\|\pi_{n_k}\|}\left(\psi_{n_k}^\top \widehat{x} + \widehat{x}^\top \pi_{n_k} \widehat{x}\right) \ge \lim_k \frac{1}{\|\pi_{n_k}\|}\left(-c\|\widehat{x}\| + \widehat{x}^\top \pi_{n_k} \widehat{x}\right) = \widehat{x}^\top \pi \widehat{x},$$

and we deduce (4.5.1). □

Proof of Theorem 4.4.1. Recall that we assume $u = 0$. The equivalence between conditions (i) and (ii) follows if we can prove $D_{\mathbb{R}}(T) = V(T)$. In view of Lemma 4.5.5, the remaining statement will then follow if we prove the inclusion $\mathcal{S}(D_{\mathbb{R}}(T)) \subset D_{\mathbb{C}}(T)$.

Let us prove $V(T) \subset D_{\mathbb{R}}(T)$, the reverse inclusion being already established in Lemma 4.5.5. Pick $(v, w) \in V(T)$ and define θ^* by

$$\theta^* := \sup\{\theta \in \mathbb{R} : (v, w - \theta \mathrm{Id}) \notin D_{\mathbb{R}}(T)\}.$$

We wish to show that θ^* lies in $[-\infty, 0)$. Note that for all large $\theta \in \mathbb{R}$, $(v, w-\theta\mathrm{Id}) \in \mathbb{R}^d \times \mathbb{S}_-^d \subset D_{\mathbb{R}}(T)$ according to Lemma 4.5.1. Therefore $\theta^* \ne \infty$. It remains now to exclude $\theta^* \in [0, \infty)$. Assume for contradiction that $\theta^* \ge 0$ and define $w^* := w - \theta^* \mathrm{Id}$. Since $D_{\mathbb{R}}(T)$ is open, $(v, w^*) \notin D_{\mathbb{R}}(T)$. Now take a sequence $T_n \uparrow T^*(w^*) \le T$, so that $\lim_n \|\pi(T_n, w^*)\| = \infty$. By continuity of π we can find $\theta_n \downarrow \theta^*$ such that $\|\pi(T_n, w_n) - \pi(T_n, w^*)\| < 1$, and hence

$$\lim_n \|\pi(T_n, w_n)\| = \infty, \tag{4.5.2}$$

where $w_n = w - \theta_n \mathrm{Id} \in D_{\mathbb{R}}(T)$. The flow property of π and the comparison lemma (Lemma 4.5.3) give

$$\pi(T, w_n) = \pi(T - T_n, \pi(T_n, w_n)) \ge \mathrm{e}^{-\kappa^\top (T-T_n)} \pi(T_n, w_n) \mathrm{e}^{-\kappa(T-T_n)}. \tag{4.5.3}$$

Another application of the comparison lemma gives $\pi(T_n, w_n) \ge \pi(T_n, w_1)$, so that the eigenvalues of $\pi(T_n, w_n)$ are bounded from below, uniformly in n. Together with (4.5.2) and (4.5.3) this implies $\lim_n \|\pi(T, w_n)\| = \infty$ and that the eigenvalues of $\pi(T, w_n)$ are uniformly bounded from below. Thus, with $\widehat{x}$ as in Lemma 4.5.6, we use the left-continuity (with respect to $\le$) of $G(t, x, v, \cdot)$ and the fact that $(v, w_n) \in D_{\mathbb{R}}(T)$ for each n together with Lemma 4.5.5 to get

$$G(T, \widehat{x}, v, w^*) = \lim_n G(T, \widehat{x}, v, w_n) = \lim_n F(T, \widehat{x}, v, w_n) = \infty. \tag{4.5.4}$$

Therefore (v, w^*) cannot lie in $V(T, \widehat{x})$ and hence not in $V(T)$. On the other hand, as $w^* \preceq w$ and G is non-decreasing (with respect to $\preceq$), we have

$$G(T, x, v, w^*) \le G(T, x, v, w) < \infty$$

for all $x \in \mathbb{R}^d$. But this contradicts (4.5.4), hence $\theta^* < 0$, so that $(v, w) \in D_{\mathbb{R}}(T)$ as desired.

Finally, we prove $\mathcal{S}(D_{\mathbb{R}}(T)) \subset D_{\mathbb{C}}(T)$. To this end, let $(v, w) \in D_{\mathbb{R}}(T)$, and pick any $(\overline{v}, \overline{w}) \in \mathbb{R}^d \times \mathbb{S}^d$. Define

$$(v_\theta, w_\theta) = (v, w) + \mathrm{i}\theta(\overline{v}, \overline{w}),$$

and let $\theta^* = \inf\{\theta > 0 : (v_\theta, w_\theta) \notin D_{\mathbb{C}}(T)\}$. We wish to show that $\theta^* = \infty$, and assume for contradiction that $\theta^* < \infty$. Since $D_{\mathbb{C}}(T)$ is open, $(v_{\theta^*}, w_{\theta^*}) \notin D_{\mathbb{C}}(T)$. Now take a sequence $t_n \uparrow T^*(w_{\theta^*}) \le T$, so that $\lim_n \|\pi(t_n, w_{\theta^*})\| = \infty$. By continuity of π we can find $\theta_n \uparrow \theta^*$ with $(v_{\theta_n}, w_{\theta_n}) \in D_{\mathbb{C}}(T)$ for all n, such that $\|\pi(t_n, w_{\theta_n}) - \pi(t_n, w_{\theta^*})\| < 1$. Then $\lim_n \|\pi(t_n, w_{\theta_n})\| = \infty$. Now, (with similar arguments as in the first part of the proof, see (4.5.3) and following), one checks that the eigenvalues of $\pi(T, w_{\theta_n})$ are bounded from below uniformly in n, and thus, $\lim_n \|\pi(T, w_{\theta_n})\| = \infty$. Therefore, with $\widehat{x}$ as in Lemma 4.5.6, we have

$$F(T, \widehat{x}, v_{\theta^*}, w_{\theta^*}) = \lim_n F(T, \widehat{x}, v_{\theta_n}, w_{\theta_n}) = \infty. \tag{4.5.5}$$

On the other hand, $D_{\mathbb{R}}(T)$ is also open, so for some $p > 1$, $(pv, pw) \in D_{\mathbb{R}}(T) \subset D_{\mathbb{R}}(t)$ for all $t \le T$. But then by Lemma 4.5.5,

$$\mathbb{E}_x\left[\left(\mathrm{e}^{v^\top X_t + X_t^\top w X_t}\right)^p\right] = F(t, x, pv, pw),$$

which is bounded in $t \in [0, T]$ by continuity of the right side. Consequently, the family of random variables $\{\mathrm{e}^{v^\top X_t + X_t^\top w X_t} : t \in [0, T]\}$ is bounded in L^p, hence is uniformly integrable. Together with path continuity of X_t, this implies that $t \mapsto G(t, x, v_{\theta^*}, w_{\theta^*})$ is continuous on $[0, T]$. But since $(v_{\theta^*}, w_{\theta^*}) \in D_{\mathbb{C}}(t) \cap \mathcal{S}(D_{\mathbb{R}}(t))$ for $t < T^*(w_{\theta^*})$, Lemma 4.5.5 gives $G(t, x, v_{\theta^*}, w_{\theta^*}) = F(t, x, v_{\theta^*}, w_{\theta^*})$. In view of (4.5.5), we obtain the desired contradiction. It follows that $\theta^* = \infty$, and the theorem is proved. □

4.6 Calibration Study

In this section we demonstrate the use of our quadratic Gaussian model by conducting a calibration test, in which the two-factor specification is fitted to

traded German electricity option data. The author would like to thank Markus Regez for his comments and for preparing and providing the data.

The data

The option data for our calibration study comes from the European Energy Exchange (short: EEX) and is publicly available[3]. It contains European-style options which are written on German electricity forwards with a calendar year delivery period[4]. The quotation date is 27th of November 2020. The options displayed on EEX are calls and puts with a very large range of strikes, and are written on calendar year forwards, i.e. Cal21, Cal22, Cal23 and Cal24. Moreover, several option expiry dates are possible on the same underyling forward, which is a unique feature of the energy markets. Note however that electricity options on calendar year forwards are not very liquidly traded. Not all strikes with their settlement prices of the options displayed on the EEX are really traded. Thus, in order to obtain traded market data, we process the raw data as follows:

- We remove all options (and their prices) which have zero open interest both in calls and their correponding puts[5]. If only calls (or their corresponding puts) are traded (in the sense of positive open interest), then through the put-call parity, the puts (or calls) will also be considered traded options.
- We compute the Black-Scholes delta of both puts and calls options with various strikes and remove those with delta $< 1\%$. This allows us to remove very far out-of-the-money options. At the same time it also removes the very deep in-the-money options via the put-call parity.

For the calibration study, we only use the call options. This is not restrictive, as the put options with non-zero open interest are converted to call options via the put-call parity. A visualization of the data is presented in Figure 1 and a detailed description in Table 1. We see that the data exhibits the so called Samuelson's

[3]The option data is available under https://www.eex.com/en/market-data/power/options.

[4]We only consider options on calendar year forwards as they are the the most liquidly traded options in mid to long-term electricity markets. Moreover, they are directly comparable without further adjustments such as the seasonality adjustments.

[5]A corresponding put has the same strike and option expiry date, and the same undelying forward contract.

effect[6].

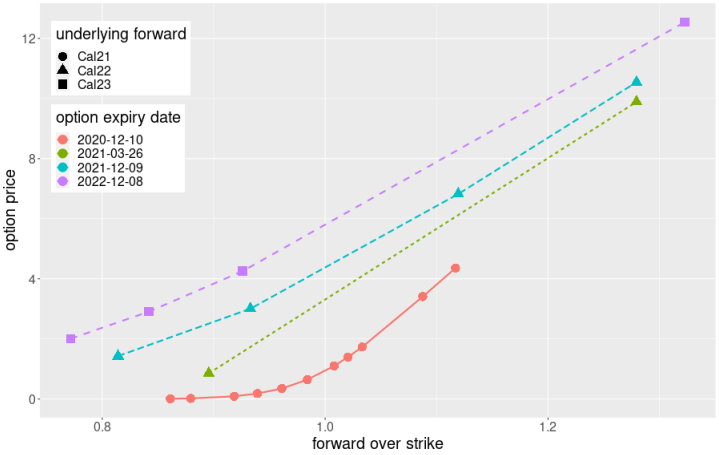


Figure 1: Traded EEX German call options that are written on calendar year forwards, priced as of 27th November 2020. The option prices at EEX are quoted in premium. They are plotted here with respect to the *moneyness* (=forward over strike).

product	delivery period	option expiry	forward	number of options
calls on Cal21	2021-01-01 - 2021-12-31	2020-12-10	41.33	11
calls on Cal22	2022-01-01 - 2022-12-31	2021-03-26	44.78	2
calls on Cal22	2022-01-01 - 2022-12-31	2021-12-09	44.78	4
calls on Cal23	2023-01-01 - 2023-12-31	2022-12-08	46.30	4

Table 1: Detailed descriptive information on the German electricity call options, which are traded on EEX, quoted as of 27th November 2020. The number of options refers to the number of different strikes traded for a given calendar year call option.

Calibration procedure

Let's $t = 0$ denote quotation date. The available financial instruments for the calibration are forwards $F(0, T_1, T_2, X_0)$ and call options $E[(F(T, T_1, T_2, X_T) -$

[6]The volatility is increasing with decreasing time-to-maturity of options.

$K)^+|\mathcal{F}_0]$ which are written on $F(T,T_1,T_2,X_T)$ for different T, T_1 and T_2. As $F(T,T_1,T_2,X_T)$ is unknown on date $t = 0$, the best approximation is given by $E[F(T,T_1,T_2,X_T)|\mathcal{F}_0] = F(0,T_1,T_2,X_0)$. Hence, the calibration procedure consists of two parts: calibration of model-based forwards and model-based calls. For the model implementation, we use the two-factor specification, Specification 4.2.2, for which we have seen that it fits the electricity forward curves very well. The model-based forward $F(0,T_1,T_2,X_0)$ can be easily implemented using the moment formula, Theorem 2.3.2. The implementation of the model-based option prices requires several steps, which we present in the following.

Implementation of model-based option prices

The model-based option price $E[(F(T,T_1,T_2,X_T) - K)^+|\mathcal{F}_0]$ can be implemented using the Fourier approach (Theorem 4.3.4), so that for each $z = \zeta + i\lambda$ with $\lambda \in \mathbb{R}$, $\widehat{q}_{\text{call}}(z)$ is computed via the exponential quadratic transform formula (Theorem 4.4.1). Theoretically either side of the transform formula (4.4.1) can be used for the computation: its left hand side suggests the use of a Monte Carlo estimator and its right hand side needs a Riccati solver. In practice however, a Monte Carlo estimator yields huge approximation error while the solution to a Riccati system can be computed very accurately; see Figure 2 for a sanity test, which visualizes this fact. Thus, we evaluate option prices by solving the Riccati system (4.4.1). For solving the Riccati system (4.4.1), we need appropriate choices of parameters and initial values in order to guarantee that the maximal lifetime of the system will not be reached.

For the evaluation of the Fourier integral, the Fourier integrand needs to be solved for many z. We need to repeatedly solve the Riccati equations for different initial values. These initial values are themselves the products of z and the solutions of linear ODEs of the forward pricing formula. However, the linear ODEs only need to be solved once. In detail, for each $z = \zeta + i\lambda$, $\lambda \in \mathbb{R}$, we solve the Riccati system (4.4.1) for initial values

$$\phi(0) = u_0 z, \qquad \psi(0) = v_0 z, \qquad \pi(0) = w_0 z,$$

where u_0, v_0 and w_0 are the solutions of (4.3.4) implied by the linear ODEs (2.3.6), and are the same for all Fourier integrands.

Regarding the numerical integration, there exists various build-in functions in any programming language which can handle it. However, the convergence of the numerical sum can be slow due to the strong oscillating property of the

Fourier integrand. A possible way to speed it up is to add and subtract another term, which has closed form moment generating function, so that the numerical integration only needs to be computed on the difference between this term and $\widehat{q}_{\text{call}}(z)$. A normal distributed random variable Y which has the same first and second moment as $F(T, T_1, T_2, X_T)$ can be a good choice. In the following, we abbreviate $F(T, T_1, T_2, X_T)$ to F_T.

To formalize the idea, let $\widetilde{Y} \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ be a normal distributed random variable. Then the call price at strike K can be expressed as

$$\begin{aligned}\mathbb{E}[(F_T - K)^+] &= \mathbb{E}[(F_T - K)^+ - (\widetilde{Y} - K)^+] + \mathbb{E}[(\widetilde{Y} - K)^+] \\ &= R(F_T, \widetilde{Y}, K) + \mathbb{E}[(\widetilde{Y} - K)^+],\end{aligned}$$

where

$$\begin{aligned}R(F_T, \widetilde{Y}, K) :&= \mathbb{E}[(F_T - K)^+ - (\widetilde{Y} - K)^+] \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \operatorname{Re}\left[\frac{1}{(\zeta + i\lambda)^2}\left(\mathbb{E}\left[e^{(\zeta+i\lambda)(F_T-K)}\right] - \mathbb{E}\left[e^{(\zeta+i\lambda)(\widetilde{Y}-K)}\right]\right)\right] d\lambda \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \operatorname{Re}\left[\frac{1}{(\zeta + i\lambda)^2}\left(\mathbb{E}\left[e^{(\zeta+i\lambda)(F_T-K)}\right] - e^{(\tilde{\mu}-K)(\zeta+i\lambda)+\frac{\tilde{\sigma}^2(\zeta+i\lambda)^2}{2}}\right)\right] d\lambda.\end{aligned}$$

Note that the second term in the last equation is the explicit expression of the characteristic function[7] of the normal distributed $(\widetilde{Y} - K) \sim \mathcal{N}(\tilde{\mu} - K, \tilde{\sigma}^2)$.

This approach splits the computation of the call price into the computation of a call price where a normal distributed random variable models the underlying dynamics, and the residual term, which is the difference between the call price of our model and that of the normal distributed $\widetilde{Y}$. Adding and subtracting $\mathbb{E}[(\widetilde{Y} - K)^+]$ has two advantages. On the one hand, the characteristic function of $\widetilde{Y}$ and thus also of $\widetilde{Y} - K$ are fully explicit. Thus, computing this additional term does not require much additional effort, but it helps to dampen the oscillating effect of a Fourier integral. On the other hand, $\mathbb{E}[(\widetilde{Y} - K)^+]$ has a closed form solution as Y is a normal distributed random variable. We will see this in the following proposition.

Proposition 4.6.1. *For any normal distributed underlying dynamics $S \sim \mathcal{N}(\mu, \sigma^2)$, the call price at strike K is given by*

$$\mathbb{E}\left[(S - K)^+\right] = (\mu - K)\Phi(-d) + \sigma f(d),$$

[7] Here we define the characteristic function on the complex numbers $\zeta + i\lambda$, $\lambda \in \mathbb{R}$.

with $d := \frac{K-\mu}{\sigma}$, $\Phi(\cdot)$ *and* $f(\cdot)$ *denote the cumulative distribution function and the density function of the standard normal distribution respectively.*

Proof. Let $Z \sim \mathcal{N}(0,1)$, then $S \overset{\text{law}}{=} \mu + \sigma Z$. Thus,

$$\begin{aligned}\mathbb{E}\Big[(S-K)^+\Big] &= \mathbb{E}\Big[(S-K)\mathbb{1}_{\{S>K\}}\Big] \\ &= \mathbb{E}\Big[(\mu+\sigma Z-K)\mathbb{1}_{\{Z>d\}}\Big] \\ &= (\mu-K)\Big(1-\Phi(d)\Big) + \sigma\mathbb{E}\Big[Z\mathbb{1}_{\{Z>d\}}\Big].\end{aligned}$$

with $d := \frac{K-\mu}{\sigma}$. The first summand can be further simplified using the symmetry $\Phi(-d) = 1-\Phi(d)$. And for the second summand, we use $f'(z) = (-z)\cdot\frac{1}{2\pi}e^{-z^2/2} = -zf(z)$, and thus

$$\sigma\mathbb{E}\Big[Z\mathbb{1}_{\{Z>d\}}\Big] = \sigma\mathbb{E}\Big[Z\Big] - \sigma\mathbb{E}\Big[Z\mathbb{1}_{\{Z\le d\}}\Big] = 0 - \sigma\int_{-\infty}^{d} zf(z)dz = \sigma\int_{-\infty}^{d} f'(z)dz = \sigma f(d).$$

This gives the claim. □

It remains to discuss the choice of $\widetilde{Y}$. In general $\widetilde{Y}$ can be chosen arbitrarily, as $\widetilde{Y}$ is in the zero-sum terms which we add and subtract at the same time. One possible choice of $\widetilde{Y}$ is to let $\tilde{\mu}$ and $\tilde{\sigma}$ satisfy:

$$\tilde{\mu} = \mathbb{E}\Big[F_T\Big] \qquad \text{and} \qquad \tilde{\sigma}^2 = \mathbb{V}\Big[F_T\Big]. \tag{4.6.1}$$

In the following we compute (4.6.1) for the two-factor model, Specification 4.2.2. Recall that in the two-factor model, X_t evolves according to a two-dimensional Ornstein-Uhlenbeck process. Thus, at any t, it is a normal-distributed random variable, i.e. $X_t \sim \mathcal{N}\big(m(t), \Sigma(t)\big)$, with

$$m(t) = e^{-\kappa t}X_0, \qquad \Sigma(t) = \int_0^t e^{-\kappa s} a\, e^{-\kappa^\top s} ds,$$

where $\Sigma(t)$ solves [8]

$$-\kappa\Sigma(t) - \Sigma(t)\kappa^\top = e^{-\kappa t} a\, e^{-\kappa^\top t} - a.$$

[8] Alternatively, one can also solve the ODE $\Sigma'(t) = -\kappa\Sigma(t) - \Sigma(t)\kappa^\top$.

Moreover, as $F_T = u_0 + X_T^\top w_0 X_T$ (see (4.3.5)), we have[9]

$$\tilde{\mu} = u_0 + \mathbb{E}\big[X_T^\top w_0 X_T\big] = u_0 + \mathrm{Tr}\big[w_0 \Sigma\big] + m^\top w_0 m,$$
$$\tilde{\sigma}^2 = \mathbb{V}\big[X_T^\top w_0 X_T\big] = \mathrm{Tr}\big[w_0 \Sigma\,(w_0 + w_0^\top)\Sigma\big] + m^\top (w_0 + w_0^\top)\Sigma(w_0 + w_0^\top)\, m,$$

where $m = m(T)$ and $\Sigma = \Sigma(T)$.

After the implementation of the two-factor model, Specification 4.2.2, we calibrate it to the real option data using the least square method. In order to speed up the fitting, we first compute the model-based forwards and call approximations (Proposition 4.6.1), and optimize over the least squares of all forwards and options. Once the discrepancy is within a reasonable range, we then evaluate Fourier integral $R(F_T, \widetilde{Y}, K)$ and compute the accurate model-based call prices and the least square over all forwards and accurate calls. The model fitting is visualized in Figure 3, and goodness of fit is given Table 2. The parameters of the calibration are reported in Table 3.

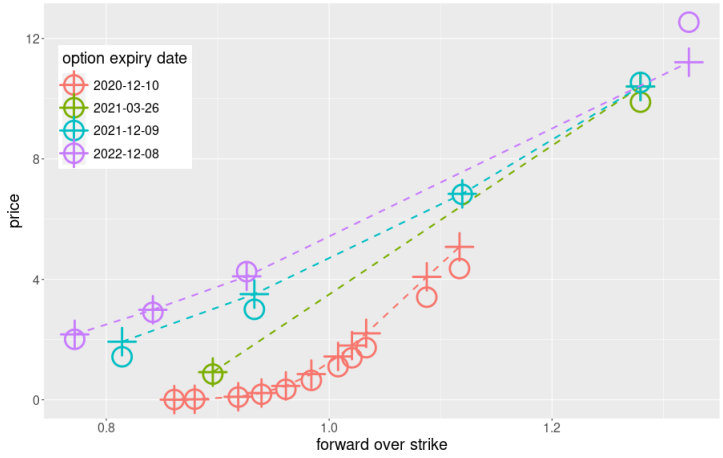


Figure 3: A visualization of the calibration study. The circles represents the traded EEX option data, and the plus represents the model-based option prices. The calibration uses $\zeta = 0.0001$. The calibration can roughly fit the option data while it fits the forward data accurately.

[9]This uses well-known results of a normal distributed random variable; see e.g. Section 8.2.2 of Petersen and Pedersen [2008].

	Cal21		Cal22			Cal23	
	forward	option	forward	option I	option II	forward	option
absolute error	0.7492	0.2725	0.6238	0.2917	0.2894546	0.0913	0.4332
relative error	1.812%	26.052%	1.393%	6.469%	13.348%	0.197%	6.306%

Table 2: Table of Goodness of fitting. The overall relative error is 1.38506% for the forwards and 18.01% for the options.

c	32.4598177
α	9.7723320
β	4.9730813
κ_Z	0.3377648
κ_Y	0.8018996
σ_Z	0.7739308
σ_Y	0.3598896
ρ	0.8297327
z_0	1.0911264
y_0	0.2491788

Table 3: Estimated parameters of Specification 4.2.2.

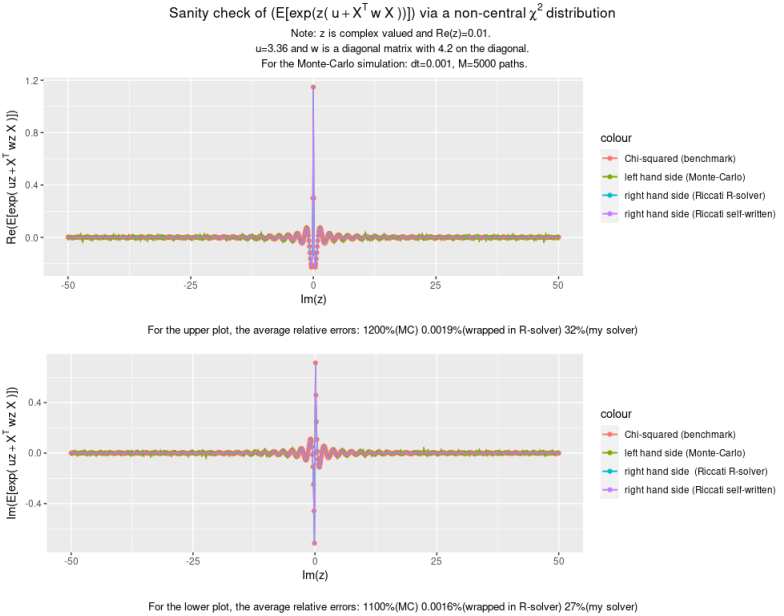


Figure 2: A visual comparison of accuracy between the left hand side and the right hand side of the exponential quadratic formula (4.4.1). This comparison is based on a special case, in which $\widehat{q}_{\text{call}}(z)$ is etual to the characteristic function of a non-central χ^2 distributed random variable $Y := u + X_T^\top w X_T$. The Monte Carlo estimator uses 5000 simulation paths and 1000 time steps. For solving the Riccati system, we use two different solvers for compasion, the deSolve-package of R and a self-written solver based on an Euler-scheme. We see that the deSolve solver gives almost accurate result, and Monte Carlo has too significant errors (1000% relative error).

Chapter 5

A machine learning approach to gas storage optimization

5.1 Introduction

The prices of natural gas exhibit distinct yearly seasonal pattern. Due to the partial storability and large fluctuation in demand of gas, its prices are generally lower in summer, and higher and spikier in winter. In order to take advantage of the seasonality of gas, physical storage facilities are required. And thus, producers and other market players are highly motivated to own or to contract storage facilities, creating worldwide a high demand in the underground gas storage facilities. As an example, the US working gas[1] in underground storage in 2020 is record high compared with that of the past five years; see Figure 1. Given the rising price of gas storage and the smoothing effect of gas price spreads through the use of storage facility, it is essential to optimize the use of the storage in order to generate profit for a market player. In this chapter, we focus on the optimization of gas storage and use machine learning approaches, in particular

[1] Working gas refers to the total volume of gas in the storage at a particular point in time. It is computed as the total gas volume minus the base gas.

deep hedging in the sense of Buehler, Gonon, Teichmann, and Wood [2019], to solve it.

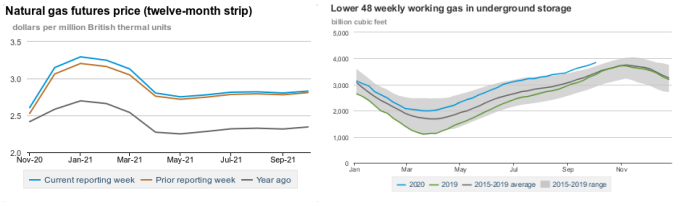


Figure 1: US natural gas (not LNG) futures curve and storage information provided by EIA (US Energy Information Administration), as of October 2, 2020; plots are from https://www.eia.gov/naturalgas/storage/dashboard/. Left plot shows one year natural gas futures curves consisting of twelve monthly futures contracts with delivery period of months ranging from November 2020 to October 2021; we see that there is a pronounced seasonal pattern, namely high prices in winter months and low prices in summer.The right plot shows the lower bound of underground gas storage in US in lower 48 weekly working gas; it is obvious that the gas storage this year (2020) is record high compared with that of the past five years.

There are three types of underground natural gas storage: depleted natural gasfield/oilfields, aquifers, and salt caverns. They are often located close to a pipeline, which makes the delivery of physical transactions more convenient. Compared to the storage through conversion to LNG, the underground natural gas storage are bigger and cheaper, but are restricted to regional use. In Table 1, we provide an overview of the main stylized characteristics of a gas storage, that are relevant for modeling and optimization.

Over the last decades, various literature contributes to the modeling and optimization of energy storage. For standard references see section 12.6 of Geman [2009], section 5.3.4 of Fiorenzani et al. [2012], and Holland [2007, 2008]. Other references include e.g. De Jong [2015], Boogert and De Jong [2008], Safarov and Atkinson [2017], Cummins et al. [2017], Carmona and Ludkovski [2010], Bjerksund et al. [2011], Thompson et al. [2009], Hénaff et al. [2018], Malyscheff and Trafalis [2017]. Much of the literature puts more emphasis on the modeling (and prediction) of gas prices rather than on developing algorithm for the optimization problems. On the optimization side, storage was optimized via the Least Squares Monte Carlo approach (LSMC) or support vector machine regression (SVR), considered as a stochastic control problem

storage optimisation constraints with unit: therm or MWh	
initial storage	0 units (plus cushion gas)
terminal storage	0 units (plus cushion gas)
capacity	c
injection rate on day k	u_k units ($u_k > 0$)
withdrawal rate on day k	ℓ_k units ($\ell_k < 0$)
injection cost	$\kappa \in [0, 1]$
withdrawal cost	$\kappa \in [0, 1]$
overhead (one time expense)	K \$

Table 1: This table gives the most important characteristics for modeling the gas storage. Note that in the underground storage, there is usually a cushion gas (base gas) which is the volume of natural gas that is intended as permanent (and not withdrawable) inventory to maintain minimal pressure, which for modeling purpose can be neglected. For simplicity the injection and withdrawal costs are assumed to be a constant proportional cost of the injection and withdrawal respectively. In reality, the cost depends on the pressure in the underground storage, which itsself depends on the level of the working gas.

with HJB equations, or an application of real option theory. To the best of the author's knowledge, almost none of the machine learning techniques has been applied to gas storage and related problems. And thus, our applications of *deep hedging* from Buehler, Gonon, Teichmann, and Wood [2019] to the gas storage is very unique. It demonstrates the potential of machine learning techniques in storage-related modeling and optimization problems. We provide two models based on this machine learning concept, which are of the intrinsic valuation type. The first model (Model I) uses spot-proxy (the day-ahead forwards) as hedging instruments. The second model (Model II) uses additionally the monthly forwards with delivery periods as hedging instruments. The second model is very unique in the sense that it features the real gas trading very well, due to the use of the tradable monthly forwards. The traditional approach is to use an artificial daily forward curve for simplicity, which is implied from the tradable monthly forwards with delivery period. The author strongly believes that the use of tradables, and thus the use of Model II, are more appropriate for the gas storage optimization, as the main purpose of gas storage optimization is to maximize the P&L or to maximize the utility for a storage manager rather than a valuation for risk-management purpose.

This chapter is structured in the following way: In Section 2 we review the deep hedging approach. In Section 3, we present our first model (Model I), an intrinsic spot model which uses deep hedging. We compare our model in

numerical tests against a set of benchmark strategies optimized using LSMC. In Section 4, we present our second model (Model II), which uses additionally monthly forwards with delivery periods, and investigate its performance in numerical tests against Model I and benchmark strategies using LSMC.

Throughout this chapter, we use a discrete setting: Let $\mathbb{T} = \{0, 1, 2, \cdots, N-1\}$ for some $N \in \mathbb{N}$ be the trading horizon in days, and let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $\mathbb{F} = (\mathcal{F}_k)_{k\in\mathbb{T}}$ be a filtered probability space with real-world measure $\mathbb{P}$. And we assume the existence of an equivalent risk neutral measure $\mathbb{Q}$.

This chapter is based on the joint work with Thomas Krabichler, Josef Teichmann and Hanna Wutte. The author would like to thank Vlatka Komaric and Michael Kettler of Axpo Solutions for posing this interesting problem, for fruitful discussions and for providing data.

5.2 Deep hedging

The concept of *deep hedging* goes back to the beautiful paper of Buehler, Gonon, Teichmann, and Wood [2019]. In the following we review this concept. The notations and formulations we use are very similar to that of the paper and of the lecture notes of "Machine Learning in Finance" by Christa Chuchiero and Josef Teichmann; see https://people.math.ethz.ch/~jteichma/index.php?content=teach_mlf2019 for more details.

Let us first recall the definition of a neural network.

Definition 5.2.1 (neural network, Buehler et al. [2019])**.** *Let L denote the number of layers, let $N_0, N_1, \ldots, N_L \in \mathbb{N}$ denote the dimension of layers respectively. In particular, $N_1, \ldots, N_{L-1}$ denote the dimension of the hidden layers, and N_0 and N_L the dimension of the input and output layers. Further, let $\hat{\sigma} : \mathbb{R} \to \mathbb{R}$ denote the (sigmoidal) activation function. For any $l = 1, ..., L$, let $W_l : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$ be an affine function. A function $g : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ defined as*

$$g(x) = W_L \circ F_{L-1} \circ \cdots F_1 \text{ with } F_l = \hat{\sigma} \circ W_l \text{ for } l = 1, ..., L-1$$

is called a (feed forward) neutral network with L layers.

The main idea of deep hedging is that at any time point t, the hedging or trading strategy is constructed and parametrized via a neural network. The input variables are chosen by the user, and can be e.g. the current price of an asset or the strategy used in the past. The training is based on a hedging criterion

which can be translated into the minimization of a loss function. Therefore, the outputs of the neural network which are used to train the neural network, can be set to zero. In other words, the deep hedging can be seen as an approach of the *supervised learning*. In the following, we make the idea mathematically more precise.

Let X_k be the price of an asset on day k, and h_k the hedging strategy on day k. In terms of gas storage optimization, h_k is the action of storage, and describes the rate of injection or withdrawal on day k. A hedging strategy H over the whole trading horizon is then $H = \{h_0, ..., h_{N-1}\}$, and its value is given by $(H \bullet X)_{N-1} := \sum_{k=0}^{N-1} h_k \Delta X_k = \sum_{k=0}^{N-1} h_k (X_{k+1} - X_k)$.

Suppose we want to hedge a claim $f(X_{N-1})$ and consider a quadratic hedging criterion such as the mean-variance criterion, then the hedging problem can be formulated as the following optimization problem:

$$\inf_{H \in \mathcal{H}} \mathbb{E}\Big[\big(f(X_{N-1}) - \pi - (H \bullet X)_{N-1}\big)^2\Big], \tag{5.2.1}$$

where $\pi := \mathbb{E}_{\mathbb{Q}}[f(X_{N-1})]$ denote the fair price of $f(X_{N-1})$, and $\mathcal{H}$ denotes the set of all predictable strategies H.

Next, on each day $k \in \mathbb{T}$, we model the hedge h_k as a function of the current price X_k and approximate it by a neural network, i.e. $h_k = g_k(X_k)$. For the whole trading horizon, we have N neural networks. The training of g_k is conducted via a loss function $\mathcal{L}$ implied by (5.2.1). Therefore, the hedging problem can be formulated as the following supervised learning problem:

- Input: M trajectories of underlying : $(X_k^i)_{k \in \mathbb{T}; i=1,...,M}$;
- Output: M 0's;
- Training object: hedging strategy of the whole trading horizon, that is N neural networks $g_0, ..., g_{N-1}$, each of which has L layers;
- Training criterion: minimize loss function $\mathcal{L}$, i.e. for $i = 1, ..., M.$,

$$\mathcal{L}(i) = \Big(f(X_{N-1}^i) - \pi - \sum_{k=0}^{N-1} g_k(X_k^i) \cdot [X_{k+1}^i - X_k^i] - 0\Big)^2.$$

The strength of deep hedging is that it allows to solve a high dimensional hedging and optimization problem with complex underlying dynamics very efficiently.

5.3 Model I: intrinsic spot trading

In the following, we introduce a deep hedging model for gas storage optimization, which is based on trading the day-ahead prices of gas. Note that in the Commodities markets, the day-ahead futures or forward is seen as a close proxy of the spot price. And therefore, we will refer to the trading activities of the day-ahead price of gas as the spot trading throughout the chapter. For simplicity, we assume no discounting (r=0) and zero transaction cost, i.e. $\kappa = K = 0$; for a more general formulation including costs, see Remark 5.3.1.

Let $S_k = \big(F(k, k+1, k+1)\big)_{k\in\mathbb{T}}$ denote the $\mathbb{F}$-adapted gas spot price, and h_k^S the $\mathcal{F}_k$-measurable *action* on day k. $h_k^S > 0$ refers to an *injection* of $|h_k^S|$ MWh and $h_k^S < 0$ refers to a *withdrawal* of $|h_k^S|$ MWh. A trading strategy $\widetilde{H}$ over the whole trading horizon is then $\widetilde{H}^S = \{h_1^S, ..., h_{N-1}^S\}$ and its value is given by $(\tilde{H}^S \bullet S)_{N-1} := \sum_{k=0}^{N-1} h_k^S S_k$. Moreover, the *storage level* (or working gas) H_n^S on day n is given by

$$H_n^S := \sum_{k=0}^{n-1} h_k^S,$$

with initially empty storage, i.e. $H_0^S := 0$.

Suppose the preference of the storage manager can be expressed through a utility function $U : \mathbb{R} \to \mathbb{R}$, which we want to maximize. That is we maximize:

$$\mathbb{E}_{\mathbb{P}}\big[U(W_{N-1})\big], \tag{5.3.1}$$

where

$$W_{N-1} := \sum_{k=0}^{N-1} -h_k^S S_k \tag{5.3.2}$$

denotes the terminal profit and loss[2] (P&L). The optimisation is subject to the following constraints:

$$H_N^S = 0, \tag{5.3.3}$$

$$0 \le H_k^S \le c, \qquad \text{and} \qquad \ell_k \le h_k^S \le u_k, \tag{5.3.4}$$

for all $k \in \mathbb{T}$. Alternatively, the daily constraints (5.3.4) can be expressed iteratively using

$$\widetilde{\ell}_k := \max\big\{\ell_k, -H_k^S\big\}, \qquad \text{and} \qquad \widetilde{u}_k := \min\big\{u_k, c - H_k^S\big\},$$

[2]The negative sign is added in analogy to the direction of cashflow.

so that

$$\widetilde{\ell}_k \le h_k^S \le \widetilde{u}_k. \tag{5.3.5}$$

Next, we train the action h_k^S via deep neural networks. On day $k \in \mathbb{T}$, the neural net g_k gets the current spot price S_k and time k as input, i.e. $h_k^S = g_k(k, S_k)$. This means that in total we have N neural networks for the storage schedule $\widetilde{G}(S) = \{g_0(0, S_0), ..., g_{N-1}(N-1, S_{N-1})\}$. Note that the number of neural networks can be significantly smaller than the number of trading days. However, for the ease of notation, we set number of neural networks equal to N, the number of trading days. The network-based storage level G_n on day n is given by

$$G_n(S) = \sum_{k=0}^{n-1} g_k(k, S_k), \qquad G_0 = 0.$$

The training of g_k is conducted via a loss function $\mathcal{L}$ implied by (5.3.1)–(5.3.5). In the following, we reformulate the gas storage problem in the sense of deep hedging using an exponential utility as an example. Recall that the exponential utility is formally defined as $U(x) := 1 - e^{-rx}$ with a risk aversion rate $r \in \mathbb{R}^+$ for $x \in \mathbb{R}$. Moreover, maximizing $U(x)$ is equivalent to minimizing $\widetilde{U}(x) := e^{-rx}$. Then, the model can be formulated as:

- Input: time horizon of storage $\mathbb{T}$, M trajectories of the spot $(S_k^i)_{k\in\mathbb{T};i=1,...,M}$;
- Training object: storage action (withdrawal or injection rate) of the whole storage horizon, that is N neural networks $g_0, ..., g_{N-1}$, each of which has L layers:
$$h_k^{i,S} = g_k^S(k, S_k^i).$$
- Training criterion: minimize the loss function $\mathcal{L}$, i.e. for $i = 1, ..., M.$,
$$\min_{\widetilde{G}(S^i)\in\mathcal{G}^i} \mathcal{L}(i) := \min_{\widetilde{G}(S^i)\in\mathcal{G}} \widetilde{U}\big(W_{N-1}^i\big),$$
where
$$W_{N-1}^i = \sum_{k=0}^{N-1} -g_k(k, S_k^i) \cdot S_k^i,$$
$$\mathcal{G}^i = \Big\{\widetilde{G}(S^i) \mid G_N(S^i) = 0;\ 0 \le G_k(S^i) \le c;\ \ell_k \le g_k(k, S_k^i) \le u_k \ \text{ for } k \in \mathbb{T}\Big\}.$$

Remark 5.3.1. *For the full case as described in Table 1, where κ and K are non-zero, the terminal profit and loss is given by*

$$W_{N-1} := \left(\sum_{k=0}^{N-1} -h_k^S S_k - \left| h_k^S S_k \right| \cdot \kappa \right) - K$$

For numerical testing, spot curves of gas as well as benchmark strategies were provided by Axpo Solutions. The spot prices is provided as a 1000×351 matrix, with $M = 1000$ scenarios each of which has prices of $N = 351$ trading days. The corresponding benchmark strategies are provided in the same format. They are computed using the LSMC.

We implemented the model above by mainly using the Keras module of Python. In total, there are $N = 351$ neural networks, each of which models the storage action on a single trading day. For each neural network g_k, we used the sigmoidal function as the activation function; the daily constraints ℓ_k and u_k, and the action h_k were transformed using a linear interpolation to 0, 1 and $\frac{h_k - \ell_k}{(u_k - \ell_k)} \in [0, 1]$ respectively. Moreover, we create iterative tensor operations which uses functions on abstract tensors as well as the sublayering technique in the high level implementation of Keras. For incorporating the daily constraints, we used the iterative expression (5.3.4). For the final zero storage constraint, we check for every k:

$$H_k^S \leq \sum_{k+1}^{N-1} \ell_k, \tag{5.3.6}$$

to make sure that $H_N^S = 0$ will still be realistic given the storage level on day k; if the condition is violated for any k, the upper bound for action will be set equal to the lower bound, forcing a complete withdrawal of storage for all days starting from day k. In reality of course, it is possible to leave a non-empty storage by paying some penalization fee. Yet for our modeling, the zero final storage rule holds strictly. Figure 2 visualizes the constraints: left plot shows the normalized zero storage constraint, the right hand side of (5.3.6), and the right plot shows the daily injection and withdrawal bounds.

After having implemented the model, we train our model of h_k based on the spot prices. We split our data set into a training set of 900 scenarios and a validation set of 100 scenarios in order to perform in-sample and out-of-sample tests. All of the 1000 benchmark strategies were optimized using LSMC. Thus they serve as the optimal solution. We want to examine how fast and how close we get to a reasonably well solution. For that, we varied the length of training,

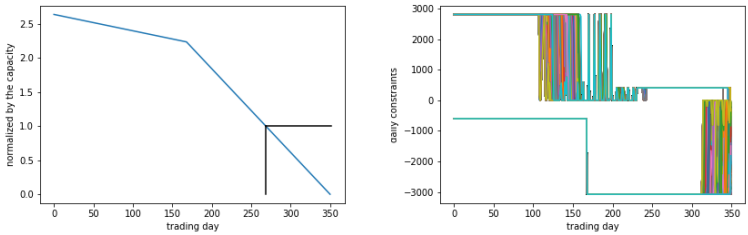


Figure 2: The daily constraints of a gas storage. The left figure visualizes the empty final storage constraint, i.e. the left hand side of (5.3.6), and is normalized by the total capacity c. The critical boundary is reached on the trading day 269. In other words, from day 269 until day $N-1$ ($N=351$) the daily action is only to withdrawal natural gas at the maximal withdrawal rate. The right figure visualizes the daily injection and withdrawal constraints.

the depth and the number of the neural networks ($\leq N$) as well as the learning rate and the batch size of training to see what setting yields the best performance of Model I. The training is very fast, and is good managable on a notebook with 8 cores. As an example, for an implementation of the model using as much as N neural networks, it needs less than 10 minutes for training on 900 scenarios for 1000 iterations. Thus regarding the speed, it is comparable with the LSMC. Moreover, note that it is not necessary to build the above model with N neural networks. In fact, we find that the implementation with 12 neural networks, $L=2$, $N_0=16$ and $N_1=1$ already gives a decent model. After 1000 training iterations on 900 scenarios with learning rate=0.05, batch size=64 and risk aversion rate $r=3$, it gets reasonably close to the benchmark solution. Figure 3 provide a visualization of the P&L comparison between Model I and benchmark in the in-sample and out-of-sample tests, as well as a visualization of storage strategies of Model I and that of benchmark respectively. A descriptive statistics comparison of the terminal P&L between Model I and the benchmark is reported in the table at the bottom of Figure 6.

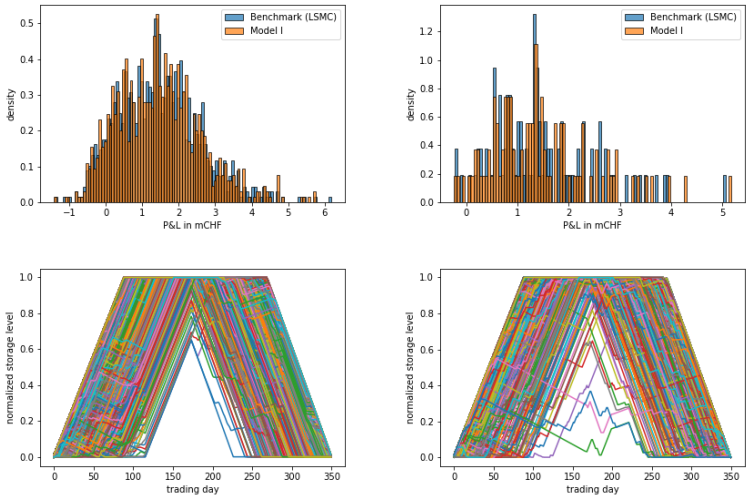


Figure 3: A comparison between performance of Model I and that of benchmark (LSMC). The upper plots compare the terminal P&L between Model I and benchmark in million CHF: upper left gives the P&L of the training set and upper right gives that of the test set. We see that in both results are quite close to the benchmark. The bottom plots shows the storage action computed by the neural networks of Model I, and the bottom right plot shows the benchmark actions. Both are normalized by the capacity c. We see that very often the strategy is to inject gas until the storage capacity is reached, and then to withdrawal until the storage is empty. This is in accordance with the underlying seasonality pattern: one store throughout summer months when prices are comparably low, and withdraw and sell at higher prices in winter months until the storage is empty.

5.4 Model II: intrinsic spot and forward trading

In the following, we extend the previous deep hedging model for gas storage by trading additionally on the front month rolling forwards with delivery period of a whole month. A front month rolling forward curve contains at any point in time the first nearby monthly forward. We assume here that a monthly forward contract is only traded before its delivery period starts, and the delivery obligations is valued using the spot prices that has delivery day within the delivery period. Note that we only consider those forwards that have delivery

months within the time horizon of the storage problem. A visualization of the forward rolling mechanism is presented in Figure 4.

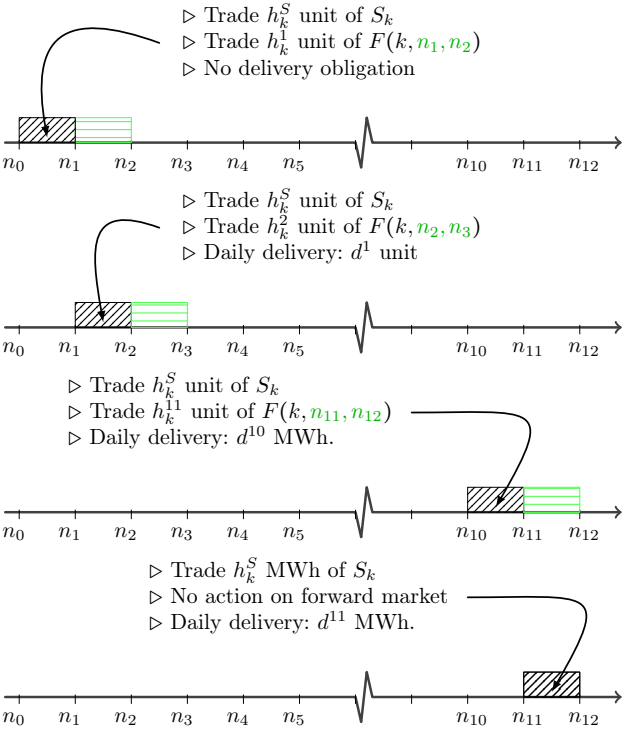


Figure 4: The mechanism of the rolling strategies in Model II. Note that contrary to the rolling mechanism in Figure 3.5.1, we allow physical settlement of forward thanks to the storage facility.

Let $0 = n_0 < n_1 < ... < n_m < N$ be the first days of the months $\mathcal{J} = \{0, 1, ..., m\}$ respectively. Let h_k^J with $J \in \mathcal{J}$ denote the action on day k on the forward $F(k, n_J, n_{J+1} - 1)$, which has delivery period $[n_J, n_{J+1} - 1]$. $h_k^J > 0$ refers to buying and $h_k^J < 0$ refers to selling $F(k, n_J, n_{J+1} - 1)$. The above assumption implies that $h_k^J = 0$ for $k < n_{J-1}$ and for $k \geq n_J$; in particular $h_k^m = 0$ for all $k \in \mathbb{T}$. We aim to maximize

$$\mathbb{E}_{\mathbb{P}}\big[U(W_{N-1})\big], \tag{5.4.1}$$

with the terminal P&L

$$W_{N-1} := W^S_{N-1} + W^F_{N-1}.$$

Here W^S_{N-1} denotes the terminal P&L from the spot trading and is given by (5.3.2), where h^S_k denotes only activity of spot trading, which results in schedule of gas storage next day. And W^F_{N-1} denotes the terminal P&L from trading the monthly forward, and is defined as:

$$W^F_{N-1} = \sum_{J=1}^{m-1} \sum_{k=n_{J-1}}^{n_J-1} \big(- h^J_k F(k, n_J, n_{J+1}-1)(n_{J+1}-n_J)\big). \tag{5.4.2}$$

For a forward with the delivery period $[n_J, n_{J+1}-1]$, the daily delivery quantity d^J is fixed on day $n_J - 1$ for $J \geq 0$, and is given by:

$$d^J := \sum_{k=n_{J-1}}^{n_J-1} h^J_k, \qquad \text{for } J > 0,$$

and $d^0 := 0$. The *storage level* H_n on day n depends on activities from both the spot and the monthly forward trading. For $n \in [n_{I-1}, n_I)$, H_n is given by:

$$H_n := \sum_{k=0}^{n-1} h^S_k + \sum_{J=1}^{I-2} \big(d^J(n_{J+1}-n_J)\big) + d^{I-1}(n - n_{I-1} + 1)$$

with initially empty storage, i.e. $H_0 := 0$. The optimisation of (5.4.1) is subject to the following constraints:

$$H_N = 0, \tag{5.4.3}$$

and for $n_J \leq k \leq n_{J+1}$, $J \leq m-1$:

$$0 \leq H_k \leq c, \qquad \text{and} \qquad \ell_k - d^J \leq h^S_k \leq u_k - d^J, \tag{5.4.4}$$

and for $\alpha \in [0,1]$:

$$h^J_k \leq \alpha \frac{c}{n_{J+1} - n_J} \tag{5.4.5}$$

Alternatively, the daily constraints ((5.4.4)) can be expressed as iterative daily bounds for the physical storage using

$$\widetilde{\ell}_k := \max\{\ell_k, -H_k\}, \qquad \text{and} \qquad \widetilde{u}_k := \min\{u_k, c - H_k\},$$

so that

$$\widetilde{\ell}_k \leq h_k + d^J \leq \widetilde{u}_k. \tag{5.4.6}$$

Remark 5.4.1. *In the model above, the action of storage on day k is $(h_k^S + d^J)$ for $n_J \le k < n_{J+1}$. And thus, the action of pure spot trading is restricted by the daily delivery amount d^J, which is a result of $h_{\tilde{k}}^J$ with $n_{J-1} \le \tilde{k} < n_J$. In other words, the forward trading activity has a delayed effect on the spot trading, but spot trading does not affect forward trading: the daily delivery quantity will only be fixed for the following month after the forward trading of that contract is finished; and the delivery obligations then have an effect on the spot trading activity of that month, as the sum of daily delivery and the spot trading is bounded by the daily withdrawal and injection rate of the storage problem.*

The constraint (5.4.5) makes sure the maximal amount traded can be stored in case of no spot trading; It can also be seen as some liquidity constraints; moreover, with scaling factor $\alpha \in [0,1]$, we can bound the volume of forward trading and maintain certain balance between the spot and forward trading.

Next, we train the storage action using deep neural networks. The approach is similar to Model I of Chapter 5.3, , except that we additionally model the action on monthly forward using different deep neural networks. For the ease of notation, we abbreviate monthly forward $F_k = F(k, n_J, n_{J+1})$. The model can be formulated as:

- Input: time horizon of storage $\mathbb{T}$, M trajectories of the spot $(S_k^i)_{k\in\mathbb{T};i=1,\dots,M}$, and of rolling month forward $(F_k^i)_{k\in\mathbb{T};i=1,\dots,M}$ respectively;
- Training object: trading strategy which can be split into action of spot h_k^S and action of rolling month forward h_k^J. Let $(g_k^S)_{k\in\mathbb{T}}$ denote the neural networks for $(h_k^S)_{k\in\mathbb{T}}$, and let $(g_k^J)_{0\ge k<n_M}$ denote the neural networks for $(h_k^J)_{k\in\mathbb{T}}$, so that

$$h_k^{i,S} = g_k^S(k, S_k^i), \qquad \text{and} \qquad h_k^{i,J} = g_k^J(k, F_k^i).$$

  In total, we have $N + n_M$ neural networks, each of which has L layers.
- Training criterion: minimize the loss function $\mathcal{L}$, i.e. for $i = 1, \dots, M$ minimize:

$$\mathcal{L}(i) := \widetilde{U}\big(W_{N-1}^i\big) = \widetilde{U}\big(W_{N-1}^{i,S} + W_{N-1}^{i,F}\big),$$

  according to (5.4.2)–(5.4.5).

For the numerical testing, 1000 scenarios of spot as well as 1000 scenarios of monthly forward price curves of 12 months are provided by Axpo Solutions.

From the latter, we computed 1000 rolling monthly forward curves, each of which contains only the first nearby contract on any trading day. The data set is then split into 900 scenarios of spot and of forward prices for training the model, and 100 scenarios for testing the model. For this model, we use the same setup for the modeling of spot trading as in Model I. To recall, model I uses 12 neural networks with $L = 2$, $N_0 = 16$ and $N_1 = 1$. For the modeling of forward trading, we use the exact same setup for 11 months, as in the 12th month, the front month contract delivers beyond the trading horizon of the storage. In detail, for forward trading, we use 11 neural networks with $L = 2$, $N_0 = 16$ and $N_1 = 1$. After the model implementation which is more involved due to larger amount of constraints, it was trained for 1000 iterations, with learning rate=0.05 and batch size=100. Figure 5 visualizes the trading strategy in terms of storage level with respect to different choice of α. Figure 6 provides a visualization and a detailed summary statistics for comparing this model with Model I and benchmark case for Model I. The comparison is based on the same setup for the spot strategy part and on the same training conditions with the exception of risk aversion rate[3]. Compared with the Model I, this model shows in average a significantly higher P&L, and is also more volatile (see *std*). Moreover the P&L is mostly generated from the forward trading activities, and thus, it is highly sensitive w.r.t the choice of α from (5.4.5). Thus it is a delicate desk to set the bound α.

[3]The risk aversion rate r in the loss function can be used to shrink the variance of P&L. As the use of forwards in Model II increases P&L and variance significantly, we use a larger r to restrict the variance to certain extent. For the setup we use $r = 3$ for Model I and $r = 10$ for Model II.

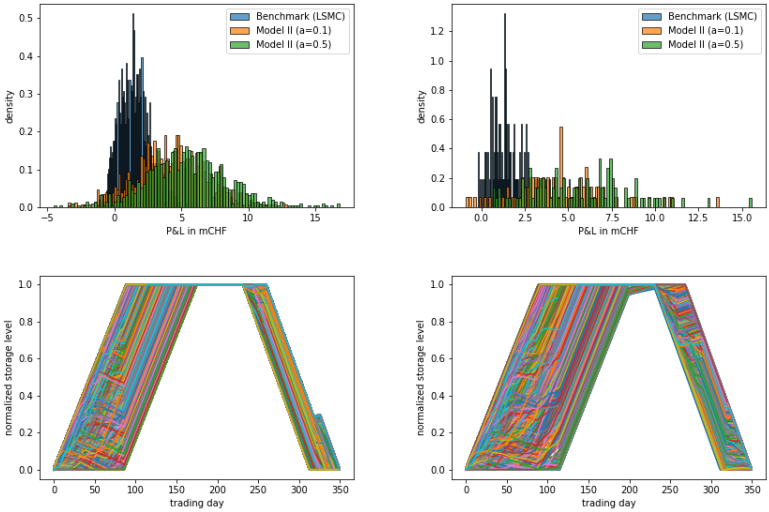
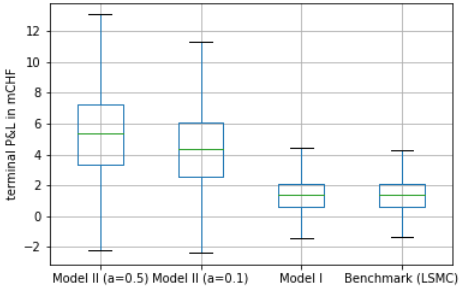


Figure 5: A comparison between performance of Model II (with different $\alpha's$ and that of benchmark (LSMC). The upper plots compare the terminal P&L between Model II ($\alpha = 0.1$ and $\alpha = 0.5$) and benchmark in million CHF: upper left gives the P&L of the training set and upper right gives that of the test set. We see that in both plots the P&L of Model II are significantly larger than that of the benchmark. The bottom plots visualize the storage actions computed using model II with $\alpha = 0.1$ (bottom left plot) and with $\alpha = 0.5$ (bottom right plot). The storage levels are normalized by the capacity c. Similar as in model I, the seasonal pattern of prices and the trading strategies are reflected by the pattern of storage level: buy at lower summer prices and store until capacity is reached; then sell at relatively high winter prices until storage is completely emptied.

We see in Figure 6 that Model II, which allows trading activities on forwards, is clearly the most favourable choice in terms of maximizing the terminal P&L. When compared to Model I, it is slightly more complex, but in terms of computational time and effort, it is still very good manageable on a personal notebook. The high P&L generated by using the Model II comes with high standard deviation of P&L of all scenarios, and is very sensitive to the limitation on forward activity, expressed by α. Thus, one possible direction of future work is to improve the restriction for the forward trading in order to create a more balanced wealth generated from spot and forward trading activities on the one hand, and to generate high P&L with less risk on the other hand.



terminal P&L in mCHF	training set			test set			benchmark
	$\alpha = 0.5$	$\alpha = 0.1$	Model I	$\alpha = 0.5$	$\alpha = 0.1$	Model I	spot only
average	5,429,539	4,284,428	1,440,326	5,655,886	4,425,810	1,436,510	1,448,475
median	5,362,543	4,341,783	1,391,877	5,576,047	4,258,077	1,284,445	1,382,435
std	3,111,824	2,708,985	1,132,284	2,839,238	2,526,534	1,019,118	1,109,619

Figure 6: A comparison between performance of Model I, Model II ($\alpha = 0.1$ and $\alpha = 0.5$) and that of benchmark, both provided as visualization and as detailed summary table with means, medians and standard deviations (std) of the terminal P&L. In the boxplot, we have united the training and testing sets, and have removed the outliers that are outside 1.5 times the interquartile range above the third and below the first quantile. We observe that the terminal P&Ls of Model II are both higher and more volatile than that of Model I or the benchmark (LSMC). Moreover, they depends on the choice of α: the higher the α, the higher the terminal P&L and its std. For Model I, $\alpha = 0$, and thus both P&L and std are the lowest. Note that the benchmark (LSMC) is only the benchmark case for the Model I and is not fully appropriate as benchmark for Model II, as it does not allow for forward trading. When comparing performance on the training sets with that on the test sets for Model II $\alpha = 0.5$, we see that the test set yields better results. This suggest that there is still room for improvement, e.g. on the setup of neural networks or setup of training, or simply continuing training (more iterations). However, for comparison with Model I, with the exception of risk aversion rates for the loss function ($r = 3$ for Model I and $r = 10$ for Model II), we use the same setup for spot strategy and training conditions for all models.

Another possible direction of future work is to include $K > 1$ forward curves for the gas storage problem to maximize the P&L and thus, the utility of the storage manager.

Chapter 6

Long-term model risk in energy market

6.1 Introduction

In a financial institution, the valuation and risk management of financial products rely heavily on the use of models. With the technological advances and needs of higher standards, the amount of models used as well as its complexity has grown significantly. With that, the need of model risk quantification becomes increasingly important. The famous Statistician George E.P. Box once stated: *"All models are wrong, but some are useful."* Indeed all models use some simplifying assumptions as a trade-off to tractability, and capture some important but not all features that can be observed in reality. There is an increasing amount of literature dedicated to this topic, each of which captures some model related risk aspects. Yet, it is important to note that there is no unanimous definition for model risk, as it has a different meaning for different group of professions and people. Instead, model risk is a collection of important aspects in relation to the creation and the use of a model.

This chapter provides a review of model risk, with a strong focus on applicable methods related to derivative pricing and hedging, in particular for long-term energy markets. It is structured as follows: in the remaining part of introduction, we briefly discuss different definitions and aspects of model risk set by the

financial regulators, famous practitioners and some academics. In Section 2, we conduct a survey on model risk literature, with a strong focus on those approaches that have great potential to be applied to adhoc problems in the financial industry. Here we discuss three categories of model risk quantification approaches: the pairwise model comparison, the Bayesian model averaging and the worst case approaches. In Section 3, we then present two applications to quantify model risk in modeling and hedging long-term energy options. In the first application we quantify the model risk in hedging a long-term energy call using a misspecified model. In the second application we quantify the model risk in the modeling of a long-term energy call using an entropy approach, which is proposed by Glasserman and Xu [2014].

For this chapter, we denote by X the stochastic elements of a model (e.g. a random variable, a random vector, or a random process), denote by F_t (instead of $F(X_t, T, T_1, T_2)$) the time-t forward, and denote by $V(X)$ a payoff (e.g. of a call option on F_t, and thus on X).

This chapter is based on joint work with Martin Larsson. The author would like to thank Vlatka Komaric and Markus Regez for fruitful discussions and for providing energy option data.

6.1.1 Regulatory definitions on model risk

In US, the office of the Comptroller of the Currency (OCC) of the federal reserve published in 2011 the *Supervisory Guidance on Model Risk Management*. This document provides the first regulatory definition of model risk, which has since then become the industry standard worldwide. In this document,

"... the use of models invariably presents model risk, which is the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports ..."

According to the document, the wrong model output and the wrong interpretation of model output are identified as the main sources of model risk. This definition however indirectly includes all the errors in the life circle of a model implementation starting from wrong design. The federal reserve further set detailed policies of model validation process which banks and other financial institutions should comply with.

In 2013, the European Banking Authority (EBA) also published their definition of model risk, which is similar to the one published by the OCC, but explicitly stresses particular phases in the life cycle of a model. According

to the article 3.1.11 in the *Capital Requirement Directive (CRD IV)*, model risk is

"... the potential loss an institution may incur as a consequence of decisions that could be principally based on the output of internal models, as a result of errors in the development, implementation or use of such models ..."

6.1.2 Types of model risk

Long before the regulators give their definition and set industry standards, there are already researchers looking at various aspects of model risk.

In a well-known Goldman Sachs research paper published in 1996 and also in Derman [1996], Derman pioneered in identifying and elaborating all aspects of model risk considerations that were crucial to him. While a unanimous definition for model risk was not possible among practitioners, this paper gives a good summary of all aspects considered for model risk in the industry. According to Derman the main aspects of model risk are: the inapplicability of modeling, the use of an incorrect model, the incorrect use of a correct model, the inappropriate use of a correct model, data issue, calibration and re-calibration errors.

In our view, it is questionable whether some of the aspects (incorrect use of a correct model, inappropriate use of a correct model, data issue and calibration errors) should be seen as model risk or rather operational risk. We think that the aspect of incorrect model is crucial. That refers to either incorrect model-based assumptions, e.g. on the dynamics (normal vs. log-normal vs. fat-tailed distribution, or one vs. multi-factors), or the misspecification of variables (models) and parameters.

6.1.3 Model risk vs. model uncertainty

In some but not all academic literature, people differentiate between the so called *model risk* and *model uncertainty*, expressing the modeler's aversion to risk and his aversion to ambiguity respectively. In the former one, the modeler does not have the knowledge of the exact model but knows the probabilities of each models be the true model; in the latter one, the likelihood of each of the model to be the true model is unknown. They are in analogy to the classical Knightian uncertainty (Knight [1921]), Ellberg's paradox (Ellsberg [1961]); see e.g. Uppal and Wang [2003], Gilboa and Schmeidler [1989], Bannör and Scherer [2014]. To demonstrate the idea, consider the situation of having a whole set of

models $\mathcal{S} := \{S^i; i \in I\}$ to choose from for modeling; see Figure 1. We call the risk of choosing a wrong model *model uncertainty*, if there is either no knowledge on the probability measure or many specifications of probabilities are possible. Otherwise, if there is a "known" or "canonical" or "natural" probability measure $\mathbb{P}$ on the set of models $\mathcal{S}$, the risk is referred to as the model risk. Following this definition, model risk can be seen as a special case of model uncertainty. However, we want to stress that the definition stated in this section is one way of model risk consideration. In much of the literature, the two terms are used interchangeable. People argue that it is possible to pre-define a unique distribution on the set $\mathcal{S}$, thus the *model uncertainty* can then be treated the same as *model risk*.

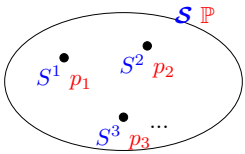


Figure 1: The risk of choose a wrong model out of set $\mathcal{S}$ is called model risk or model uncertainty, depending on whether or not a unique distribution $\mathbb{P}$ exists.

6.2 Model risk quantification - a review

In much of the literature two aspects of model risk have been addressed frequently: the *risk of model misspecification* and the *parameter uncertainty*. Given a rich pool of models $\mathcal{S}$, the risk of have chosen the incorrect model $S^i \in \mathcal{S}$ is referred to as model misspecification; Similarly, given a model with a rich pool of parametrizations, $\mathcal{S}^\Theta := \{S(\theta); \theta \in \Theta \subset \mathbb{R}^d\}$, the risk of having chosen the incorrect θ (and hence use the incorrect model $S(\theta)$), is referred to as *parameter uncertainty*. These two aspects can be seen as equivalent, if the model $\mathcal{S}^\Theta$ is flexible and general enough, so that every parametrization $\theta_i \in \Theta$ corresponds to a unique distribution, thus represents a model $S^i := S(\theta_i)$.

The general approach of model risk quantification or in identifying the best model is to first choose a set of models $\mathcal{S}$ that we consider. Then we define a mapping f, to quantify risk of each model $f(S)$ with $S \in \mathcal{S}$ or the risk of all models in $\mathcal{S}$. Based on the valuation, one can choose the best model or estimate the averaged or the worst case of risk.

The mapping f is referred to as the *risk measure*, which is defined as follows:

Definition 6.2.1 (Risk measure, cf. Bannör and Scherer [2014], Chap. 3.1)**.** *Let $\mathcal{X}$ be a collection of random variables on probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$, i.e. risk-exposed quantities. Let $\pi : \mathcal{H} \to \mathbb{R}$ be a linear mapping on a subcollection of random variables $\mathcal{H} \subset \mathcal{X}$ and let $V : \mathcal{X} \to \mathbb{R}$ be a function*[1]*.*

V is called a risk measure w.r.t. π, if V fulfills the following axioms:

- *V is monotone, i.e. for $X, Y \in \mathcal{X}$ and $X \geq Y$, $V(X) \geq V(Y)$;*
- *V is π-translation invariant, i.e. for $X \in \mathcal{X}$ and $Y \in \mathcal{H}$ the equality $V(X + Y) = V(X) + \pi(Y)$ holds.*

Furthermore, V may have these additional properties:

- *V is called convex, if for $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$, $V(\lambda X + (1 - \lambda)Y) \leq \lambda V(X) + (1 - \lambda)V(Y)$ holds;*
- *V is called coherent, if it is convex and positively homogeneous, i.e. for $X \in \mathcal{X}$, and $c > 0$, $V(cX) = cV(X)$ holds.*

In the following, we briefly sketch three different approaches.

6.2.1 Approach I: pair-wise model comparison, parameter perturbation

Suppose a trader wants to hedge a derivative, say an european-style call. The call was priced using the correct model, say *model A*, which is not known to the trader. He then assumed a distribution which is unfortunately incorrect, Subsequently, he ended up choosing a different model than *model A*, say *model B*. How bad is then his hedge due to the model misspecification? This question motivates the following type of pair-wise model comparison.

Let $\mathcal{S} = \{A, B\}$ be the set of models we consider, with *model* $A = S(\theta_1)$ and *model* $B = S(\theta_2)$. We assume that:

- *model A* is the true model;
- *model B* is a different model that will be tested.

[1] $\mathcal{H}$ can be thought of as the set of risk-less positions or bank accounts and the mapping π as the identity mapping.

Note that here we keep the assumptions reasonably simple and generic and thus omit any assumption on the relationship between *model A* and *model B*. In some papers, the *model A* is assumed to be quite different and significantly more complex than *model B*. In others, they are two models from a model class. As an example, *model B* is a completely different parametrization or a perturbation of *model A* in one parameter. In the latter case, we have that $\exists j$: $\theta_1^{(j)} \neq \theta_2^{(j)}$ and $\theta_1^{(k)} = \theta_2^{(k)}$ for $k \neq j$ and $k = 1, ..., d$.

The procedure of investigating the model misspecification can be sketched as:

- Determine the parameters of *model A* by fitting it to the market data
- Calibrate *model B* to the data generated by *model A*
- Compare the pricing or hedging performance of *model B* and *model A*

This approach has been explored in various papers, mostly in simulation based analysis; see (among others) e.g. Hull and Suo [2002], Longstaff et al. [2001], Melino and Turnbull [1995], Green and Figlewski [1999], Driessen et al. [2003], Hilscher et al. [2020], Schröter et al. [2012]. To the best of the author's knowledge, Karoui et al. [1998] is the only paper that derived an analytic solution of pricing error for an European-style call option due to perturbation of implied volatility in the classical Black-Scholes setting.

The main shortage of this approach is that the knowledge gained from the pair-wise model comparison is specific to the choice of models considered and does not allow broad knowledge transfer. Nevertheless, it is a good approach in selecting between alternative models, and in model risk quantification regarding hedging using a misspecified model.

We will later use this approach to investigate model risk regarding hedging, when *model A* is used for pricing and *model B* is used for hedging a derivative; see Chapter 6.3.1 for more details.

6.2.2 Approach II: Bayesian model averaging

As the name already suggests, the Bayesian approach relies heavily on the use of conditional probabilities. The underlying assumption is the narrow definition of model risk provided in Section 6.1.3: given a set of models $\mathcal{S}$ and a set of observations x, there exists a probability distribution $\mathbb{P}$ on $\mathcal{S}$, which reflects the subjective beliefs about the likelihood of each of the models to be the true

model[2]. Moreover, one assumes that the relationship between the distribution $\mathbb{P}$ and data y is not static. Therefore a dynamic Bayesian process is needed in which the knowledge of the model distribution $\mathbb{P}$ is constantly enhanced and updated.

Let $\mathcal{S} = \{S^i(\theta); i \in \{1, ..., N\}, \theta_i \in \Theta_i \subset \mathbb{R}^d\}$ be the set of models, and let $\Theta := \Theta_1 \times \Theta_2 \times ... \times \Theta_N$ be the set of parametrizations for $\mathcal{S}$, where each Θ_i is the (finite) set of all possible parametrizations for model S^i. Hence, there are two variables we optimize and update for in the Bayesian approach: the likelihood of each model S^i to be the true model, and the probability of using a specific θ_i from Θ_i given the model S_i. We start with prior model weights $\mathbb{P}(S^i)$ and prior (density) of the model parameter given model S^i, i.e. $p(\theta_i|S^i)$. Given a data set y, the posterior probability for model S^i is

$$\mathbb{P}(S^i|\,y) = \frac{p(y|S^i)\mathbb{P}(S^i)}{\sum_{j\in I} p(y|S^j)\mathbb{P}(S^j)},$$

where $p(y|S^j) = \int_{\Theta_j} \mathbb{P}(y|\theta_j, S^j)p(\theta_j|S^j)d\theta_j$ expresses the likelihood of having data y given model S^i. Figure 2 briefly sketches the updating procedure with an example of using second moment computation to measure dispersion across models of $\mathcal{S}$. For general reference on Bayesian statistics, see e.g. Bernardo and Smith [2009], for model-averaging references, see e.g. Hoeting et al. [1999], Raftery et al. [1997], Kass and Raftery [1995], for model risk specific reference see e.g. Bannör and Scherer [2014], Cont [2006].

This technique is widely used as a model selection method on (mostly) simple model structures, i.e. linear models or simple regression-type of models. The updating procedure improves predictive ability of some model-dependent target quantity $V(X)$, and the result is very stable. It is not used in quantifying model risk, nonetheless it is widely considered a method of model risk minimization, selection of best model, and improvement of model predictability. The main critics of this technique are twofold: on the one hand this technique required prior information, which would need certain probabilistic sophistication on the modeler side; on the other hand, if we apply it on Black-Scholes type of stochastic models, the computational effort is significant and cannot be directly applied to a big sample, as the Baysian estimator does not have a closed-form solution;

[2]This idea splits statisticians into two camps: The Bayesian statistician and the frequentist statistician who believes in a true but unknown model that exists and that one cannot assign probabilities to different "candidate models". This discussion is similar to the one we had previously in Section 6.1.3.

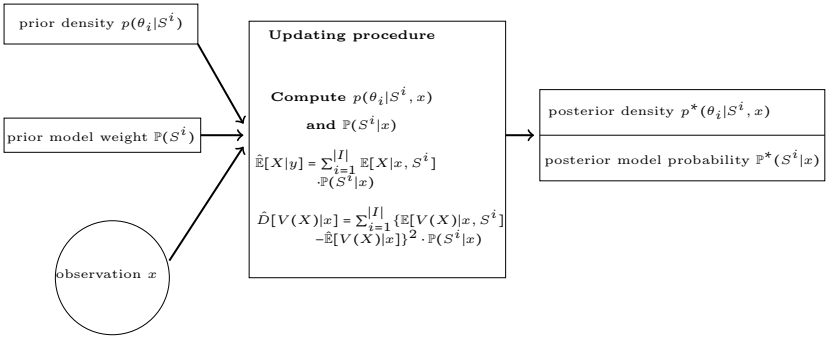


Figure 2: The Bayesian updating process. Suppose we have forward data x and want to compute the a vanilla K-striking call with the payoff $V(X) := (x - K)^+$. In each round of the updating process, it evaluates the call X in every model, and computes the weighted average value over all models ($\hat{\mathbb{E}}[V(X)|y]$) and the dispersion $\hat{D}[V(X)|y]$, where the weights were the posterior probability from the last round.

see e.g. Jacquier and Jarrow [2000] where specific inference on information of parameters and the Markov-Chain Monte-Carlo method are needed.

6.2.3 Approach III: worst case approach

This approach is the most popular one in our community. It is motivated by the narrow definition of *model uncertainty* of Section 6.1.3, namely given a set of models $\mathcal{S}$, the likelihood of each model is not known or cannot be identified. And thus, the best one can do is to compute the all valuation of $V(X)$ and take the worst case valuation as a measure for model risk.

The most influential early work goes back to Gilboa and Schmeidler [1989], where they provide a firm axiomatic system, in which a risk-averse agent facing uncertainty/ambiguity can choose among a set $\mathcal{A}$ of feasible alternatives by the means of a max-min expected utility theory, i.e. $\max_{X \in \mathcal{A}} \min_{S^i \in \mathcal{S}} \mathbb{E}^{S^i}[U(V(X))]$. Here the risk aversion of the modeling agent is expressed through the utility function U, and the aversion to ambiguity is captured by taking the minimum of all models in $\mathcal{S}$. In another notable work, Cont [2006] sets out a quantitative framework with a particular focus on derivative pricing. Note that for option pricing only risk-neutral measures are used. To reflect that, Cont assumes that a set of risk-neutral probability measures $\mathcal{S}$ is available for valuation of some claim

(option) X, $\mathcal{S} = \{\mathbb{Q} : \mathbb{Q} \text{ is a risk-neutral measure}\}$, and no further information is available (in the sense of model uncertainty). Here, each risk-neutral measure $\mathbb{Q}$ defines a model $S^{\mathbb{Q}}$, therefore the set $\mathcal{S}$ is a set of models. Then, with no information about likelihood at hand, the worst-case bounds (best and worst prices) are:

$$l(X) = \inf_{\mathbb{Q}\in\mathcal{S}} \mathbb{E}^{\mathbb{Q}}[V(X)], \qquad u(X) = \sup_{\mathbb{Q}\in\mathcal{S}} \mathbb{E}^{\mathbb{Q}}[V(X)],$$

which can be interpreted as the *bid-ask* prices. Note that u fulfills the axiom of a coherent risk measure[3] (according to Definition 6.2.1); see e.g. Föllmer and Schied [2011]. Moreover, the model risk can be quantified as

$$\xi(X) = u(X) - l(X) = \sup_{\mathbb{Q}\in\mathcal{S}} \mathbb{E}_{\mathbb{Q}}[V(X)] - \inf_{\mathbb{Q}\in\mathcal{S}} \mathbb{E}_{\mathbb{Q}}[V(X)],$$

which is the maximal impact that model risk can have on $V(X)$. In case that one might have additional information about the trustworthiness of some model $\mathbb{Q}$, Cont [2006] suggests to extend the existing bounds by adding a penalization term $\alpha : \mathcal{S} \to \mathbb{R}_0^+$. In essence, the bounds are:

$$\tilde{l}(X) = \inf_{\mathbb{Q}\in\mathcal{S}} \mathbb{E}^{\mathbb{Q}}[V(X)] - \alpha(\mathbb{Q}), \qquad \tilde{u}(X) = \sup_{\mathbb{Q}\in\mathcal{S}} \mathbb{E}^{\mathbb{Q}}[V(X)] - \alpha(\mathbb{Q}),$$

where $\tilde{u}$ is a convex risk measure; see e.g. Föllmer and Schied [2002] for details on convex risk measure. The coherent risk measure of Cont [2006] has been applied (among others) in a slight modification to quantify model risk for gas storage valuation problems in Henaff et al. [2013]. They examined historically estimated parameter risk associated with the storage valuation. Instead of a set of benchmark instruments as in Cont [2006], Henaff et al. [2013] start with two proposed spot price models for gas with spikes, calibrate them to the historical prices using a maximum likelihood method, and then perturb the optimal parameters with certain constraints to obtain the set of models. The model risk (parameter risk) is then measured as the difference between the highest value and the lowest value of $V(X)$ among the perturbed models, normalized by the value of $V(X)$ using the base model.

Another different and very notable method for studying model risk quantification was the concept of *risk-captured prices*; see e.g. Bannör and Scherer [2013, 2014], Bannör et al. [2016]. Let $\mathcal{S}^{\Theta}$ be a parametrized family of models ,

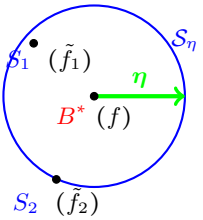
[3] Example of such measures are the average value at risk, or simply value at risk assuming a normal distribution.

where the parameter space Θ is equipped with a distribution $\mathbb{P}$ expressing the likelihood of validity of each model $S^\theta \in \mathcal{S}^\Theta$. Let ξ be a convex risk measure, then the *risk-captured prices* of claim X is defined as

$$\Gamma(X) := \xi(\theta \mapsto \mathbb{E}_\theta[V(X)]), \quad \hat{\Gamma}(X) := -\Gamma(-X).$$

$\Gamma(X)$ can be interpreted as the *ask* price and $\hat{\Gamma}(X)$ as the *bid* price. This notion provides a nice framework for model risk quantification. In particular, Henaff et al. [2013] showed (see Proposition 4.2) that if ξ is chosen as the average value at risk (AVaR), then the AVaR induced risk-captured price is continuous w.r.t. the topology of weak convergence on the parameter distribution when the pricing function $\theta \mapsto \mathbb{E}_\theta[X]$ is continuous and bounded. This provide justification for the use of the asymptotic $\hat{\mathbb{P}}$ instead of the $\mathbb{P}$ in case $\hat{\mathbb{P}}$ is more tractable and convenient to use, and makes this framework very application friendly. As a consequence, this method has been applied (among others) for valuation of a gas-fired power plant; Bannör et al. [2016] proposed a multi-factor structural model for gas-fired power plant, and examined the impact each model parameter has and expressed the model risk in terms of relative width of bid-ask-spreads[4].

Last but not least, a very elegant robust approach of model risk quantification using relative entropy was proposed by Glasserman and Xu [2014], which requires relatively small computational efforts beyond the modeling of a baseline model. The procedure is outlined in the following:



1. Fix a "nominal model" or "base" model, denoted B^*
2. Consider a set of models $\mathcal{S}_\eta$ with a given "distance" η to the base model B^*
3. "Model risk" is given by the worst case (out of all models within a given distance) for a given payoff

Moreover, they suggest to use relative entropy to measure the "distance" between distributions (and thus between models): the relative entropy (in the Bayesian sense) measures information gained through additional data, and thus it is *"a measure of the additional information required to make a perturbed model preferable to a baseline model"*. Following Glasserman and Xu [2014], let f be the density of the nominal model for modeling a contingent claim $V(X)$ and $\tilde{f}$

[4] In essence, they computed $\left(\Delta = \frac{\text{price}_{\text{bid}} - \text{price}_{\text{ask}}}{\text{price}_{\text{mid}}}\right)$. They concluded that spike risk is the most important parametric risk within their model.

the density of the alternative model. Further, assume the *likelihood ratio* $m = \frac{\tilde{f}}{f}$ is well-defined. Then the *relative entropy*[5] (or Kullback-Leibler Divergence) is given by

$$\mathbb{E}[m \ln m] := \int \frac{\tilde{f}(x)}{f(x)} \ln \frac{\tilde{f}(x)}{f(x)} f(x) dx.$$

m is also known as the Radon-Nikodym-density, which can be interpreted as a measure change from nominal model B^* (with density f) to an alternative model, e.g. S_1 (with $\tilde{f}_1$).

Glasserman & Xu consider alternative models described by a set $\mathcal{S}_\eta$ of likelihood ratios m for which $\mathbb{E}[m \ln m] \leq \eta$. The trick is that for an alternative model with density $\tilde{f}$, the valuation of $V(X)$ w.r.t. $\tilde{f}$, $\mathbb{E}^{\tilde{f}}[V(X)]$, can be transformed back to a valuation w.r.t. the base model:

$$\mathbb{E}^{\tilde{f}}[V(X)] = \int V(x)\tilde{f}(x)dx = \int V(x)\frac{\tilde{f(x)}}{f(x)} f(x)dx = \mathbb{E}[m(X)V(X)].$$

And hence, the model risk is bounded by

$$\inf_{m \in \mathcal{S}_\eta} \mathbb{E}[m(X)V(X)] \quad \text{and} \quad \sup_{m \in \mathcal{S}_\eta} \mathbb{E}[m(X)V(X)]$$

The upper bound $\sup_{m \in \mathcal{P}_\eta} \mathbb{E}[m(X)V(X)]$ can be solved explicitly using the Lagrangian multiplier[6]. The constraints of the upper bounds are $\mathbb{E}[m(X) \ln m(X)] \leq \eta$ and $\mathbb{E}[m(X)] = 1$. Thus, with $\theta > 0$ and $\mu > 0$, the Lagrangian is given by:

$$\sup_m \mathbb{E}\Big[\ m(X)V(X) - \frac{1}{\theta} \cdot \{m(X) \ln m(X) - \eta\} + \mu \cdot m(X)\ \Big].$$

The first order condition yields $V(x) - \frac{1}{\theta}(\ln m(x) + 1) + \mu = 0$. Rearranging gives then

$$m(x) = e^{\theta V(x)} \cdot e^{\theta\mu - 1}.$$

Moreover, as m is a likelihood ratio ($\mathbb{E}[m(X)] = 1$), we obtain

$$\mathbb{E}[e^{\theta V(X)}] = e^{1-\theta\mu}.$$

[5] This actually goes back to extensive prior literature on model "robustness" (robust statistical estimation and robust optimal control), which is used in economics but received little attention in finance.

[6] The lower bound can be solved similarly (with negative θ), and thus we omit that.

Hence, for a fixed θ:

$$m^*(\theta, X) = \frac{e^{\theta V(X)}}{\mathbb{E}[e^{\theta V(X)}]}$$

And thus, the *worst case* (or upper bound) is given by:

$$\mathbb{E}[m^*(\theta, X)V(X)] = \frac{\mathbb{E}[V(X) \cdot e^{\theta V(X)}]}{\mathbb{E}[e^{\theta V(X)}]}. \tag{6.2.1}$$

We see that the worst case model risk is characterized by an exponential change of measure defined through V and $\theta > 0$. Note that here the entropy (or level of uncertainty) η is now linked to θ in the following way: for a given $\eta > 0$, we can find an optimal $\hat{\theta}$ with $\hat{m}(\hat{\theta})$ such that

$$\eta = \mathbb{E}\big[\hat{m}(\hat{\theta}) \ln \hat{m}(\hat{\theta})\big].$$

The advantage of Glasserman and Xu [2014] become very obvious now, namely the computation of model risk, (6.2.1), is fully explicit; Moreover, it is essentially the estimation of $V(X)$ under the baseline model (with e.g. a Monte-Carlo simulation), without the estimation of all alternative models within the level of uncertainty η, and thus is computationally very efficient. Similar approaches have been explored by e.g. Feng and Schlögl [2018], Feng et al. [2018]. We will give an example on energy options using this method; see Chapter 6.3.2.

The worst-case approaches are in general very robust, and require less sophisticated inputs compared to the model averaging approach. The main concern here is that it gives a too conservative model risk estimate.

6.3 Application to energy markets

Hull and Suo [2002] states that in liquid markets *the specification of a model is not usually a significant issue in the pricing and hedging of "plain vanilla" instruments*. They argue that if a market has sufficient liquidity, a trader can back out a whole implied volatility surface from all liquidity traded instruments and update them frequently, and thus the model misspecification has much less impact on the option pricing and on the hedging performance. While this is certainly true in a liquid market like the Equity markets, these assumptions do not apply in the illiquid mid- to long-term energy markets, where the trader solely relies on a model he believes in. And thus, the risk of have chosen the wrong model (that is an insufficient model which does not fully capture all necessary

features) is a significant concern. In the following we shows two examples which illustrate the approaches of model risk quantification in option pricing (Chapter 6.3.2) and hedging (Chapter 6.3.1).

6.3.1 Application I: model risk in hedging energy option

Imagine we hold an illiquid energy call option with payoff $V(X)$ and want to hedge it using a model. Here the real energy dynamics is very complex and can only be described by a sophisticated and unknown model, say *model A*. Moreover, we don't have the knowledge of the model that describes the reality or don't have the sophistication to work with the reality, and thus, we choose to use a simple, tractable model for hedging, say *model B*. How big is the error of model risk (model misspecification), if we price and hedge w.r.t. *model B*? It is essential to notice that there are actually two types of errors when comparing the model-based hedge to the real option price:

- model-based tracking error of the hedge: this shows how well the model-based hedge tracks the model based price; let's denote it $\tilde{\varepsilon}_t$.
- model misspecification error: this shows how good the model based price tracks the real price; let's denote it $\tilde{\gamma}_t$.

To be more precise, let ε_t denote the discrepancy between the value of the hedge strategy $H^B(X)$ (using model B) and real call value at time t (with model A). That is:

$$\varepsilon_t := H^B(X_t) - \mathbb{E}^A[V(X_t)]. \tag{6.3.1}$$

Then ε_t can be written as $\varepsilon_t = \tilde{\varepsilon}_t + \tilde{\gamma}_t$, with

$$\tilde{\varepsilon}_t = H^B(X_t) - \mathbb{E}^B[V(X_t)], \qquad \tilde{\gamma}_t = \mathbb{E}^B[V(X_t)] - \mathbb{E}^A[V(X_t)]$$

For a simple example, we assume the reality is described by the *Heston model*, i.e. *model A = Heston model*, and we misspecify it as the *Black-Scholes model*, i.e. *model B = BSM*. That is, BSM is used to calibrate option price and to conduct delta hedging[7]; see Apendix 6.A for more details on the models. For the

[7] In order to make Heston model and BSM comparable, we choose $\tilde{\sigma} = \sqrt{\theta}$, where θ expresses the long-term variance of the stochastic variance process v_t.

computation of $\tilde{\varepsilon}_t$ in the BSM, Karoui et al. [1998] gives a simple semi-analytical formula:

$$\begin{aligned}\tilde{\varepsilon}_t &= \frac{1}{2}e^{rt}\int_0^t \left([\tilde{\sigma}^2 - v(u)]\cdot F_u^2\cdot\Gamma(u,F_u)\right)du \\ &\approx \frac{1}{2}e^{rt}\sum_{i=1}^{N-1}([\tilde{\sigma}^2 - v_{i\Delta t}]\cdot F_{i\Delta t}^2\cdot\Gamma_{i\Delta t}\cdot\Delta t),\end{aligned} \tag{6.3.2}$$

where $\Gamma(t,F_t)$ is the Black-Scholes gamma with $\Gamma(t,F_t) = \dfrac{N'(d_1)}{\tilde{\sigma}F_t\sqrt{T-t}}$ and $\Delta t = T/N$. With that, it is easy to quantify the model risk as the residual error, namely $\tilde{\gamma}_t = \varepsilon_t - \tilde{\varepsilon}_t$.

Table 1 gives a summary statistics of all three errors ε_T, $\tilde{\varepsilon}_T$ and $\tilde{\gamma}_T$, at maturity T of the simulation; we find that the effect of model risk $\widetilde{\gamma}_t$ outweighs the model-based tracking error of the hedge $\widetilde{\varepsilon}_t$. Here, the parameter set is chosen arbitrarily. To see if this observation holds more generally, we conduct sensitivity analysis w.r.t. each of the parameters $(\kappa, \theta, \tau, \sigma)$ while having the remaining parameters unchanged. Figure 6.B.1 in Apendix 6.B provides a visualization of the result in QQ-plot, which confirms the observation from Table 1.

	Min	1st Quantile	Median	Mean	3rd Quantile	Max
ε_T	−0.03512	−0.004373	−0.000011	0.0001766	0.004561	0.03743
$\widetilde{\varepsilon}_T$	−0.008187	−0.000936	−0.000137	−0.000298	0.000423	0.004253
$\widetilde{\gamma}_T$	−0.03743	−0.004237	0.000296	0.0004746	0.004962	0.03833

Table 1: Summary statistics of ε_T, $\tilde{\varepsilon}_T$ and $\tilde{\gamma}_T$. The procedure is as follows: on a discretized time grid with N time points, we simulate M forward paths according to the Heston model using an Euler-scheme; At each of I points in time (every N/I-th time point) and for each forward path, we compute the Black-Scholes delta hedge, the accumulated value process of the hedge $H^B(X_t)$; the total error ε_t can be computed via (6.3.1), the tracking error of the hedge $\tilde{\varepsilon}_t$ can be computed via (6.3.2); The model error is then simply the residual. In the simulation, we set $M = 1000$, $N = 10,000$, $I = 1000$. Other parameters used for the samples are: $F_0/K = 1$, $\tilde{\sigma} = 0.8$, $r = 0$, $\kappa = 0.5$, $\theta = \tilde{\sigma}^2$, $\rho = 0.5$, $\sigma = 0.5$, $\tau = 1$.

6.3.2 Application II: model risk in pricing energy option

In long-term energy markets, the illiquidity is a big issue. In terms of the valuations of forwards and options, this means that they highly depend on the (subjective) choice of model one makes, and thus is vulnerable towards model risk. Therefore, it makes sense to look at a large set of models (with a certain specification to precisely express how large) and without pre-specifying any likelihood of the validity of each of the models. In this context, the Glasserman and Xu [2014] approach comes in handy, as we want to consider a set of models as large as possible, yet we wish to evaluate them so that the computational efforts used are minimal. In the following we give an example to illustrate the application of Glasserman and Xu [2014] to a long-term energy option.

Our data set[8] contains the volatility surface (market quotes) of the Calendar year 2023 German power options (priced as of 23.August 2017). Based on the data, we calibrate the *Bates model*, the *Variance Gamma Scaled-Self Decomposable model (VGSSD)* and the simple *Black-Scholes model (BSM)* to the option volatility surface, where we assume the *Bates model* to be our baseline model; see Appendix 6.C for more details on the models. A visual comparison of these three calibrated models in terms of market implied densities is provided in Figure 3: the *Bates model* seems to have the best balance between capturing the "unsmoothness" and the the fat-tail behavior (log-normal distribution) of the market data. Moreover, we observe that the distance between VGSSD and the baseline model is larger than the distance between BSM and the baseline model. We then determine the entropy distance η's between distributions of the alternative models and the baseline model. For each uncertainty level η, we compute the worst-case option price using the robust Monte-Carlo method proposed by Glasserman & Xu; see Apendix 6.D for more details of the robust Monte-Carlo. The results are listed in the Table 2. We see that the deviation of VGSSD from the baseline model is bigger (as $\eta_{\text{VGSSD}} > \eta_{\text{BS}}$ and also $\theta_{\text{VGSSD}} > \theta_{\text{BS}}$). Therefore, the amount of models considered within the distance η_{VGSSD} is larger than that of η_{BS}, and thus the upper bound $\sup \mathbb{E}[m(\theta_{\text{VGSSD}})V(X)]$ is larger. The relationship between η, $\mathbb{E}[V(X)]$ and $\sup_m \mathbb{E}[m(X)V(X)]$ is shown in Figure 4: the bigger η, the higher the upper bound for $V(X)$. This approach can easiliy put "a number" to the model risk. Yet, it is still an open research question about how to choose the entropy level

[8] The data set was provided by Axpo Solutions, for which we are very grateful.

(uncertainty level) for a sufficient model risk consideration. One possible way of answering this question is to draw link from entropy budget to some familiar conditions: e.g. one calibrate the baseline model to the perturbed power option price levels (say 10% increase or decrease of the prices), calculate the entropy distance of the purturbed distribution η and then compute the worst-case power option prices of that η.

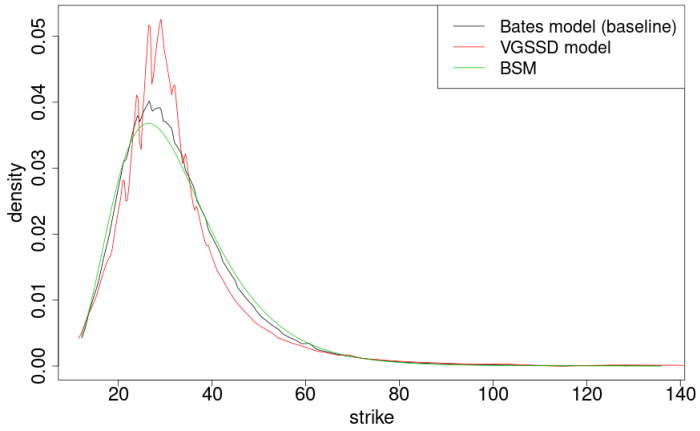


Figure 3: This figure shows a model comparison in terms of the market implied densities of the different models w.r.t. the strikes for German Cal23 power calls. This computation relies on the formula of *Breeden and Litzenberger*. We briefly illustrate the procedure: given a fixed maturity T (here: from 23/08/2017 to 01/01/2023, expressed in years), one extracts the call strike K and its corresponding market price $C(K,T)$, and calibrates the model to the data using some minimization technique such as least-squares. Once calibrated, one computes the model-based call prices for each model and creates the call price curve. Note that it is common practice to imply Black-Scholes volatility for each strike considered from each calibrated model and create volatility curves. In case of any missing quotes or gaps to fill, one can use interpolation techniques. Next, the market implied density (of each model) $f(K)$ can be computed by the *Breeden and Litzenberger* formula, namely $f(K) = e^{rT}\frac{\partial^2 C(K,T)}{\partial K^2} \approx e^{rT}\frac{C(K+\Delta_K,T) - 2C(K,T) + C(K-\Delta_K,T)}{(\Delta_K)^2}$, where Δ_K denotes the strike grid size. We see that VGSSD is best in capturing the "unsmoothness" of the market data but is not capturing the slow-decaying property, while BSM is best in capturing log-normal but is completely smooth. Bates provides a good balance between capturing the "unsmoothness" and the slow-decaying distribution property. Moreover, BSM is closer to Bates in distribution than VGSSD model, suggesting the information gap in terms of relative entropy between BSM and Bates are smaller than that between VGSSD and Bates.

	Bates Model (baseline)	VGSSD	BSM
η	0	$1.82 \cdot 10^{-6}$	$4.59 \cdot 10^{-7}$
θ_η	0	$1.92 \cdot 10^{-4}$	$9.41 \cdot 10^{-5}$
$\sup \mathbb{E}[m(\theta)V(X)]$	4.934	4.953	4.943
model risk	0	0.019	0.009

Table 2: This table gives the results of the example using the Glasserman and Xu [2014] approach. From the calibrated models, one first computes the likelihood ratios m between the alternative models and the baseline model (Bates). Based on that, one can then compute the uncertainty level η, the penalization factor θ as well as the upper bound of the at the money option price $V(X)$ given the η, i.e. $\sup \mathbb{E}[m(\theta_\eta)V(X)]$. The row $\sup \mathbb{E}[m(\theta_\eta)V(X)]$ provides the highest valuation of $V(X)$ given the maximal entropy distance in row η (or the corresponding value in row θ). As an example, if we consider all models with a maximal distance of $\eta_{\text{VGSSD}} = 1.82 \cdot 10^{-6}$ to the baseline model, then the highest price for $V(X)$ is 4.953. From the table we see that the VGSSD is further away from the baseline model, as η_{VGSSD} is bigger (also seen in Figure 3). And therefore, the upper bound of $V(X)$ given the maximal distance η_{VGSSD} is higher. In other words, the bigger deviation on information (in η) we allow for our model selection, the more models we are considering in valuation of $V(X)$, and thus the larger the upper bound of valuation of $V(X)$.

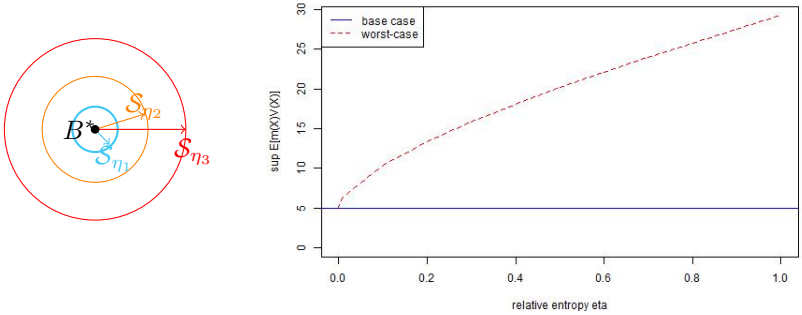


Figure 4: The figures show the relationship between the entropy level η and the worst-case model risk number $\sup \mathbb{E}[m(\theta_\eta)V(X)]$: the higher the η, the higher the upper bound of $V(X)$. Left plot gives a visualization of this model risk consideration, where the amount of models considered is restricted to each ball: the bigger the radius η, the more models are considered, thus the bigger the upper bound. Right plot gives the upper bound for the prices of the at the money Cal23 German power option w.r.t. different η.

6.A Models for Example I

In the following, we briefly review the Heston model and the BSM. In the BSM, proposed in [Black, 1976], F_t is a geometric Brownian motion, given by

$$\frac{dF_t}{F_t} = rdt + \tilde{\sigma}dW_t$$

The well known Black-Scholes formula states that the time-t call price with payoff $V(X)$ and maturity T, is given by

$$\mathbb{E}[V(X)] = N(d_1)F_t - N(d_2)K \cdot e^{-r(T-t)},$$

where N denotes the cumulative distribution function of a standard normal distribution and

$$d_1 = \frac{1}{\tilde{\sigma}\sqrt{T-t}}\left(\ln(\frac{S_t}{K} + (r + \frac{\tilde{\sigma}^2}{2})(T-t)\right) \quad \text{and} \quad d_2 = d_1 - \tilde{\sigma}\sqrt{T-t},$$

The biggest weakness of this model is that the implied volatility is constant, irrespective of quotation date, delivery date, expiry date and strike[9].

The Heston model[10] was proposed in Heston [1993] which extends the BSM by a stochastic volatility factor $\sqrt{v_t}$. The time-t forward F_t is given by

$$\begin{aligned}\frac{dF_t}{F_t} &= rdt + \sqrt{v_t}dW_t^1 \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_t^2,\end{aligned}$$

where $d\langle W^1, W^2\rangle_t = \rho dt$. Note that the stochastic variance modeled v_t is non-negative, and for Monte-Carlo simulation where this cannot be guaranteed, one might need additional conditions (such as $v_t = max(v_t, 0)$ or $v_t = |v_t|$).

[9] In reality the implied volatility is different for options with different characteristics, and also changes from one day to the next.

[10] In terms of electricity option modeling, this model can capture the long-term smile usually quite well but not the short-term smile.

6.B Sensitivity analysis of Example I

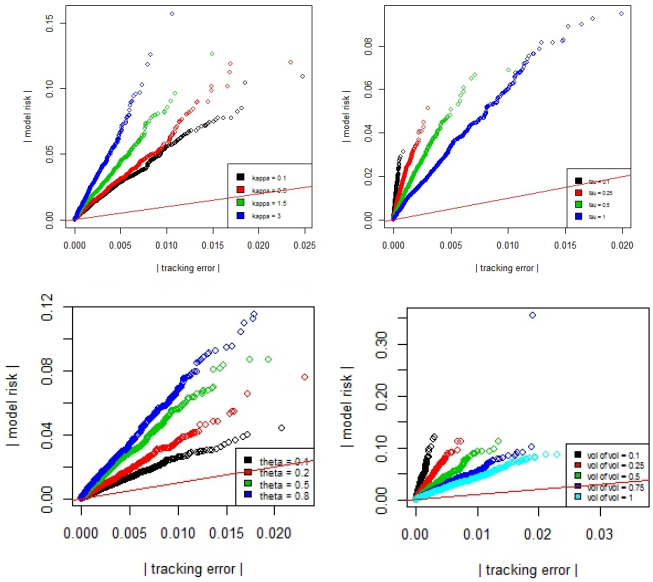


Figure 6.B.1: Sensitivity analysis of parameters κ (top left), τ (top right), θ (bottom left), σ of Heston v_T (bottom right). Each one is a QQ-plot, plotting the absolute tracking error $|\tilde{\varepsilon}_T|$ against absolute model risk $|\tilde{\gamma}_T|$, when one parameter is varied while the remaining parameters of Table 1 stays unchanged. The red line provides the diagonal. In all plots we see that the model risk outweighs the tracking error, which confirms what was seen in Table 1.

6.C Models for Example II

In the following, we briefly review the Bates and the VGSSD model.

The Bates model was proposed in Bates [1996], which combines the stochastic volatility of Heston [1993] and pure jump process of Merton [1976]. The time-t

forward F_t is given by

$$\begin{aligned}\frac{dF_t}{F_t} &= -\left(e^{\mu_J - \frac{1}{2}\sigma_J}\right)\lambda dt + \sqrt{v_t}dW_t^1 + dZ_t \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_t^2 \\ Z_t &= \sum_{i=1}^{N_t} D_i, \text{ with } D_i \overset{iid}{\sim} \mathcal{N}(\mu_J, \sigma_J),\end{aligned}$$

where $d\langle W^1, W^2\rangle_t = \rho dt$. This model can fit short-term smiles (due to Merton) and long-term smiles (due to Heston) well.

The VGSSD process is a time-changed variance gamma process, and can be constructed from the variance gamma process: define the scaled stochastic process $F(t)$ such that it is in law equal $t^\gamma F_{VG_1}$, where F_{VG_1} is a variance gamma random variable at unit time. The characteristic function is given by

$$\phi F_t(u) = \phi F_{VG_1}(ut^\gamma) = \left(1 - iut^\gamma v\theta + \frac{1}{2}u^2 t^{2\gamma} v\sigma^2\right)^{-\frac{1}{v}}.$$

Compared to variance gamma processes[11], the VGSSD has constant skewness and kurtosis of returns. This allows it to fit the short-term smile as well as the long-term smile well with only four parameters. For details on VGSSD process see e.g. Carr et al. [2002], O'Sullivan et al. [2010] and for variance gamma process see e.g. Madan et al. [1998].

6.D Robust Monte Carlo of Glasserman and Xu [2014]

We start by generating N independent realizations of X: $(X_1, X_2, ..., X_N)$. The standard Monte Carlo estimator of $\mathbb{E}[V(X)]$ is given by

$$\mathbb{E}[V(X)] \approx \frac{1}{N}\sum_{i=1}^{N} V(X_i).$$

For any fixed and given θ, $(\eta(\theta), \mathbb{E}[m(\theta)V(X)])$ is a straightforward computation according to (6.2.1). The Monte-Carlo estimator $\mathbb{E}[m(\theta)V(X)]$ is simply

$$\mathbb{E}[m(\theta)V(X)] = \frac{\mathbb{E}[V(X)\cdot e^{\theta V(X)}]}{\mathbb{E}[e^{\theta V(X)}]} \approx \frac{\sum_{i=1}^N V(X_i)\cdot e^{\theta V(X_i)}}{\sum_{i=1}^N e^{\theta V(X_i)}}.$$

[11] The variance gamma process has decreasing skewness and kurtosis over time, which makes it look like a normal distribution over long-term. Thus, it cannot fit long-term vol smiles well.

However, since we only know the relative entropy η and not the corresponding level θ, an estimator for the entropy at θ, $\eta(\theta)$, is needed. In essence, θ can be implicitly computed from

$$\hat{\eta}(\theta) \approx \frac{1}{N}\sum_{i=1}^{N}\hat{\eta}_i(\theta) = \frac{1}{N}\sum_{i=1}^{N}\hat{m}_i(\theta)\cdot\ln\hat{m}_i(\theta),$$

where the likelihood ratio $\hat{m}_i(\theta)$ for each X_i, $i = 1, ..., N$ is:

$$\hat{m}_i(\theta) = \frac{e^{\theta V(X)}}{\mathbb{E}[e^{\theta V(X)}]} \approx \frac{e^{\theta V(X_i)}}{\frac{1}{N}\sum_{j=1}^{N}[e^{\theta V(X_j)}]}.$$

Therefore, the procedure is the following: given a relative entropy budget η^*, one computes $\left(\hat{\eta}(\theta), \mathbb{E}[m(\theta)V(X)]\right)$ for multiple θ; Then the upper (or lower) bound of $\mathbb{E}[V(X)]$ is the highest (or lowest) estimate of $\mathbb{E}[m(\theta^*)V(X)]$ for $\eta(\theta^*) \leq \eta^*$.

References

Damien Ackerer and Damir Filipović. Linear credit risk models. *arXiv preprint arXiv:1605.07419*, 2016.

Damien Ackerer and Damir Filipovic. Option pricing with orthogonal polynomial expansions. *arXiv preprint arXiv:1711.09193*, 2017.

Damien Ackerer, Damir Filipović, and Sergio Pulido. The jacobi stochastic volatility model. *Finance and Stochastics*, 22(3):667–700, 2018.

David Ardia, Juan Ospina Arango, and Norman Giraldo Gomez. Jump-diffusion calibration using Differential Evolution. *Wilmott Magazine*, 55:76–79, 2011a. URL http://www.wilmott.com/.

David Ardia, Kris Boudt, Peter Carl, Katharine M. Mullen, and Brian G. Peterson. Differential Evolution with DEoptim: An application to non-convex portfolio optimization. *The R Journal*, 3(1):27–34, 2011b. URL https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Ardia~et~al.pdf.

David Ardia, Katharine M. Mullen, Brian G. Peterson, and Joshua Ulrich. *DEoptim: Differential Evolution in R*, 2016. URL https://CRAN.R-project.org/package=DEoptim. version 2.2-4.

Karl Bannör, Rüdiger Kiesel, Anna Nazarova, and Matthias Scherer. Parametric model risk and power plant valuation. *Energy Economics*, 59:423–434, 2016.

Karl F Bannör and Matthias Scherer. Capturing parameter risk with convex risk measures. *European Actuarial Journal*, 3(1):97–132, 2013.

Karl F Bannör and Matthias Scherer. Model risk and uncertainty–illustrated with examples from mathematical finance. In *Risk-A Multidisciplinary Introduction*, pages 279–306. Springer, 2014.

Martin Barlow, Yuri Gusev, and Manpo Lai. Calibration of multifactor models in electricity markets. *International Journal of Theoretical and Applied Finance*, 7(02):101–120, 2004.

David S Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1):69–107, 1996.

Fred Espen Benth and Paul Krühner. Derivatives pricing in energy markets: an infinite-dimensional approach. *SIAM Journal on Financial Mathematics*, 6 (1):825–869, 2015.

Fred Espen Benth and Thilo Meyer-Brandis. The information premium for non-storable commodities. *Journal of Energy Markets*, 2(3):111–140, 2009.

Fred Espen Benth and Jūratė Šaltytė-Benth. The normal inverse gaussian distribution and spot price modelling in energy markets. *International journal of theoretical and applied finance*, 7(02):177–192, 2004.

Fred Espen Benth and Maren Diane Schmeck. Pricing futures and options in electricity markets. In *The Interrelationship Between Financial and Energy Markets*, pages 233–260. Springer, 2014.

Fred Espen Benth, Jan Kallsen, and Thilo Meyer-Brandis. A non-gaussian ornstein–uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, 14(2):153–169, 2007a.

Fred Espen Benth, Steen Koekebakker, and Fridthjof Ollmar. Extracting and applying smooth forward curves from average-based commodity contracts with seasonal variation. *Journal of Derivatives*, 15(1):52, 2007b.

Fred Espen Benth, Jurate Saltyte Benth, and Steen Koekebakker. *Stochastic modelling of electricity and related markets*, volume 11. World Scientific, 2008a.

Fred Espen Benth, Álvaro Cartea, and Rüdiger 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, 2008b.

Fred Espen Benth, Rüdiger Kiesel, and Anna Nazarova. A critical empirical study of three electricity spot price models. *Energy Economics*, 34(5):1589–1616, 2012.

Fred Espen Benth, Marco Piccirilli, and Tiziano Vargiolu. Mean-reverting additive energy forward curves in a heath–jarrow–morton framework. *Mathematics and Financial Economics*, 13(4):543–577, 2019.

José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.

Francesca Biagini and Yinglin Zhang. Polynomial diffusion models for life insurance liabilities. *Insurance: Mathematics and Economics*, 71:114–129, 2016.

Petter Bjerksund, Heine Rasmussen, and Gunnar Stensland. Valuation and risk management in the norwegian electricity market. In *Energy, natural resources and environmental economics*, pages 167–185. Springer, 2010.

Petter Bjerksund, Gunnar Stensland, and Frank Vagstad. Gas storage valuation: Price modelling v. optimization methods. *The Energy Journal*, 32(1), 2011.

Fischer Black. The pricing of commodity contracts. *Journal of financial economics*, 3(1-2):167–179, 1976.

Alexander Boogert and Cyriel De Jong. Gas storage valuation using a monte carlo method. *The journal of derivatives*, 15(3):81–98, 2008.

Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.

Derek W Bunn and Dipeng Chen. The forward premium in electricity futures. *Journal of Empirical Finance*, 23:173–186, 2013.

Markus Burger, Bernhard Klar, Alfred Müller, and Gero Schindlmayr. A spot market model for pricing derivatives in electricity markets. *Quantitative finance*, 4(1):109–122, 2004.

René Carmona and Michael Coulon. A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*, pages 41–83. Springer, 2014.

René Carmona and Valdo Durrleman. Pricing and hedging spread options. *Siam Review*, 45(4):627–685, 2003.

René Carmona and Michael Ludkovski. Valuation of energy storage: An optimal switching approach. *Quantitative finance*, 10(4):359–374, 2010.

Peter Carr, Hélyette Geman, Dilip B Madan, and Marc Yor. The fine structure of asset returns: An empirical investigation. *The Journal of Business*, 75(2): 305–332, 2002.

Zhuliang Chen and Peter A Forsyth. A semi-lagrangian approach for natural gas storage valuation and optimal operation. *SIAM Journal on Scientific Computing*, 30(1):339–368, 2008.

Panagiotis Christodoulou, Nils Detering, and Thilo Meyer-Brandis. Local risk-minimization with multiple assets under illiquidity with applications in energy markets. *International Journal of Theoretical and Applied Finance*, 2018.

Rama Cont. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical finance*, 16(3):519–547, 2006.

Christa Cuchiero. Polynomial processes in stochastic portfolio theory. *Stochastic Processes and their Applications*, 2018.

Christa Cuchiero, Damir Filipović, Eberhard Mayerhofer, Josef Teichmann, et al. Affine processes on positive semidefinite matrices. *The Annals of Applied Probability*, 21(2):397–463, 2011.

Christa Cuchiero, Martin Keller-Ressel, and Josef Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16(4):711–740, 2012.

Mark Cummins, Greg Kiely, and Bernard Murphy. Gas storage valuation under lévy processes using the fast fourier transform. *Journal of Energy Markets*, 10(4):43–86, 2017.

Cyriel De Jong. Gas storage valuation and optimization. *Journal of Natural Gas Science and Engineering*, 24:365–378, 2015.

Freddy Delbaen and Hiroshi Shirakawa. An interest rate model with upper and lower bounds. *Asia-Pacific Financial Markets*, 9(3-4):191–209, 2002.

Emanuel Derman. Model risk. *RISK-LONDON-RISK MAGAZINE LIMITED-*, 9:34–38, 1996.

Joost Driessen, Pieter Klaassen, and Bertrand Melenberg. The performance of multi-factor term structure models for pricing and hedging caps and swaptions. *Journal of Financial and Quantitative Analysis*, 38(3):635–672, 2003.

Philip H Dybvig. *Bond and bond option pricing based on the current term structure*. Olin School of Business, University of Washington, 1988.

Daniel Ellsberg. Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669, 1961.

Alvaro Escribano, J Ignacio Pena, and Pablo Villaplana. Modelling electricity prices: International evidence. *Oxford bulletin of economics and statistics*, 73 (5):622–650, 2011.

Yu Feng and Erik Schlögl. Model risk measurement under wasserstein distance. *arXiv preprint arXiv:1809.03641*, 2018.

Yu Feng, Ralph Rudd, Christopher Baker, Qaphela Mashalaba, Melusi Mavuso, and Erik Schlögl. Quantifying the model risk inherent in the calibration and recalibration of option pricing models. *Available at SSRN 3267775*, 2018.

Damir Filipovic. *Term-Structure Models. A Graduate Course.* Springer, 2009.

Damir Filipović and Martin Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, Oct 2016. ISSN 1432-1122. doi: 10.1007/s00780-016-0304-4.

Damir Filipović and Martin Larsson. Polynomial jump-diffusion models. *Swiss Finance Institute Research Paper*, (17-60), 2019.

Damir Filipović and Martin Larsson. Polynomial jump-diffusion models. *Stochastic Systems*, 10(1):71–97, 2020.

Damir Filipović and Sander Willems. A term structure model for dividends and interest rates. 2018.

Damir Filipović, Elise Gourier, and Loriano Mancini. Quadratic variance swap models. *Journal of Financial Economics*, 119(1):44–68, 2016.

Damir Filipović, Martin Larsson, and Anders B Trolle. Linear-rational term structure models. *The Journal of Finance*, 72(2):655–704, 2017.

Stefano Fiorenzani, Samuele Ravelli, and Enrico Edoli. *The handbook of Energy trading.* John Wiley & Sons, 2012.

H Föllmer and Martin Schweizer. Hedging of contingent claims. *Applied stochastic analysis*, 5:389, 1991.

Hans Föllmer and Alexander Schied. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002.

Hans Föllmer and Alexander Schied. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter, 2011.

Hans Follmer and Dieter Sondermann. Contributions to mathematical economics. *North Holland*, 1986.

Dennis Frestad. Common and unique factors influencing daily swap returns in the nordic electricity market, 1997–2005. *Energy Economics*, 30(3):1081–1097, 2008.

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

Fabio Genoese, Massimo Genoese, and Martin Wietschel. Occurrence of negative prices on the german spot market for electricity and their influence on balancing power markets. In *2010 7th International Conference on the European Energy Market*, pages 1–6. IEEE, 2010.

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

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with a non-unique prior. 1989.

Paul Glasserman. Shortfall risk in long-term hedging with short-term futures contracts. *Option Pricing, Interest Rates and Risk management*, pages 477–508, 2001.

Paul Glasserman and Xingbo Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014.

T Clifton Green and Stephen Figlewski. Market risk and model risk for a financial institution writing options. *The Journal of Finance*, 54(4):1465–1499, 1999.

Rangga Handika, Stefan Trück, et al. The relationship between spot and futures prices: An empirical analysis of australian electricity markets. In *3rd IAEE Asian Conference*, 2012.

David Heath, Eckhard Platen, and Martin Schweizer. *Numerical comparison of local risk-minimisation and mean-variance hedging*. Australian National University, Centre for Mathematics and its Applications, School of Mathematical Sciences, 1999.

David Heath, Eckhard Platen, and Martin Schweizer. A comparison of two quadratic approaches to hedging in incomplete markets. *Mathematical Finance*, 11(4):385–413, 2001.

Patrick Henaff, Ismail Laachir, and Francesco Russo. Gas storage valuation and hedging. a quantification of the model risk. *arXiv preprint arXiv:1312.3789*, 2013.

Patrick Hénaff, Ismail Laachir, and Francesco Russo. Gas storage valuation and hedging: A quantification of model risk. *International Journal of Financial Studies*, 6(1):27, 2018.

Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

Jens Hilscher, Robert A Jarrow, and Donald R van Deventer. The valuation of corporate coupon bonds. *Available at SSRN 3277092*, 2020.

Jennifer A Hoeting, David Madigan, Adrian E Raftery, and Chris T Volinsky. Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401, 1999.

Alan Holland. Optimization of injection/withdrawal schedules for natural gas storage facilities. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 287–300. Springer, 2007.

Alan Holland. A decision support tool for energy storage optimization. In *2008 20th IEEE International Conference on Tools with Artificial Intelligence*, volume 2, pages 299–306. IEEE, 2008.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

John Hull and Wulin Suo. A methodology for assessing model risk and its application to the implied volatility function model. *Journal of Financial and Quantitative Analysis*, pages 297–318, 2002.

Eric Jacquier and Robert Jarrow. Bayesian analysis of contingent claim model error. *Journal of Econometrics*, 94(1-2):145–180, 2000.

Patrick Jaillet, Ehud I Ronn, and Stathis Tompaidis. Valuation of commodity-based swing options. *Management science*, 50(7):909–921, 2004.

Jan Kallsen and Johannes Muhle-Karbe. Exponentially affine martingales, affine measure changes and exponential moments of affine processes. *Stochastic Processes and their Applications*, 120(2):163–181, 2010.

Ioannis Karatzas and Steven E Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1998.

Nicole El Karoui, Monique Jeanblanc-Picquè, and Steven E Shreve. Robustness of the black and scholes formula. *Mathematical finance*, 8(2):93–126, 1998.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

Rüdiger Kiesel, Gero Schindlmayr, and Reik H Börger. A two-factor model for the electricity forward market. *Quantitative Finance*, 9(3):279–287, 2009.

Xi Kleisinger-Yu, Vlatka Komaric, Martin Larsson, and Markus Regez. A multifactor polynomial framework for long-term electricity forwards with delivery period. *SIAM Journal on Financial Mathematics*, 11(3):928–957, 2020.

Tino Kluge. Pricing swing options and other electricity derivatives. 2006.

Frank Hyneman Knight. *Risk, uncertainty and profit*, volume 31. Houghton Mifflin, 1921.

Christopher R Knittel and Michael R Roberts. An empirical examination of deregulated electricity prices. *POWER Working Paper No. PWP-087*, 2001.

Steen Koekebakker and Fridthjof Ollmar. Forward curve dynamics in the nordic electricity market. *Managerial Finance*, 31(6):73–94, 2005.

Nikola Krečar, Fred E Benth, and Andrej F Gubina. Towards definition of the risk premium function. *IEEE Transactions on Power Systems*, 2019.

Robert Litterman and Jose Scheinkman. Common factors affecting bond returns. *Journal of fixed income*, 1(1):54–61, 1991.

Francis A Longstaff, Pedro Santa-Clara, and Eduardo S Schwartz. Throwing away a billion dollars: The cost of suboptimal exercise strategies in the swaptions market. *Journal of Financial Economics*, 62(1):39–66, 2001.

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

Dilip B Madan, Peter P Carr, and Eric C Chang. The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105, 1998.

Alexander M Malyscheff and Theodore B Trafalis. Natural gas storage valuation via least squares monte carlo and support vector regression. *Energy Systems*, 8(4):815–855, 2017.

Angelo Melino and Stuart M Turnbull. Misspecification and the pricing and hedging of long-term foreign currency options. *Journal of International Money and Finance*, 14(3):373–393, 1995.

Robert C Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976.

Alain Monfort, Jean-Paul Renne, and Guillaume Roussellet. A quadratic kalman filter. *Journal of Econometrics*, 187(1):43–56, 2015.

Katharine Mullen, David Ardia, David Gil, Donald Windover, and James Cline. DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 2011. URL http://www.jstatsoft.org/v40/i06/.

Anthony Neuberger. Hedging long-term exposures with multiple short-term futures contracts. *The Review of Financial Studies*, 12(3):429–459, 1999.

Conall O'Sullivan, Michael Moloney, et al. *The variance gamma scaled self-decomposable process in actuarial modelling.* UCD Centre for Economic Research, 2010.

KB Petersen and MS Pedersen. The matrix cookbook. november 2008, 2008.

Dragana Pilipovic. *Energy risk: Valuing and managing energy derivatives.* McGraw Hill Professional, 2007.

Kenneth V. Price, Rainer M. Storn, and Jouni A. Lampinen. *Differential Evolution - A Practical Approach to Global Optimization.* Natural Computing. Springer-Verlag, January 2006. ISBN 540209506.

Adrian E Raftery, David Madigan, and Jennifer A Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997.

William Thomas Reid. *Riccati differential equations*. Elsevier, 1972.

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

Nemat Safarov and Colin Atkinson. Natural gas storage valuation and optimization under time-inhomogeneous exponential lévy processes. *International Journal of Computer Mathematics*, 94(11):2147–2165, 2017.

Till Schröter, Michael Monoyios, Mario Rometsch, and Karsten Urban. Model uncertainty and the robustness of hedging models, 2012.

Eduardo Schwartz and James E Smith. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911, 2000.

Eduardo S Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of finance*, 52(3):923–973, 1997.

Martin Schweizer. Risk-minimality and orthogonality of martingales. *Stochastics: An International Journal of Probability and Stochastic Processes*, 30(2):123–131, 1990.

Martin Schweizer. A guided tour through quadratic hedging approaches. Technical report, Discussion Papers, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 1999.

James M Steeley. Modelling the dynamics of the term structure of interest rates. 1990.

Rainer Storn and Kenneth Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, Dec 1997. ISSN 1573-2916. doi: 10.1023/A:1008202821328. URL https://doi.org/10.1023/A:1008202821328.

Matt Thompson, Matt Davison, and Henning Rasmussen. Natural gas storage valuation and optimization: A real options application. *Naval Research Logistics (NRL)*, 56(3):226–238, 2009.

Raman Uppal and Tan Wang. Model misspecification and underdiversification. *The Journal of Finance*, 58(6):2465–2486, 2003.

Niyaz Valitov. Risk premia in the german day-ahead electricity market revisited: The impact of negative prices. *Energy Economics*, 82:70–77, 2019.

Johannes Viehmann. Risk premiums in the german day-ahead electricity market. *Energy policy*, 39(1):386–394, 2011.

Pablo Villaplana. Pricing power derivatives: A two-factor jump-diffusion approach. In *EFMA 2004 Basel Meetings Paper*, 2003.

Peter Volkmann. Über die invarianz konvexer mengen und differentialungleichungen in einem normierten raume. *Mathematische Annalen*, 203(3):201–210, 1973.

Tony Ware. Polynomial processes for power prices. *Applied Mathematical Finance*, 26(5):453–474, 2019.

Rafał Weron. Market price of risk implied by asian-style electricity options and futures. *Energy Economics*, 30(3):1098–1115, 2008.