Technische Universität München
Lehrstuhl für Finanzmathematik

Incorporating parameter risk into derivatives prices – bid-ask pricing and calibration

Karl Friedrich Bannör

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Univ.-Prof. Claudia Czado, Ph.D.
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Matthias Scherer
	2.	Univ.-Prof. Dr. Ralf Korn, Technische Universität Kaiserslautern
	3.	Lecturer Dr. Christoph Reisinger, University of Oxford/Vereinigtes Königreich

Die Dissertation wurde am 16. Mai 2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 26. Juli 2013 angenommen.

To my beloved, all-time supporting wife Bianka,
who made this work possible.

Abstract

In this thesis, a new methodology using convex risk measures is developed to incorporate parameter risk into prices of financial derivatives, provided that a distribution on the parameter space is given. In this context, weak continuity properties of convex risk measures w.r.t. the underlying probability measure on the parameter space are analyzed. Parameter risk arising from time series estimation is discussed in extensive numerical case studies and large-sample approximations for certain parameter risk-captured prices are stated. A technique to induce a parameter distribution in case of calibration to market prices is presented, allowing to conduct a comparison of parameter risk in different financial market models and of different exotic options. For the calibration to quoted bid-ask prices, a non-parametric calibration approach to broad classes of distortion risk measures is developed and a calibration to quoted bid-ask prices – comparing the non-parametric approach to given parametric suggestions – is assessed.

Zusammenfassung

In dieser Doktorarbeit wird eine neue Methodik basierend auf konvexen Risikomaßen entwickelt, um Parameterrisiko in den Preisen von Finanzderivaten zu berücksichtigen. In diesem Kontext werden schwache Stetigkeitseigenschaften von konvexen Risikomaßen als Funktion des unterliegenden Wahrscheinlichkeitsmaßes auf dem Parameterraum untersucht. Es wird Parameterrisiko resultierend aus Zeitreihenschätzung in numerischen Fallstudien untersucht und Approximationen für parameterrisikoadjustierte Preise im Falle von großen Stichproben angegeben. Im Falle der Parameterschätzung durch Kalibrierung an Marktpreise wird eine Technik entwickelt, ein Wahrscheinlichkeitsmaß auf dem Parameterraum zu konstruieren, um einen Vergleich des Parameterrisikos in verschiedenen Finanzmarktmodellen sowie verschiedenen exotischen Optionen durchzuführen. Zur Kalibrierung an marktquotierte Bid-Ask-Preise wird ein nichtparametrischer Ansatz zur Kalibrierung an Verzerrungsrisikomaße entwickelt und eine Kalibrierung an marktquotierte Bid-Ask-Preise zum Vergleich mit existierenden parametrischen Methoden durchgeführt.

Acknowledgements

At first, I want to thank my supervisor Prof. Dr. Matthias Scherer for his continuous support during all the time. He always had an open door for questions and fruitful discussions about the thesis, came up with useful ideas, and helped me with his kind and patient nature to make my step into the research community. I think that this thesis would not have been possible without him.

I also want to thank Prof. Dr. Rudi Zagst for providing a possibility to finance my PhD studies with work at the Chair of Mathematical Finance as well as giving me the opportunity to present my research to a broader audience at various international conferences. Furthermore, I want to thank Prof. Dr. Rüdiger Kiesel for inviting me to Essen for a joint research project, which allowed me to gain further insight into electricity markets.

During the time of my PhD studies, I could always stock up my creativity and exchange ideas with my appreciated colleagues, among them I found wonderful friends. So, I thank my roommates Daniela Selch and Thorsten Schulz for bearing me as their officemates, Lexuri Fernández and Peter Hieber for nice distraction during creative breaks, German Bernhart, Maximilian Gass, Dr. Asma Khedher, Mirco Mahlstedt, Franz Ramsauer, and Steffen Schenk for great table football matches, PD Dr. Aleksey Min for telling me great stories, and Tim Friederich, Prof. Dr. Kathrin Glau, Julia-Stefanie Kraus, Mikhail Krayzler, Maximilian Mair, Daniela Neykova, and Natalia Shenkman for a good time together at the Chair of Mathematical Finance.

Since they have provided an excellent environment for building the foundation of my education during my childhood, I want to thank my parents Monika and Karl Heinz Hofmann. Last, but not least, I want to thank my dearly-loved wife Bianka Bannör: Due to all her support in my private life and her sacrifice, I was able to write this thesis.

Contents

1 Introduction **9**

1.1 The principles of mathematical modeling in finance 9

1.2 Model and parameter risk and uncertainty 11

1.3 Literature overview 15

1.4 Contributions 18

2 Mathematical preliminaries **23**

2.1 Theory of convex risk measures 23

2.2 Choquet integration theory 29

3 Model and parameter risk – a convex risk measure ansatz **37**

3.1 Model and parameter uncertainty in derivatives pricing 37

3.2 Capturing model and parameter risk by using convex risk measures 43

3.3 Examples: AVaR- and entropic-driven risk-captured prices 48

3.4 Market-implied distributions 52

4 Convergence properties of risk-captured prices **55**

4.1 Convergence results of risk-capturing functionals 55

4.2 Counterexamples 60

4.3 Outlook 60

5 Application: Bid-ask prices implied by estimation risk **63**

5.1 Estimation risk-captured prices for consistent estimators 63

5.2 Asymptotics of risk-capturing functionals 65

5.3 Case study: Estimation risk for Margrabe options in a two-dimensional Black-Scholes market 68

5.4 The parameter risk-captured valuation of a gas power plant 74

6 Application: Parameter risk induced by calibration to market prices **89**

6.1 Calibration to market prices 91

6.2 Parameter risk from calibration to market prices 92
6.3 Case study: Comparing parameter risk of different models and exotic options 99
6.3.1 Parameter risk in different models 104
6.3.2 Parameter risk profiles of different exotics 112
6.4 Outlook 113

7 Calibration of risk-captured model prices to bid-ask market prices 119
7.1 Introduction 119
7.2 The bid-ask calibration problem 120
7.3 A non-parametric calibration scheme for bid-ask price 125
7.3.1 General results 125
7.3.2 Application to parameter risk-captured prices 134
7.4 Application to data 134
7.5 Calibration to discontinuous distortion functions 137
7.6 Outlook 144

8 Conclusion 149
8.1 Summary 149
8.2 Critical reflection 151
8.3 Outlook 152

Bibliography 155

1 Introduction

In this introductory chapter, we first give a short insight into mathematical modeling in finance and explain the cruciality of model and parameter risk and uncertainty in mathematical finance. We provide a literature overview on existing works on model and parameter risk and uncertainty in mathematical finance and briefly summarize the main contributions of this thesis.

1.1 The principles of mathematical modeling in finance

Today, mathematical modeling plays an important role in many different areas like, e.g., geoscience, engineering, empirical social science, and, last but not least, finance. In physics and engineering, mathematical modeling of real-world phenomena goes back to, e.g., Sir Isaac Newton and even to the ancient Greeks. Contrasting, in finance, mathematical and particularly stochastic modeling is a rather recent trend, starting around 1900 with Louis Bachelier's seminal PhD thesis (cf. (Bachelier, 1900)).

When regarding the financial world instead of modeling phenomena from classical mechanics (like, e.g., in engineering), one immediately recognizes that the whole system is much more complex in the sense that many different forces drive the market in many different ways. When describing the fall of a stone to the ground in a laboratory, there are undoubtedly many different forces apart from earth gravitation that actually have an influence (e.g. the aerodynamic resistance, the gravitation of different objects in the laboratory). But their magnitude is so small compared to the magnitude of earth gravitation that not considering them eventually does not matter too much for the model.

Contrasting, when modeling financial markets (e.g. stock markets for the purpose of derivatives pricing), there are many different market participants that actually influence asset prices in many different ways. Hence, a model trying to capture the whole market microstructure with all action and interaction of market participants would be extremely complicated, modeling many different dimensions with myriads of parameter. Thus,

such an approach is challenging from a modeling and computational perspective. But, contrastingly, there are several other good reasons not to model the microstructure of financial markets. First, financial markets cannot be put under laboratory conditions and therefore models cannot be tested reliably. Second, it is impossible to observe all market participants' behavior and interaction simultaneously. Third, many market participants exhibit irrational and erratic behavior which may be difficult to model even when modeling only a single market participant.[1] Finally and maybe most crucial, the whole system is dynamic, with new market participants entering and leaving the system in a continuous manner. Even if one could observe the market participants's behavior and collect huge data to exploit, in every second, new market participants enter the financial markets and behave differently, such that the predictions coming from a possible collected data set are outdated and do not match the new market environment.[2] Hence, the typical approach to model stock markets is to forget about the market microstructure and to model asset prices directly in a stochastic manner.

Since the advent of Louis Bachelier's thesis (Bachelier, 1900) and the seminal papers (Samuelson, 1965; Black and Scholes, 1973), the predominant principle to model the prices of financial assets has been the following: The events on the financial markets and the time-increasing information is modeled by a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, the filtration $(\mathcal{F}_t)_{t\geq 0}$ corresponding to the flow of information. Then, the asset prices are modeled as a stochastic process $S = (S_t)_{t\geq 0}$ which is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. To obtain a useful asset price model, one should demand that the stochastic process S fulfills stylized facts such as:

- The stock price process, abbreviated by $S = (S_t)_{t\geq 0}$, is always positive;
- Returns (yields) of stock prices are scattered around 0 (or around somewhere close to 0) and behave roughly similar and independent from each other.

Furthermore, the availability of (semi-)closed-form pricing formulae is crucial for quick and efficient evaluation of common products.

The most famous model, the Black–Scholes model (cf. (Black and Scholes, 1973), awarded with the Nobel Memorial Prize in Economic Sciences in 1997), models S by a geometric Brownian motion with drift parameter $\mu \in \mathbb{R}$ and volatility $\sigma > 0$. For derivative

[1] Some groundbreaking work has, e.g., been done by (Kahneman and Tversky, 1979), where "irrational" behavior is partly explained by the *Prospect Theory*. Daniel Kahneman was awarded the Nobel Memorial Prize in Economic Sciences in 2002 for his work.

[2] In financial markets, one can even argue that relying too much on collected data may result in overconfidence, since the data may not be representative any more to model future use.

pricing purposes, the drift term μ is already determined by the risk-free interest rate. Furthermore, many other models, capturing more stylized facts of financial market time series such as volatility clustering, the so-called "leverage effect" (cf. (Christie, 1982)), or jumps, have been developed to overcome known shortcomings of the Black–Scholes model. Examples are, e.g., the jump-diffusion models of (Merton, 1976; Kou, 2002) incorporating additional jumps in the stock price[3], stochastic volatility models substituting the constant volatility σ by a stochastic process $(\sigma_t)_{t\geq 0}$ (like the Stein and Stein model (Stein and Stein, 1991), the Heston model described in (Heston, 1993), or the jump-diffusion style model of Barndorff-Nielsen and Shephard combining both approaches (cf. (Barndorff-Nielsen and Shephard, 2001; Nicolato and Venardos, 2003)), and models driven by processes other than Brownian motions (e.g. Lévy models like the Variance Gamma model of (Madan et al., 1998), which are extensively treated in the textbook (Cont and Tankov, 2004), Sato models as in (Carr et al., 2006)). Hence, since the Black–Scholes model falls behind to model stock prices accurately due to the vast generalization (returns are normally distributed), a whole zoo of models competing for usage has been developed and it is difficult to decide which model to favor. There is vivid research to provide new models for different situations, but, until now, a "gold standard" model for derivatives pricing has not yet emerged.

1.2 Model and parameter risk and uncertainty

When setting up a stochastic model, one typically observes a complicated situation where the outcome to model behaves in a more or less erratic manner (which is the reason for employing stochasticity). In some cases, a simple and accurate description of the behavior may be provided easily, as, e.g., the throw of a fair dice, where the relative frequency of different results behaves like the (discrete) uniform distribution on $\{1,\ldots,6\}$. But, typically, the objects (e.g. stock prices, interest rates, or FX rates) to capture in finance behave much more complicated. Hence, it is not clear from the beginning that the choice of one stochastic model $(\Omega,\mathcal{F},P)$ is a good choice or whether a different model $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{P})$ might be more suitable to model the stock price by some stochastic process $(S_t)_{t\geq 0}$. Since the asset price process $(S_t)_{t\geq 0}$ is typically used without directly specifying the stochastic basis, the specification of the filtered measurable space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0})$ is often done implicitly. So, choosing a model mostly means specifying

[3]The model of (Merton, 1976) employs normally distributed jumps, while the model described in (Kou, 2002) uses double exponentially distributed jumps, also allowing for asymmetrically distributed jumps.

a probability measure P on an implicitly given filtered stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$, since the dynamics of the stochastic process $(S_t)_{t\geq 0}$ are specified by the probability measure P. Hence, the environment can be mathematically described as a situation where a whole set of probability measures $\mathcal{P}$ (which may typically be infinite) is available for modeling, i.e. the dynamics of $(S_t)_{t\geq 0}$ follow different trajectories or the probabilities of trajectories shift. Sometimes, the set of possible probability measures (i.e. different stochastic models) $\mathcal{P}$ may be parameterized in a canonical way by a parameter space Θ, i.e. $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

To provide a concise wording to different situations that may occur when different models $\mathcal{P}$ are available, we first provide a short excursion into the literature. In the seminal dissertation (Knight, 1921), the situation where different objects $x_1, \ldots, x_N$ are possible outcomes was analyzed. (Knight, 1921) distinguishes between two possible situations that may occur:

1. One does not know any quantification about which outcome might occur
2. One knows the probability of each possible outcome $x_1, \ldots, x_N$

The former situation, where basically no information is available, is called *uncertainty* by (Knight, 1921), while the latter one, which at least provides a probabilistic description, is called *risk*. Obviously, facing risk is a special case of uncertainty (one could always forget about the probabilities) and a more comfortable situation compared to facing real uncertainty. One can try to deal with a risky situation by *risk management*, i.e. exploiting the information about the probabilities of the different outcomes $x_1, \ldots, x_N$ and acting such that a certain risk functional may be minimized.

Transferring the concepts of risk and uncertainty to stochastic modeling, the situation of having a whole set of models $\mathcal{P}$ to choose from for modeling is generally referred to as *model uncertainty*. In case of an available and feasible parameterization of $\mathcal{P}$ by some parameter space Θ, one speaks about *parameter uncertainty*. From a mathematical point of view, model and parameter uncertainty are equivalent notions, since one may parameterize $\mathcal{P}$ by some suitable parameter set Θ and a bijection $\Theta \to \mathcal{P}$. But, practically, the set Θ can often be chosen such that treating different parameters $\theta \in \Theta$ allows for more convenient interpretation in the real world than treating the corresponding model P_θ. If we additionally have a probability measure R given on the set of possible models $\mathcal{P}$ (resp. on the set of possible parameters Θ) quantifying the likelihood of each model (resp. parameter), then we are in a setting of *model risk* (resp. *parameter risk*), which (trivially) is a special case of model (resp. parameter) uncertainty.

As we have mentioned above, during the last years, there has been developed a large battery of different models (Lévy models, jump-diffusion models, stochastic volatility models and many others) that may be used for derivatives pricing purposes. But, until now, none of these models has turned out to become an accepted reference capturing asset price movements completely. Conversely, the whole setting in finance is completely dynamic: Market participants learn from historical mistakes, behave differently in new market situations, and suddenly enter/exit the market. Hence, models describing the dynamics of today's financial markets quite decently may not be valid tomorrow due to the completely changed setting. Furthermore, different models that are fitted to data (typically via calibration to prices of liquid derivatives, e.g. European options) deliver tremendously different results when pricing exotic derivatives (as, e.g., barrier options) with them, as pointed out in (Schoutens et al., 2004). Hence, financial market modeling (e.g. for the valuation of derivatives) is a situation where model uncertainty is prominent. This was outlined, e.g., in (Figlewski, 1998; Green and Figlewski, 1999), and insistently stressed in the seminal paper (Cont, 2006).

Another source of uncertainty, namely parameter uncertainty, plays an equally prominent role for pricing derivatives. Typically, in practice, a parametric model (e.g. Black–Scholes model, Heston model, Variance Gamma model etc.) is assumed to hold for the asset price process. But the parameters of the model θ (like, e.g., the Black–Scholes volatility, the numerous parameters for the Heston model, etc.) have to be specified, coming from a (possibly rich) parameter space Θ. Some of the parameters (e.g. today's asset price, risk-free interest rates) may be available in liquid markets, but the crucial ones for derivatives pricing are typically not.[4] Hence, the model-specific parameters have to be specified. Typically, there are two methods how to specify the parameters of the model:

1. If a trustworthy time series of historical asset prices, say $(s_1, \ldots, s_N)$, exists and there is statistical theory (estimation theory) available, one might try to estimate an unknown parameter θ by employing an estimator $\hat{\theta} = \hat{\theta}^{(N)} = \hat{\theta}(s_1, \ldots, s_N)$.[5]

2. If the asset $(S_t)_{t\geq 0}$ is fairly liquidly traded (as, e.g., the major stock indices like S&P 500, DAX, EURO STOXX 50, or important FX rates as EUR-USD), prices

[4] One might argue that in some mature markets, implied volatilities are readily available due to a vivid market for European options. In these cases, one typically has ambiguity between different implied volatilities calculated from different maturities and/or moneyness (volatility smile). Furthermore, as mentioned above, the presence of different implied volatilities proves that the Black–Scholes model is a questionable model.

[5] Since change of measure may affect some parameters, one has to be careful with this method for derivatives pricing.

of vanilla products (often European options) $(C_1^\star, \ldots, C_N^\star)$ are available (often in indirect quotation via Black–Scholes implied volatilities). In this situation, one might try to fit the model to mimic the market prices of these vanilla products ("calibration to market prices") and to estimate the parameter by minimizing some "aggregate error to market prices".[6]

Actually, with both procedures to obtain the parameter θ, one implicitly runs into problems, which are often not addressed by neither practictioners nor academia:

- When using historical estimation, the historical measure P induces a distribution on the parameter space Θ via the pushforward measure of the estimator $\hat{\theta}_N$, often briefly summarized by the estimator's variance and (possible) bias. When just using the point estimate (which results from plugging the time series data $(s_1, \ldots, s_N)$ into the estimation function), the distribution of the estimator $\hat{\theta}$ is completely disregarded. Hence, one runs the risk that the historical estimation procedure yields the wrong parameter.[7]
- When employing calibration to market prices, there may be the possibility that the employed optimization algorithm runs in a local minimum or that the global minimum is not unique. Since the input prices may be distorted, e.g. due to delay of quotes, one can also not be sure that the parameters which fit almost equally well might not be a good alternative. Furthermore, due to the aggregate error to market prices (which is the goal function in the optimization procedure), one has some information about the "trustworsthyness" of the parameters. In many parametric models, it is the case that the parameters play a partly interchangeable role in determining the prices of vanilla options. Hence, different parameter vectors may have similar errors to market prices.

Hence, parameter uncertainty (resp. parameter risk – when a distribution of the parameters is at hand) is an important issue when modeling the price of financial assets and, particularly, for the pricing of financial derivatives. Moreover, since models tend to become more and more complicated to capture more stylized facts (e.g. compare a Bates model to a Heston model), calibration or estimation of these models become more sophisticated and is often accompanied with more parameter risk/uncertainty.

[6] In practice, the second alternative "calibration" is typically the preferred variant for derivatives pricing since the vanilla options that are employed for calibration are *forward-looking*, i.e. implicitly express a "market consensus opinion" about future asset prices. Contrasting, historical estimation employing time series of the asset price process reflects *historic* snapshots of the asset price.

[7] Actually, in many continuous settings, the probability that the point estimate is the true parameter is actually zero!

Immediately after recognizing the vast problems that are connected to model (resp. parameter) risk and uncertainty, the question how to deal with this risk/uncertainty immediately arises.

1.3 Literature overview

The topic of model and parameter risk and uncertainty in finance has been addressed in numerous sources. Since the financial crisis of 2008, where the misspecification of models played a prominent role for the valuation of portfolio credit derivatives, the problem of model and parameter risk and uncertainty in a financial context is addressed more and more. In some sense, the presence of the smile effect, i.e. the observation that implied Black–Scholes volatilities for European options with different moneyness and/or maturity are not constant, addresses parameter uncertainty in the special case when using the Black–Scholes model. Furthermore, for risk-neutral valuation, the notion of incomplete markets (which is intensely discussed in the discrete and continuous case in the textbook of (Černy, 2009)) treats a special kind of model (resp. parameter) uncertainty, where all probability measures in doubt are equivalent to each other. A survey from an economic perspective about model risk/uncertainty in derivatives pricing is given in (Figlewski, 1998) where numerous sources of model/parameter risk and uncertainty are discussed: The smile effect is addressed as well as problems connected to other models incorporating returns with fatter tails, possible jumps, and stochastic volatility. parameter risk/uncertainty is described, in particular historical volatility estimating as well as the specification of a GARCH-type stochastic volatility model. A thorough empirical study about the risk in derivatives pricing and hedging which is associated to forecasting Black–Scholes volatility has been done in (Green and Figlewski, 1999).

In an incomplete markets setting (which is, as we have described above, a special kind of model uncertainty), more pricing (and hedging) approaches have been discussed and the original sub-/superhedging pricing has been enriched. From the hedging persperctive, weaker variants of hedging are also tackled by, e.g., mean-variance hedging (cf., e.g., (Föllmer and Schweizer, 1990; Schweizer, 1991)), quantile hedging (cf. (Föllmer and Leukert, 1999)), and efficient hedging (cf. (Föllmer and Leukert, 2000)). (Carr et al., 2001) discuss that complete sub-/superhedging even in incomplete markets may lead to non-competitive prices for derivatives traders, since derivatives traders are paid for taking some risk, which may be the risk of having selected the wrong equivalent martingale

measure. Thus, a "reduced sub-/superhedging price" is suggested. More detailed, denoting by $\mathcal{Q} := \{Q \sim P : Q \text{ martingale measure of the discounted stock price}\}$ the set of all equivalent martingale measures, one may select a subset of measures $\mathcal{Q}_0 \subset \mathcal{Q}$ and employ "reduced worst-case pricing" by setting price borders $m(X) = \inf_{Q \in \mathcal{Q}_0} \mathbb{E}_Q[X]$ and $M(X) = \sup_{Q \in \mathcal{Q}_0} \mathbb{E}_Q[X]$ for some derivative X. From a mathematical point of view, this ansatz brings coherent risk measures into play, which was generalized by (Xu, 2006) with pricing based by general convex risk measures in incomplete markets and developing hedging strategies as well as diversification issues, which a derivatives trader typically uses for her or his purposes. (Cherny and Madan, 2010) also employ convex risk measure pricing in incomplete markets by using Choquet integrals w.r.t. concave distortions of the historical measure P, which is a special class of convex risk measures. Furthermore, the convenient representation is used to calibrate to bid-ask prices (instead of mid prices as in a usual calibration setting), where parametric families developed in (Cherny and Madan, 2009) are used to determine the risk measure.

Also, from a methodological point of view, there have been some treatments and suggestions in the literature how to deal with model and parameter risk and uncertainty in derivatives pricing. A mathematical approach treating a geometric Brownian motion-style diffusion process $(S_t)_{t \geq 0}$ with stochastic, but uncertain volatility process $(\sigma_t)_{t \geq 0}$ is done in (Avellaneda et al., 1995): By means of control theory methods, a PDE-style pricing principle is developed for worst-case pricing when the volatility process $(\sigma_t)_{t \geq 0}$ is a bounded stochastic process, which is assumed to stay in some compact interval, i.e. $\sigma_t \in [\sigma_{\min}, \sigma_{\max}]$ for all $t \geq 0$. This approach is usually refered to as the *uncertain volatility model*. (Avellaneda et al., 1995) develop a Hamilton–Jacobi–Bellman-type solution to price derivatives in the uncertain volatility model, i.e. pricing is based on numerical solutions of the so-called Black–Scholes–Barenblatt PDE.

A much more general approach is described in the seminal paper (Cont, 2006), where the general pricing idea resulting from sub-/superhedges in incomplete markets is generalized. In case of the availability of a set of possible risk-neutral pricing models $\mathcal{Q}$ consisting of martingale measures w.r.t. the asset price process $(S_t)_{t \geq 0}$, i.e. $(S_t)_{t \geq 0}$ is a martingale[8] under each $Q \in \mathcal{Q}$, (Cont, 2006) advocats to do worst-case pricing, similarly to the methodology which is used for sub-/superhedging in incomplete markets: The bid and ask prices which account for model uncertainty are calculated as $m(X) = \inf_{Q \in \mathcal{Q}} \mathbb{E}_Q[X]$ resp. $M(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[X]$, leading to non-linear pricing such that the functional M is a coherent risk measure. Moreover, if the risk-neutral measures

[8] Up to the proper numéraire, which is omitted here for simplicity.

$\mathcal{Q}$ are given by calibration and the aggregated market pricing error is given by $\eta(Q)$ for some $Q \in \mathcal{Q}$, (Cont, 2006) suggests a penalized worst-case approach (which is familiar from convex optimization): The derivative's price should be calculated using the worst-case approach, penalizing every derivative price with the aggregated market error $\eta(Q)$, i.e. $M(X) = \sup_{Q\in\mathcal{Q}} \mathbb{E}_Q[X] - \eta(Q)$ resp. $m(X) = \inf_{Q\in\mathcal{Q}} \mathbb{E}_Q[X] - \eta(Q)$. From a critical point of view, the penalized approach has some methodological weaknesses: First, every derivatives price – regardless of the magnitude, e.g. the notional of the contract – is penalized by the same number, namely the aggregated error to market prices $\eta(Q)$, which does not depend on the derivative X. Second, nonperfect calibration to market prices may also be the result of *systematic underestimation* of market prices. In such a situation, reducing the market prices may not only be unintuitive, but highly dangerous, since the already underestimated price is even further lowered.

An assessment of estimation risk in portfolio credit derivatives valuation, employing a Gaussian copula model, is done by calculating quantiles (Value-at-Risk) and other risk measures by (Heitfield, 2009). Here, the massive parameter risk which comes from historical estimation of Gaussian copula models is recognized. Therefore, results about asymptotic distributions are exploited as, e.g., the Cramér–Rao bound.

Furthermore, there has been some research in treating parameter (and, to a lesser extent, model) risk for risk-neutral valuation purposes. (Lindström, 2010) assumes that the Black–Scholes volatility has a distribution (e.g. normal distribution, although the Black–Scholes volatility is positive by definition). Model prices are then computed by means of a two-step procedure: First, the price, conditioned on a fixed parameter, is calculated. Afterwards, the parameter's distribution is integrated out. (Lindström, 2010) primarily aims at explaining smiles by randomizing the Black–Scholes volatility and does not account for any risk aversion towards parameter risk in the sense of (Knight, 1921).

First ideas to incorporate the likelihood of parameters in a discrete setting have been provided by (Bunnin et al., 2000) in a Bayesian averaging setting, while (Branger and Schlag, 2004) yield ideas to measure model risk with spectral risk measure-type functionals in a discrete way. (Gupta, 2009) and (Gupta et al., 2010) generalize the work of (Cont, 2006) and (Lindström, 2010) by employing general convex risk measures for the measurement of model risk.

There have been several attempts to incorporate Bayesian ideas into derivatives pricing, we only sketch few of them (a complete overview would be out of scope). As described above, derivatives pricing is a situation where one is exposed to parameter risk (and, presumably, model risk). Hence, (Bunnin et al., 2000) and (Gupta and Reisinger, 2012)

suggest to compute the posterior distribution via Bayesian updating incorporating new data like realizations from time series and (more forward-looking) prices of plain vanilla derivatives (e.g. European options). (Gupta and Reisinger, 2012) assume that plain vanilla derivative prices follow a true model that is noised by independent error terms. A mathematical framework is suggested how this assumption is interpreted in terms of a parameter prior distribution. In particular, a local volatility framework is used and they assume that in the short run, the "implied volatilities" of at-the-money options are concise approximations for the local volatility.

1.4 Contributions

In this work, we provide a model (resp. parameter) risk framework that unifies and generalizes concepts from incomplete markets to a general model (resp. parameter) risk situation for derivatives pricing.[9] Furthermore, in the case of risk (and not true uncertainty), i.e. a probability measure R is given on the set of models $\mathcal{Q}$ quantifying the likelihood of the different models, the proposals of (Cont, 2006) and (Lindström, 2010) are generalized using the notion of convex risk measures in the spirit of (Branger and Schlag, 2004; Gupta, 2009; Gupta et al., 2010). Instead of using worst-case methodology or simply integrating out a possible distribution on the models (resp. parameters), general law-invariant convex risk measures are employed in this work to bridge the gap between expectation approaches (as in (Lindström, 2010; Bunnin et al., 2000; Gupta and Reisinger, 2012)) and (possibly penalized) worst-case approaches (as suggested in (Cont, 2006; Gupta et al., 2010)), similar to (Gupta, 2009; Gupta et al., 2010). Furthermore, the pricing suggestions for incomplete markets described in, e.g., (Carr et al., 2001; Cherny and Madan, 2010; Xu, 2006) are embedded and transferred to a general model/parameter risk situation. Our approach allows to capture model (resp. parameter) risk by explicitly incorporating the parameter's distribution, allowing a risk averse trader to acknowledge model (resp. parameter) risk (e.g. arising from historical estimation or calibration to market prices) and to reflect the amount of model (resp. parameter) risk in the width of the bid-ask spread of some derivative. Since we have more comfortable situations in case of parameter risk than model risk, i.e. we have more comprehensive examples for estimators and better intuition how estimation or calibration work, we mainly provide our results for parameter risk. In principle, the results may be equally transfered to a model risk situation as well.

[9] Although our works focus on applications in a financial context, in particular derivatives pricing, our methodology can equally be used in an actuarial context to calculate risk-captured insurance premia.

It turns out that the settings described in (Cont, 2006) and (Lindström, 2010) are both extremal cases of an application of the Average-Value-at-Risk (AVaR) w.r.t. different significance levels $\alpha \in [0, 1]$. Another possible generalization of (Cont, 2006)'s and (Lindström, 2010)'s approaches arises from the entropic risk measure. In general, we introduce the notion of risk-capturing functionals induced by a given convex risk measure. Since we assume traders to be averse towards parameter risk, we suggest to use risk-capturing functionals to determine parameter risk-captured prices in face of a non-Dirac distribution on the parameter space (e.g. the pushforward measure of an estimator). In line with the nonlinear pricing approaches of (Bion-Nadal, 2009; Carr et al., 2001; Xu, 2006; Cherny and Madan, 2010), we propose to use the price delivered by a risk-capturing functional as an ask price. The properties of the parameter's distribution are reflected in the bid-ask spread and we give some examples how to calculate bid-ask prices facing parameter risk.

A desirable feature of (parameter) risk-captured prices is that weak convergence of the parameter distribution implies convergence of the according risk-captured price to the risk-captured price computed with the limit distribution, i.e. if there is a sequence of distributions $(R_N)_{N\in\mathbb{N}}$ on the models/parameters, a risk-captured price $\Gamma(X)$ should fulfill the property that if $R_N \to R$, $N \to \infty$, in the weak sense, one might want that $\Gamma(X; R_N) \to \Gamma(X; R)$, $N \to \infty$. In particular, when this property is fulfilled, the charge for parameter risk is (eventually) decreasing in the available amount of information that is used to estimate the parameter. Actually, if the derivative's price is a continuous and bounded function of some parameter θ on a parameter space Θ, it is shown that the risk-captured price arising from broad classes of convex risk measures fulfills the described convergence property, namely all spectral risk measures (which correspond to the distortion risk measures generated by continuous distortion functions and form a superclass of, e.g., the Average-Value-at-Risk measures) and the entropic risk measure. Furthermore, we state examples where the risk-captured price does not have this property: Roughly spoken, all convex risk measures having some "worst-case part" do not provide this weak convergence. Contrasting to the seminal paper (Krätschmer et al., 2012), we always require weak convergence of the probability measures (while in (Krätschmer et al., 2012), topologies that are possibly stronger than the weak topology on the set of probability measures in the style of (Weber, 2006) are created and continuity properties of law-invariant convex risk measures are assessed w.r.t. those stronger topologies). On the other hand, we state our results for convex risk measures on the set of continuous and bounded functions $\mathcal{C}^b(\Theta)$, while (Krätschmer et al., 2012) state their results for broader domains, i.e. on general Orlicz spaces (which include the Lebesgue

spaces $L^p(\Theta), p \in [1, \infty]$ as a special case).

As a next step, we discuss in detail how to incorporate parameter risk in case of parameter estimation. In particular, we apply our results about weak convergence and obtain that, in case of a consistent estimator for the parameter, this implies that the risk-captured price converges to the price computed with the true parameter. Furthermore, we state approximations for large samples using the delta method and provide closed-form approximation formulae for the AVaR and entropic cases involving sensitivities w.r.t. the parameters, which can be applied for a large class of Maximum Likelihood estimators. For the numerical treatment, we discuss two cases where parameter estimation is a realistic situation one might run into in derivatives valuation. First, we discuss parameter risk of the valuation of an exchange option in case of estimation of the correlation coefficient in a bivariate Black–Scholes model. Second, we provide a more elaborate example assessing the parameter risk when estimating a multifactor model for the clean spark spread[10] and evaluate a gas power plant as a real option on the clean spark spread.

A problem which is mainly motivatived from best practice methodology is how to treat parameter risk in case of calibration to market prices. So, we suggest a procedure how the concept of risk-captured prices can be applied to incorporate calibration risk. In contrast to the situation of estimating parameters historically, calibration does not naturally yield a distribution on the parameter space, making it difficult to speak about true parameter risk/calibration risk. Hence, we propose a methodology how a distribution on the parameter set can be constructed that is in line with the calibration's results. We suggest transforming the function expressing the aggregate error to market prices to obtain a density which can be interpreted as a pricing error-implied distribution. First, we show some consistency properties the error function has to fulfill. Second, we focus on theoretical solutions for obtaining a continuous distribution on the parameter set (most financial market models have "continuous" parameter sets). Furthermore, we focus on the practical problem how to deal with this from a computational point of view, suggesting an algorithm how to obtain a discrete distribution on feasible parameter subsets. As a practical relevant case study, we exemplarily investigate three popular models regarding their parameter risk coming from calibration to market prices: The stochastic volatility models of Heston (see (Heston, 1993)) and Barndorff-Nielsen and Shephard (see (Barndorff-Nielsen and Shephard, 2001)) as well as the Variance Gamma model (see (Madan et al., 1998)). This is done by calculating calibration risk-captured prices for three exotic derivatives (an Asian option, an ITM barrier option, and a lookback

[10]The clean spark spread is the spread of power prices compared to gas and CO2 emissions prices.

option, all observed on a discrete time scale). Hence, this enables us to compare the exposure to parameter risk of the different models and also of the exotic options. As a result, we conclude that in our study, the Heston model bears considerably less parameter risk than the Barndorff-Nielsen–Shephard model, both having decent calibration performance. The Variance Gamma model's calibration is significantly less accurate and the parameter risk within the Variance Gamma model is much higher than in the two stochastic volatility models.

Having set up a framework how to account for parameter risk in derivatives pricing, the question remains which convex risk measure to use for calculating risk-captured prices. At least, it would be helpful to select a convex risk measure from a rich subset, e.g. the risk measures that are Choquet integrals w.r.t. distorted probabilities (a detailed treatment of Choquet integrals can be found in (Denneberg, 1994)). Inspired by (Cherny and Madan, 2010) who use parametric families of distortion risk measures in an incomplete markets setting, a starting point is to calibrate to market prices, i.e. bid-ask prices of liquid securities. We close this gap and discuss how one can re-engineer general distortion risk measures from quoted bid-ask prices. Thus, we provide existence theorems, ensuring that the bid-ask calibration problem can be solved in proper domains of distortion functions. In contrast to (Cherny and Madan, 2010), we suggest a non-parametric approach for obtaining the market-implied distortion risk measure for pricing, not restricting ourselves to a specific parametric shape of the distortion function. This ansatz is based on a piecewise linear approximation. The presented non-parametric approach provides more flexibility for solving the bid-ask calibration problem than choosing from a parametric class and remains simple to implement. Furthermore, it enables us to empirically observe the shape of possible market-implied distortion functions. We present an empirical analysis in the parameter risk setting, comparing our non-parametric calibration to a calibration within the AVaR- and *minmaxvar*-families. When using the non-parametric calibration approach, we obtain a characteristic pattern with a jump close to zero and a linear behavior afterwards. This gives rise to the introduction of an alternative parametric class of distortion functions, the ess sup-expectation convex combinations, which allow for much faster calibration than the AVaR-/*minmaxvar*-families. Furthermore, we enhance the existence theorems to larger families of distortion functions with discontinuities at zero, containing the newly introduced family.

Structure of the remaining work

The remaining thesis is organized as follows: In Chapter 2, we sketch the most important definitions, theorems, and examples from the theory of convex risk measures as well as Choquet integration theory. Chapter 3 introduces model and parameter risk and uncertainty, presents our suggested framework of risk-capturing functionals and risk-captured prices, proves basic properties, and states important examples. Chapter 4 establishes a convergence property for risk-capturing functionals and shows that several classes of functionals fulfill this property under mild technical conditions. Chapter 5 treats parameter risk arising from time series estimation, large sample approximations are stated and the estimation risk of two different financial models is assessed in numerical case studies. In Chapter 6, we discuss how to incorporate calibration risk into the framework of risk-captured prices and scrutinize the parameter risk of different exotic derivatives in different financial market models. Finally, in Chapter 7, we discuss solutions of the calibration to bid-ask prices in different classes of distortion risk measures, suggest a non-parametric calibration scheme, and perform a calibration exercise comparing different calibration methods.

The results of this dissertation are partially published in the papers (Bannör and Scherer, 2013a,c; Bannör et al., 2013), further parts can also be found in the book chapter (Bannör and Scherer, 2013b).

2 Mathematical preliminaries

The measurement of risk (resp. uncertainty) is crucial in many situations in finance. Thus, to ensure accuracy and effectivity, we need a sound mathematical machinery for risk measurement. The specific characteristics, the treatment of practitioners, and the economic interpretation of risky financial positions has given rise to the development of the theory of convex risk measures, combining results from convex analysis, functional analysis, and probability theory. In this chapter, we sketch a brief introduction into the theory of convex risk measures, provide basic theorems that are used later in the dissertation, and recall important examples for convex risk measures. Furthermore, this thesis heavily uses the subclass of distortion risk measures, i.e. convex risk measures that can be represented as a Choquet integral w.r.t. a distorted probability. Hence, we need some deeper knowledge about Choquet integration theory to understand the mathematics behind distorted probabilities and expectations w.r.t. them. Thus, we state the basic definitions, facts, and theorems on Choquet integration based on the standard reference book (Denneberg, 1994) to provide a deeper understanding of the Choquet integral.

2.1 Theory of convex risk measures

Risk, particularly in a financial or actuarial context, has increasingly been the subject of discussion in a professional environment, but also more and more in academia and during the last years even in broader circles of society. The financial market crisis of 2008, having its peak in the bankruptcy of Lehman Brothers in September 2008, has emerged a vivid discussion in society what kind of risks and to which extent risks, particularly in finance, should be taken by financial institutions. But, obviously, the ladder problem – determining to which extent financial institutions are allowed to take risks – is immediately linked to another problem, which is broadly discussed in academia: Measuring and quantifying the risk of financial positions. A popular method among practitioners and regulators has been the Value-at-Risk measure: Given a financial position X (interpreted

as a possible financial loss), which is modeled as a random variable on some probability space $(\Omega, \mathcal{F}, P)$, the Value-at-Risk w.r.t. a safety or significance level $\alpha \in (0,1)$ is given by the upper α-quantile of X, hence $\mathrm{VaR}_\alpha(X) = q_{1-\alpha}(X)$. This number aggregates the risk contained in the risky position X into one single number which has an appealing interpretation: In $100 \cdot (1-\alpha)\%$ of all possible cases, the possible loss that may be caused by the position X is overstated by the α-Value-at-Risk. Hence, determining the "risk of the position" X by the Value-at-Risk has many advantages, particularly for risk controlling and regulation.

Unfortunately, the Value-at-Risk bears some mathematical problems and exhibits some counterintuitive behavior regarding diversification effects: Given two risky positions X and Y, combining fractions of the two positions should account for positive effects since unsystematic effects may partially be hedged away due to diversification between the positions X and Y. But, the Value-at-Risk does not necessarily account for this diversification effect, since one can construct surprisingly easy situations where diversification between two positions X and Y is even punished (cf., e.g., (Artzner et al., 1999)). These effects are, amongst others, due to the fact that the Value-at-Risk only accounts for the probability of losses, but not at all for their magnitudes.

Hence, in the academic community, an ongoing discussion about which properties a functional measuring the risk of random variables should fulfill established. The seminal paper (Artzner et al., 1999) stated a system of axioms mapping the economic requirements for numbers measuring risk into mathematical properties (cash invariance, monotonicity, subadditivity, and positive homogenity). Later, subadditivity and positive homogenity were relaxed to convexity (cf. (Föllmer and Schied, 2002)) and even to quasi-convexity (cf. (Cerreia-Vioglio et al., 2011)).

We start by recalling the definition of convex risk measures on a vector space of measurable functions, as it can be found in the textbook (Föllmer and Schied, 2004). Further important contributions to this wide field have, e.g., been done in (Kusuoka, 2001; Föllmer and Schied, 2002; Acerbi and Tasche, 2002; Jouini et al., 2006; Frittelli and Scandolo, 2006; Krätschmer, 2006) and many other publications. We follow (Frittelli and Scandolo, 2006) in extending the classical definition to translation invariance w.r.t. a linear form, since we want translation invariance to hold not only for constants but also for all contingent claims without exposure towards model (resp. parameter) uncertainty.

Definition 2.1.1 (Convex risk measure)
Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{X} \subset \mathcal{L}^0(\Omega)$ be a vector space of measurable functions

on Ω. Let $\mathcal{Y} \subset \mathcal{X}$ be a sub-vector space and $\pi \in \mathcal{Y}^\star$, denoting by

$$\mathcal{Y}^\star := \{\lambda : \mathcal{Y} \to \mathbb{R} : \ \lambda \textit{ linear}\}$$

the algebraic dual space of $\mathcal{Y}$. ρ is called a convex risk measure *with π-translation invariance[1] if it fulfills the following axioms:*

1. *ρ is monotone[2]: $\forall X, Y \in \mathcal{X} : X \geq Y \Rightarrow \rho(X) \geq \rho(Y)$.*
2. *ρ is convex: $\forall X, Y \in \mathcal{X}\, \forall \lambda \in [0,1] : \rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$.*
3. *ρ is π-translation invariant: $\forall X \in \mathcal{X}\, \forall Y \in \mathcal{Y} : \rho(X+Y) = \rho(X) + \pi(Y)$.*

Furthermore, if ρ is a convex risk measure, we say that

1. *ρ is coherent, if it is additionally positively homogeneous, i.e. $\rho(cX) = c\rho(X)$ holds for all $c > 0$ and $X \in \mathcal{X}$.*
2. *ρ is normalized, if $\rho(0) = 0$ holds.*
3. *If P is a probability measure on $(\Omega, \mathcal{F})$, we call ρ P-law invariant, if $\rho(X) = \rho(Y)$ holds in case of $P^X = P^Y$, denoting by*

$$P^X(A) := P(X \in A), \quad A \in \mathcal{B}(\mathbb{R}),$$

the pushforward probability measure induced by the random variable X. If the probability measure P is canonical, we just speak about law invariance instead of P-law invariance.

The theory of convex risk measures has its origin in the shortcoming of the Value-at-Risk (VaR): This risk measure (which is essentially a quantile) was advocated by J.P. Morgan in the 1990s, but does not satisfy the convexity property (as pointed out by (Artzner

[1] If we do not mention the translation invariance w.r.t. a specified linear form π and the sub-vector space $\mathcal{Y}$, we always assume that $\mathcal{Y}$ is the vector space of constant functions and π is the canonical linear form with $\pi(1) = 1$.

[2] Most contributions (e.g. (Föllmer and Schied, 2004; Artzner et al., 1999; Krätschmer, 2006; Frittelli and Scandolo, 2006)) define convex risk measures to be anti-monotone and anti-translation invariant. To match our purposes and for the sake of elegance, we follow (Cont, 2006) by employing ordinary monotonicity and translation invariance.

et al., 1999)). Hence, the Average-Value-at-Risk (AVaR)[3] was developed (see (Acerbi and Tasche, 2002)) to overcome this problem.

Example 2.1.2 (Average-Value-at-Risk)
Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\alpha \in (0,1]$ a given significance level*, and $X \in \mathcal{L}^1(P)$ a random variable. Then we define the* α-Value-at-Risk *to be the upper α-quantile* $\mathrm{VaR}_\alpha(X) := q_X^P(1-\alpha)$ *and the* α-Average-Value-at-Risk *to be the integrated upper tail of the random variable X, i.e.*

$$\mathrm{AVaR}_\alpha(X) := \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(X)\,\mathrm{d}\beta.$$

Obviously, the Average-Value-at-Risk dominates the Value-at-Risk due to

$$\begin{aligned}\mathrm{AVaR}_\alpha(X) &= \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(X)\,\mathrm{d}\beta \\ &\geq \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\alpha(X)\,\mathrm{d}\beta = \mathrm{VaR}_\alpha(X), \quad \alpha \in (0,1].\end{aligned}$$

Furthermore, for a normally distributed variable $X \sim \mathcal{N}(\mu, \sigma^2)$, the Average-Value-at-Risk w.r.t. some significance level $\alpha \in (0,1]$ can be expressed in a closed-form expression (cf. (McNeil et al., 2005, p. 45)) via

$$\mathrm{AVaR}_\alpha(X) = \mu + \sigma\frac{\varphi(\Phi^{-1}(1-\alpha))}{\alpha}, \tag{2.1}$$

as usually denoting by Φ^{-1} the quantile function of the standard normal distribution and by φ the density function of the standard normal distribution.

It can be shown that the Average-Value-at-Risk w.r.t. some significance level $\alpha \in (0,1]$ is a coherent, law-invariant risk measure (cf. (Acerbi and Tasche, 2002)). Moreover, for $\alpha \searrow 0$ and X P-a.s. bounded, $\mathrm{AVaR}_\alpha(X)$ converges to the P-essential supremum (cf. (Föllmer and Schied, 2004, Remark 4.45)). Hence, the definition of the Average-Value-at-Risk can sensibly be extended by defining

$$\mathrm{AVaR}_0(X) := \operatorname{ess\,sup} X, \quad X \in L^\infty(P).$$

The functionals examined in the theory of convex risk measures exhibit some attractive properties: Convex analysis has introduced the notion of the Fenchel–Moreau transform

[3] In some publications, the Average-Value-at-Risk (AVaR) is also referred to as the Conditional-Value-at-Risk (CVaR), Tail-Value-at-Risk (TVaR), or Expected Shortfall (ES). Although there are slight differences in the definitions, which matter in some special cases, these alternative risk measures coincide with the Average-Value-at-Risk in most practical relevant cases (cf. (Föllmer and Schied, 2004, Corollary 4.49)).

of a function defined via the dual space of its domain. A key result of functionals is the dual representation theorem, where a canonical representation of every convex risk measure is developed and, simultaneously, a canonical construction principle for the development of new convex risk measures, is described.

Theorem 2.1.3 (Dual representation of convex risk measures)
Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{X} \subset \mathcal{L}^0(\Omega)$ be a vector space of measurable functions on Ω. Furthermore, let $\mathcal{Y} \subset \mathcal{X}$ be a sub-vector space and $\pi \in \mathcal{Y}^\star$ be a linear form on $\mathcal{Y}$. If ρ is a convex risk measure on $\mathcal{X}$ w.r.t. $(\mathcal{Y}, \pi)$, then ρ has the dual representation

$$\rho(X) = \sup_{\lambda \in \mathcal{X}_+^{\star,\pi}} \lambda(X) - \rho^\star(\lambda),$$

with $\rho^\star$ being the Fenchel–Moreau transform *(or convex dual) of ρ given by*

$$\rho^\star(\lambda) := \sup_{X \in \mathcal{X}} \rho(X) - \lambda(X)$$

and

$$\mathcal{X}_+^{\star,\pi} := \left\{ \lambda \in \mathcal{X}_+^\star : \lambda_{|\mathcal{Y}^\star} = \pi \right\}$$

being all positive linear forms on $\mathcal{X}$ extending π.

Proof
See (Krätschmer, 2006, Proposition 1.1). □

The interpretation of the dual representation is as follows: A convex risk measure can be regarded as the supremum of some linear functionals λ (which are, in many cases, expectations w.r.t. probability measures), where each linear functional is additionally penalized by the convex dual $\rho^\star(\lambda)$, which could be interpreted as the "trustworthyness" of the functional λ.

A crucial (topological) property of convex risk meaures is the so-called Fatou property.

Definition 2.1.4 (Fatou property of convex risk measures)
Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{X} \subset \mathcal{L}^0(\Omega)$ be a vector space of measurable functions on Ω. If ρ is a convex risk measure on $\mathcal{X}$, then ρ carries the Fatou property*, if for every uniformly bounded and pointwise convergent sequence of functions $(X_N)_{N\in\mathbb{N}}$ in $\mathcal{X}$ (i.e. there is a $C > 0$ such that $|X_N| \leq C$ for all $N \in \mathbb{N}$ and $X_N \to X$ for some $X \in \mathcal{X}$), $\rho(X) \leq \liminf_{N\in\mathbb{N}} \rho(X_N)$ holds.*

The interpretation of the Fatou property is as follows: If some financial position X may be approximated by other positions $(X_N)_{N\in\mathbb{N}}$ and it may be easy to calculate $\rho(X_N)$,

but the calculation of $\rho(X)$ is difficult, one can approximate $\rho(X)$ by $\rho(X_N)$ without running the risk of underestimating the risk of X, if $N \in \mathbb{N}$ is chosen large enough. In case of law invariance w.r.t. some probability measure P on (Ω, F), one can relax the condition of pointwise convergence to P-a.s. pointwise convergence (cf. (Föllmer and Schied, 2004, p. 171ff)). It can been shown that broad classes of convex risk measures exhibit the Fatou property (cf. (Jouini et al., 2006; Krätschmer, 2006)).

An important class of convex risk measures are the so-called spectral risk measures, which have been introduced in (Acerbi, 2002). As a direct generalization of the AVaR, where the tail of a distribution is weighted uniformly, spectral risk measures introduce a general weighting on the quantiles such that upper quantiles are weighted relatively stronger than lower quantiles.

Definition 2.1.5 (Spectral risk measure)
Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X \in \mathcal{L}^\infty(P)$, and $\phi : [0,1] \to \mathbb{R}_{\geq 0}$ be a decreasing and normed (i.e. $\int_0^1 \phi(x)\,\mathrm{d}x = 1$) function. Then the functional $\rho_\phi : \mathcal{L}^\infty(P) \to \mathbb{R}$ defined by

$$\rho_\phi(X) := \int_0^1 \mathrm{VaR}_\alpha(X)\phi(\alpha)\,\mathrm{d}\alpha$$

is called a spectral risk measure *with spectrum ϕ.*

A detailed discussion of spectral risk measures can be found in (Acerbi, 2002). In particular, it is shown that spectral risk measures are a subclass of coherent risk measures. (Acerbi, 2002) provides a representation theorem for spectral risk measures, showing that spectral risk measures are actually a (possibly infinite) convex combination of $(\mathrm{AVaR}_\alpha)_{\alpha\in(0,1]}$ risk measures.

Proposition 2.1.6 (Representation theorem for spectral risk measures)
Let ρ_ϕ be a spectral risk measure with spectrum ϕ. Then there exists a measure μ on $(0,1]$ such that ρ_ϕ can be calculated as

$$\rho_\phi(X) = \int_0^1 \mathrm{AVaR}_\alpha(X)\,\mu(\mathrm{d}\alpha),$$

i.e. ρ_ϕ is a (possibly infinite) convex combination of different AVaR *risk measures.*

Proof
In (Acerbi, 2002), one shows that there exists a measure μ on $[0,1]$ with the upper construction. Since $\mathrm{AVaR}_0(X) = \operatorname{ess\,sup} X$ does not have any influence on ρ_ϕ according to Definition 2.1.5, the singleton $\{0\}$ has to be a μ-nullset. □

2.2 Choquet integration theory

Classical integration theory treats how to measure areas and volumes. As a special case, one has integration theory w.r.t. probability measures where the integral w.r.t. some probability measure is exactly the expected value. One of the key properties of the regular integral is that integrals are linear functionals. But, in reality, one can set-up experiments where people which act according to some "expected value" that does not behave linearly, but exhibits sub- or superlinear behavior (e.g. (Schmeidler, 1989)). Thus, a generalization of the regular integration theory has been developed to capture sub- or superlinearity of integrals.

Corresponding to classical integration theory, where one integrates w.r.t. some (probability) measure, the integrator in Choquet integration theory can be more general – one can use monotone set functions for integration. We will not present Choquet integration theory in its most general variety, for these purposes we refer to the textbook (Denneberg, 1994). We start with some interesting properties set functions may exhibit (it may be noted that all of these properties hold for measures).

Definition 2.2.1 (Set functions)
Let Ω be a nonempty set and $\mathcal{F} \subset \mathfrak{P}(\Omega)$ be a set field, i.e. a nullset-containing collection of sets being closed w.r.t. complements and unions. Then $\mu : \mathcal{F} \to \mathbb{R}_{\geq 0}$ is called a set function*, if $\mu(\emptyset) = 0$. Furthermore, μ is called:*

1. monotone*, if for $A, B \in \mathcal{F}$, $A \subset B$, $\mu(A) \leq \mu(B)$ holds;*[4]
2. submodular*, if for $A, B \in \mathcal{F}$, $\mu(A \cap B) + \mu(A \cup B) \leq \mu(A) + \mu(B)$ holds;*
3. supermodular*, if for $A, B \in \mathcal{F}$, $\mu(A \cap B) + \mu(A \cup B) \geq \mu(A) + \mu(B)$ holds;*
4. additive*, if μ is sub- and supermodular.*

In classical integration theory, one develops the integral by introducing it for simple functions[5] and then applying a limit procedure to extend the integral to the whole set of (integrable) functions (cf., e.g., (Klenke, 2008)). Differently, the Choquet integral is defined via improper Riemann integrals over the decumulative distribution function (which is also often called survival function).

[4] In some papers and textbooks (e.g. (Föllmer and Schied, 2004)), a monotone set function is refered to as a *capacity*, which relates to an interpretation from physics. We follow the terminology of (Denneberg, 1994) and call this type of functions *monotone set functions*, since it is more comprehensive for non-expert readers.

[5] In this setting, we call $X : \Omega \to \mathbb{R}$ a *simple function*, if there exist $n \in \mathbb{N}$, sets $A_1, \ldots, A_n \in \mathcal{F}$, and

Definition 2.2.2 (Choquet integral)
Let Ω be a nonempty set, $\mathcal{F}$ a set field on Ω, and μ a monotone finite set function. Let furthermore $X : \Omega \to \mathbb{R}$ be a random variable. The Choquet integral *of X w.r.t. μ is then defined as*

$$\int_\Omega X \,\mathrm{d}\mu := \int_{-\infty}^{0} \mu(X > x) - \mu(\Omega)\,\mathrm{d}x + \int_0^\infty \mu(X > x)\,\mathrm{d}x.$$

When the domain Ω is unambiguous, we simply write $\int X \,\mathrm{d}\mu$ instead of $\int_\Omega X \,\mathrm{d}\mu$.

The above definition of the Choquet integral is an immediate generalization of the classical measure integral (provided that we restrict ourselves to finite set functions and measures), since the defining equality also holds for the classical integral w.r.t. some finite measure.[6]

In some cases, the Choquet integral may be infinite or not defined. The improper Riemann integrals of the (shifted) decumulative distribution function are always well-defined, since the decumulative distribution function $G_{X,\mu}(x) := \mu(X > x)$ is a monotone decreasing function. If both Riemann integrals are infinite, we run into a pathological situation and declare such a Choquet integral as "not defined". It can be easily shown that the Choquet integral can be represented equivalently as a Riemann–Stieltjes integral w.r.t. the cumulative distribution function, on which we rely on in some proofs later.

Proposition 2.2.3 (Choquet integral as a Riemann–Stieltjes integral)
Let Ω be a nonempty set, $\mathcal{F}$ a set field on Ω, and μ a finite monotone set function on $\mathcal{F}$. If $X : \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable function and the Choquet integral is defined, it can be equally calculated via

$$\int X \,\mathrm{d}\mu = -\int_{-\infty}^{\infty} x \, F_{-X}(\mathrm{d}x),$$

denoting – as usually – by $F_{-X}(x) = \mu(-X \leq x)$ the cumulative distribution function.

Proof
The proof follows (Denneberg, 1994, Exercise 5.3). First note that the Lebesgue–Stieltjes measure ν which is yielded by $\nu((a, b]) := \mu(X > a) - \mu(X > b)$ (use the classical

constants $c_1, \ldots, c_n \in \mathbb{R}$ such that $X = \sum_{j=1}^n c_j \mathbb{1}_{A_j}$, denoting by

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \text{otherwise} \end{cases}$$

the *indicator function* w.r.t. the set $A \in \mathcal{F}$.

[6] In most applications of Choquet integration theory, one uses normed set functions μ with $\mu(\Omega) = 1$.

Carathéodory extension procedure to obtain the whole measure), $a \leq b$, $a, b \in \mathbb{R}$, has the decumulative distribution function $G_\nu(a) = \nu((a, \infty)) = \mu(X > a)$, $a \in \mathbb{R}$. Now, the integral of the identity w.r.t. ν can be written as

$$\int_{-\infty}^{\infty} x \, \nu(\mathrm{d}x) = \int_{-\infty}^{\infty} x \, G_\nu(\mathrm{d}x),$$

denoting $G_\nu(x) := \nu((x, \infty))$. On the other hand, the Lebesgue–Stieltjes integral coincides with the Choquet integral, hence

$$\begin{aligned}
\int_{-\infty}^{\infty} x \, \nu(\mathrm{d}x) &= \int_{-\infty}^{0} G_\nu(x) - \nu(\mathbb{R}) \, \mathrm{d}x + \int_{0}^{\infty} G_\nu(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{0} \mu(X > x) - \mu(\Omega) \, \mathrm{d}x + \int_{0}^{\infty} \mu(X > x) \, \mathrm{d}x \\
&= \int_{\Omega} X \, \mathrm{d}\mu
\end{aligned}$$

holds. The assertion equally holds for F_{-X} instead of G_X due to the fact that the set of discontinuity points of a monotone function is at most countable and the improper Riemann integrals of two monotone functions that coincide on dense subsets are equal (cf. (Denneberg, 1994, Lemma 1.3)). □

The Choquet integral has some properties which generalize the properties from the regular integral w.r.t. a (probability) measure. In particular, the monotonicity and positive homogeneity of the integral prevails and the set function is enhanced in a natural manner (cf. (Denneberg, 1994, Proposition 5.1)). Furthermore, the Choquet integral exhibits properties which depend on the set function, i.e. super- (resp. sub-)modularity implies that the Choquet integral is super- (resp. sub-)additive (cf. (Denneberg, 1994, Theorem 6.3, Corollary 6.4)).

Theorem 2.2.4 (Properties of the Choquet integral)
Let Ω be a nonempty set, $\mathcal{F}$ a set field on Ω, and μ a monotone set function on $\mathcal{F}$. Let $X, Y : \Omega \to \mathbb{R}$ be μ-a.s. bounded random variables such that their Choquet integrals exist, then the following properties hold for the Choquet integral w.r.t. μ:

1. *It is monotone, i.e. if $X \leq Y$, then*

$$\int X \, \mathrm{d}\mu \leq \int Y \, \mathrm{d}\mu$$

   holds;

2. *Is is positively homogeneous, i.e. for* $c > 0$,

$$\int cX \, \mathrm{d}\mu = c \int X \, \mathrm{d}\mu$$

holds;

3. *It enhances* μ*, i.e. if* $A \in \mathcal{F}$*, then* $\mu(A) = \int \mathbb{1}_A \, \mathrm{d}\mu$ *holds;*

4. *It is comonotonic additive, i.e. if* X, Y *are comonotone, i.e. for* $\omega_1, \omega_2 \in \Omega$ $(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0$[7]*, then*

$$\int X + Y \, \mathrm{d}\mu = \int X \, \mathrm{d}\mu + \int Y \, \mathrm{d}\mu$$

follows;

5. *If* μ *is submodular, then the Choquet integral is subadditive, i.e.*

$$\int X + Y \, \mathrm{d}\mu \leq \int X \, \mathrm{d}\mu + \int Y \, \mathrm{d}\mu;$$

6. *If* μ *is supermodular, then the Choquet integral is superadditive, i.e.*

$$\int X + Y \, \mathrm{d}\mu \geq \int X \, \mathrm{d}\mu + \int Y \, \mathrm{d}\mu.$$

Proof

See (Denneberg, 1994, Proposition 5.1, Theorem 6.3, Corollary 6.4). □

An interesting property of the Choquet integral is the comonotonic additivity. If two random variables X, Y are comonotonic, i.e. the move from one state ω_1 to another state ω_2 is always in the same direction, the Choquet integral behaves just like the classical integral and exhibits classical additivity. Being put in a financial context, one may interpret two comonotonic random variables X, Y as financial positions that move co-directionally as a function of the state. Roughly spoken, the financial position X does not exhibit losses, if and only if the financial position Y does not exhibit losses. Hence, diversification between the two financial positions X and Y does not make a lot of sense.[8] Using the Choquet integral as a "risk measure" in a financial context,

[7] An equivalent definition of comonotonicity is that there is a random variable Z such that Y and X are an increasing functions of Z, i.e. there exist increasing functions $f, g : \mathbb{R} \to \mathbb{R}$ such that $X = f(Z)$, $Y = g(Z)$, cf. (Denneberg, 1994, Proposition 4.5).

[8] One might argue to account for different sizes of losses, but, at least, the positions do not provide any hedges towards each other's losses.

comonotonic additivity means that no diversification is possible between comonotonic financial positions.

One can bridge the gap between convex risk measures and Choquet integrals by the following useful representation theorem, which is the Choquet integration equivalent of the classical Daniell–Stone theorem in measure integration theory (one can even obtain the Daniell–Stone theorem as a corollary, cf. (Denneberg, 1994)). It states that *any* functional exhibiting monotonicity and comonotonic additivity is actually a Choquet integral and one can easily construct the set function from the functional.

Theorem 2.2.5 (Representation theorem of Greco and Schmeidler)
Let $\mathcal{F}$ be a σ-algebra on a set Ω and denote by $\mathcal{L}^\infty(\mathcal{F})$ the set of bounded $\mathcal{F}$-measurable functions on Ω. If $\Gamma : \mathcal{L}^\infty(\Omega) \to \mathbb{R}$ is a monotone and comonotonic additive functional, i.e.

1. *for $X, Y \in \mathcal{L}^\infty(\mathcal{F})$, $X \leq Y$, $\Gamma(X) \leq \Gamma(Y)$ holds, and,*

2. *for $X, Y \in \mathcal{L}^\infty(\mathcal{F})$ comonotonic, $\Gamma(X+Y) = \Gamma(X) + \Gamma(Y)$ holds,*

then $\mu(A) := \Gamma(\mathbb{1}_A)$ is a monotone set function and

$$\Gamma(X) = \int_\Omega X \, \mathrm{d}\mu$$

holds.

Proof
See (Denneberg, 1994, Theorem 11.2). □

Many assumptions in the above theorem can actually be relaxed, as in (Denneberg, 1994, Chapter 13), but we present it in the convenient version above. As a corollary, Theorem 2.2.5 can characterize all comonotonically additive convex risk measures as Choquet integrals (cf. (Föllmer and Schied, 2004, Theorem 4.82)).

Corollary 2.2.6
If $\mathcal{F}$ is a σ-algebra on a set Ω and $\Gamma : \mathcal{L}^\infty(\mathcal{F}) \to \mathbb{R}$ is a convex risk measure which is additionally comonotonically additive, then it is a Choquet integral w.r.t. the submodular set function $\mu(A) := \Gamma(\mathbb{1}_A)$.

Distorted probabilities

A prime example for Choquet integration, which has been richly exploited for constructing convex risk measures (cf., e.g., (Föllmer and Schied, 2004; Hürlimann, 2004; Gzyl and Mayoral, 2008; Balbás et al., 2009)), is given by distorted probabilities. Grounded in behavioral finance theory, one can exhibit that – in several situations – people that know about probabilities do not use the raw probabilities, but "distort" them, i.e. assign higher "subjective probabilities" to objectively small probabilities. It may occur that this is done systematically according to the original probability level α, i.e. all events with an objective probability α are assigned a subjective probability $\gamma(\alpha)$. In this case, mathematically spoken, a *distortion function* is applied to the probabilities.

Definition 2.2.7 (Distortion function)
Let $\gamma : [0,1] \to [0,1]$. γ is called a distortion function *if γ is monotone, $\gamma(0) = 0$, and $\gamma(1) = 1$.*

The interpretation of distortion functions is as follows: Given a probability space $(\Omega, \mathcal{F}, P)$, instead of measuring the probability of a set A classically via $P(A)$, we alternatively consider the distorted probability $\gamma(P(A))$ w.r.t. some distortion function γ. Obviously, the set function $\gamma \circ P : \mathcal{F} \to [0,1]$ is not a probability measure any more for general γ, but still preserves monotonicity w.r.t. the set order: $A \subset B$, $A, B \in \mathcal{F}$ implies $\gamma(P(A)) \leq \gamma(P(B))$. Hence, instead of calculating the expected value w.r.t. P, one can calculate the Choquet integral w.r.t. the set function $\gamma \circ P$

$$\int_\Omega X \, \mathrm{d}(\gamma \circ P) \tag{2.2}$$

for some random variable X. The Choquet integral (2.2) can be interpreted as a "subjective expected value" where the original, objective probabilities provided by the measure P are *reweighted* with the distortion function γ.

Furthermore, it can be shown that if the distortion function γ is concave, the Choquet integral is a subadditive functional (cf. (Denneberg, 1994, Example 2.1)). Thus, Choquet integrals w.r.t. concave distorted probabilities fulfill the axioms of convex risk measures (even coherent risk measures) and are often referred to as *distortion risk measures*.

Definition 2.2.8 (Distortion risk measure)
Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\Gamma : \mathcal{L}^\infty(P) \to \mathbb{R}$ be a convex risk measure on the set of P-a.s. bounded random variables. Γ is called a distortion risk measure*, if there*

exists a concave distortion function $\gamma : [0,1] \to [0,1]$ *such that* Γ *is the Choquet integral w.r.t.* $\gamma \circ P$*, i.e.*

$$\Gamma(X) = \int_\Omega X \,\mathrm{d}(\gamma \circ P) \quad \textit{for each } X \in \mathcal{L}^\infty(P).$$

Some authors (e.g. (Acerbi, 2002)) treat so-called spectral risk measures instead, which is a somewhat equivalent approach to distortion risk measures (cf. (Föllmer and Schied, 2004; Gzyl and Mayoral, 2008)). One of the well-known examples for a distortion risk measure is the Average-Value-at-Risk (AVaR), which was already mentioned in Example 2.1.2.

Example 2.2.9 (Average-Value-at-Risk)
It can easily be shown (cf., e.g., (Balbás et al., 2009)) that the AVaR_α *for some significance level* $\alpha \in (0,1]$ *can be represented as a Choquet integral w.r.t. a distorted probability. The corresponding distortion function is given by*

$$\gamma_\alpha(u) = \begin{cases} \frac{u}{\alpha}, & u \in [0,\alpha] \\ 1, & \textit{otherwise} \end{cases}$$

and γ_α *is obviously concave. For* $\alpha = 0$ *(which is the essential supremum), the distortion function is simply given by the indicator function* $\mathbb{1}_{(0,1]}$*.*

As a more elaborate example, (Cherny and Madan, 2009) introduce several parametric families of concave distortion functions and calculate positions w.r.t. the induced distortion risk measures. The following *minmaxvar*-family of concave distortions is introduced in (Cherny and Madan, 2009) and successfully employed in (Cherny and Madan, 2010).

Example 2.2.10 (*minmaxvar*-family of concave distortions)
Let $\psi_x(y) : [0,1] \to [0,1]$*,* $x \in \mathbb{R}_{\geq 0}$ *be defined by*[9]

$$\psi_x(y) := 1 - \left(1 - y^{\frac{1}{x+1}}\right)^{x+1}.$$

It can easily be verified that ψ_x *is a concave distortion function.*

[9] Choosing $x \in \mathbb{N}$ allows for the following interpretation of the *minmaxvar*-distortion functions: If a random variable $X \in L^1(P)$ has the same distribution as $\min\{Z_1, \ldots, Z_{x+1}\}$, $Z_1, \ldots, Z_{x+1}$ i.i.d., then

$$\int X \,\mathrm{d}(\psi_x \circ P) = \int \max\{Z_1, \ldots, Z_{x+1}\} \,\mathrm{d}P$$

follows.

As we have shown above, all Choquet integrals are comonotonically additive and all convex risk measures which additionally exhibit comonotonic additivity are actually Choquet integrals. Given a probability space $(\Omega, \mathcal{F}, P)$, a distortion risk measure is a somehow "natural" approach where law invariance w.r.t. P is naturally preserved. And, actually, another representation theorem yields that – under mild technical conditions – the distortion risk measures are the only Choquet integrals (hence the only comonotonically additive convex risk measures) exhibiting law invariance.

Theorem 2.2.11

Let $(\Omega, \mathcal{F}, P)$ be an atomless[10] probability space and $\Gamma : L^\infty(P) \to \mathbb{R}$ be a law invariant convex risk measure which is additionally comonotonically additive. Then there exists a concave distortion function $\gamma : [0,1] \to [0,1]$ such that Γ is a Choquet integral w.r.t. the distorted probability $\gamma \circ P$, i.e.

$$\Gamma(X) = \int_\Omega X \, \mathrm{d}(\gamma \circ P), \quad X \in L^\infty(P).$$

Proof

See (Föllmer and Schied, 2004, Theorem 4.87). □

Theorem 2.2.11 shows that the class of distortion risk measures is a natural subclass of convex risk measures, providing both the rather natural properties of comonotonic additivity and law invariance. Furthermore, calculations in the set of distortion risk measures are tractable when using the representation provided in Proposition 2.2.3.

Corollary 2.2.12 (Riemann–Stieltjes representation)

Let Γ be a distortion risk measure which is associated with some concave distortion function $\gamma : [0,1] \to [0,1]$. Then we can calculate the distortion risk measure of some claim $X \in \mathcal{L}^\infty(P)$ as a Riemann–Stieltjes integral via

$$\Gamma(X) = -\int_{-\infty}^{\infty} x \, (\gamma \circ F_{-X})(\mathrm{d}x),$$

as usually denoting by $F_{-X}(x) := P(-X \leq x)$ the cumulative distribution function of $-X$ w.r.t. probability measure P.

Proof

Follows immediately from plugging in the definition of a concave distortion function in the Riemann–Stieltjes integral representation from Proposition 2.2.3. □

[10] A probability space is called *atomless*, if it supports a standard uniformly distributed random variable X, i.e. $X \sim U[0,1]$.

3 Model and parameter risk – a convex risk measure ansatz

In this chapter, we describe and discuss our ansatz to treat model and parameter risk. Therefore, we first formally define model and parameter uncertainty and risk and state some examples (including classical ones like incomplete markets). We introduce our concept of risk-capturing functionals and risk-captured bid/ask prices, which treat model and parameter risk using convex risk measures in the spirit of (Branger and Schlag, 2004; Gupta, 2009; Gupta et al., 2010), extending the ideas of (Carr et al., 2001; Xu, 2006; Cherny and Madan, 2010) from the incomplete markets setting. By employing different convex risk measures, we state examples which generalizing both the worst-case approach of (Cont, 2006) as well as the treatment of (Lindström, 2010).

Notation

Let throughout the remaining thesis $(\Omega, \mathcal{F}, \mathbb{F})$ be a filtered measurable space and let $(S_t)_{t\geq 0}$ denote a d-dimensional $\mathbb{F}$-adapted stochastic process modelling the basic instruments with $S_t = (S_t^{(1)}, \ldots, S_t^{(d)})$, $d \in \mathbb{N}$. To simplify notation, we assume without loss of generality all claims being evaluated w.r.t. a martingale measure Q, typically denoted by X, to be already discounted by the associated numéraire process.

3.1 Model and parameter uncertainty in derivatives pricing

In this section, we define model and parameter uncertainty and briefly recall the approaches of (Cont, 2006) and (Lindström, 2010) to capture model resp. parameter uncertainty. Furthermore, we state some examples of models from derivatives pricing where model resp. parameter uncertainty is crucial.

Definition 3.1.1 (Model uncertainty)
Let $\mathcal{Q}$ be a family of probability measures on $(\Omega, \mathcal{F})$ such that all stochastic processes modelling discounted basic instruments $(S_t^{(i)})_{t\geq 0}, i = 1, \ldots, d$, are Q-martingales for all $Q \in \mathcal{Q}$. The financial market model $(\Omega, \mathcal{F}, \mathbb{F}, (S_t)_{t\geq 0}, \mathcal{Q})$ faces model uncertainty *if $|\mathcal{Q}| > 1$, where $|\mathcal{Q}|$ denotes the cardinality of $\mathcal{Q}$.*

Definition 3.1.2 (Parameter uncertainty)
Let $(P_\theta)_{\theta\in\Theta}$ be a family of pairwise different probability measures on $(\Omega, \mathcal{F})$ such that the models $(\Omega, \mathcal{F}, \mathbb{F}, (S_t)_{t\geq 0}, P_\theta)$ are arbitrage-free for all $\theta \in \Theta$. The model faces parameter uncertainty *if $|\Theta| > 1$.*[1]

Remark 3.1.3 (Interpreting parameter uncertainty as model uncertainty)
Obviously, parameter uncertainty is a special case of model uncertainty, since every $\theta \in \Theta$ induces a family of equivalent martingale measures $\mathcal{Q}_\theta$. Therefore, exposure towards model uncertainty exists w.r.t. the model family $\mathcal{Q} = \bigcup_{\theta\in\Theta} \mathcal{Q}_\theta$, just as pointed out in (Cont, 2006). In case of incomplete models P_θ, we have uncertainty in two ways: The parameter uncertainty about the true real-world model P_θ and furthermore model uncertainty about the equivalent martingale measure $Q \in \mathcal{Q}_\theta := \{Q : Q \sim P_\theta\}$ obtained by a change of measure.

As a prominent special case, which is extensively treated in the literature, model uncertainty arises in incomplete markets with *real-world measure* P, where all measures $Q \in \mathcal{Q}$ are equivalent. This special case of model uncertainty is widely understood and hedging proposals are provided.

Example 3.1.4 (Incomplete markets)
Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space, $(S_t)_{t\geq 0}$ the asset price process and let

$$\mathcal{Q} := \{Q \text{ probability measure on } (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}) : Q \sim P, (S_t)_{t\geq 0} \text{ is } Q\text{-martingale}\}$$

be the set of P-equivalent martingale measures. General arbitrage pricing theory has shown (cf. (Harrison and Pliska, 1983)) that there is model uncertainty w.r.t. $\mathcal{Q}$ if and only if there exist contingent claims that cannot be hedged by a dynamic trading strategy. Thus, this situation is called an incomplete market*. This kind of model uncertainty primarily arises from having different sources of stochasticity in the postulated real-world model P (as, e.g., in stochastic volatility models where one cannot treat "volatility" as a*

[1] As our notation in Definition 3.1.2 suggests, parameter uncertainty may arise from uncertainty of the *real-world measure* and then transfers to uncertainty of the *risk-neutral measure*.

tradeable asset) or from some kind of stochasticity in the real-world model that may be distorted in many ways (as, e.g., in non-Brownian Lévy models where the processes parameters can be affected by equivalent measure change). Interpreting the Radon–Nikodým derivative $\mathrm{d}Q/\mathrm{d}P$ *of the measure change* $Q \sim P$ *as the "market price of risk parameter", one can also think of incomplete markets as a situation of parameter uncertainty in a natural way.*

An illustrating example for parameter uncertainty can be stated in every model where a hidden parameter vector θ has to be estimated and is not directly quoted by the market (as, e.g., spot rates or interest rates). A simple example not coming from incomplete markets is the famous Black–Scholes model with uncertain Black–Scholes volatility.

Example 3.1.5 (Parameter uncertainty in financial market models)

1. *Examining the risk-neutral version of the* Black–Scholes model*, the dynamics of a stock price follow a geometric Brownian motion, i.e. the SDE*

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t, \quad S_0 > 0,$$

with $(W_t)_{t\geq 0}$ *being Brownian motion,* r *the risk-free interest rate, and* σ *the stock's volatility. While the initial stock price* S_0 *and the risk-free rate* r *are typically market-quoted quantities (and therefore known), one does not have direct information about the volatility* σ*. Hence,* a priori*, every positive number* $\sigma > 0$ *can be taken. Usually, one uses market data (e.g. historical estimation, calibration to market prices) to specify the volatility* σ*.*

2. *In the (risk-neutral)* Heston model*, the stock price dynamics follow the coupled SDEs*

$$\begin{aligned} \mathrm{d}S_t &= rS_t\,\mathrm{d}t + \sigma_t S_t\,\mathrm{d}W_t^{(1)}, \quad S_0 > 0, \\ \mathrm{d}\sigma_t^2 &= \kappa(\sigma_t^2 - \sigma_{\text{long}}^2)\,\mathrm{d}t + \xi\sigma_t\,\mathrm{d}W_t^{(2)}, \quad \sigma_0^2 > 0, \end{aligned}$$

with $(W_t^{(j)})_{t\geq 0}$*,* $j = 1, 2$*, being Brownian motions with correlation* $\rho \in [-1, 1]$*. Compared to the Black–Scholes model, the parameter vector of unknown parameters has a higher dimension. Again, the initial stock price* S_0 *and the risk-free rate* r *are known by market quotation. Furthermore, one may argue that the initial volatility* σ_0 *may be given by some volatility index like the VIX as described in (Guillaume and Schoutens, 2011). On the other hand, the mean reversion speed* $\kappa > 0$*, the long-term volatility* $\sigma_{\text{long}}^2 > 0$*, the vol-of-vol* $\xi > 0$*, and the correlation* $\rho \in [-1, 1]$ *are typically not given and have to be obtained in an estimation or calibration*

procedure. Hence, in the Heston model, one is typically exposed towards parameter uncertainty at least w.r.t. the quadruple $(\kappa, \sigma^2_{\text{long}}, \xi, \rho) \in \mathbb{R}^3_{\geq 0} \times [-1, 1]$.

(Cont, 2006) proposes to handle model uncertainty by a worst case approach: Restricting the set of contingent claims to the vector space

$$\mathcal{C} := \left\{ X \in \bigcap_{Q \in \mathcal{Q}} L^1(Q) : Q \mapsto \mathbb{E}_Q[X] \in \mathcal{L}^\infty(\mathcal{Q}) \right\},$$

one can define an upper bound for the price of a contingent claim $X \in \mathcal{C}$ by $\Gamma^u(X) := \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[X]$ and analoguously a lower bound $\Gamma^l(X) := -\Gamma^u(-X) = \inf_{Q \in \mathcal{Q}} \mathbb{E}_Q[X]$. This definition clearly quantifies the extremes of model (resp. parameter) uncertainty, which is the main purpose of (Cont, 2006).

A special case of (Cont, 2006)'s general setting is the *uncertain volatility model* of (Avellaneda et al., 1995), where a stochastic control approach is employed to calculate the upper and lower price bounds.

Example 3.1.6 (Pricing with uncertain volatility)
Let $(S_t)_{t \geq 0}$ *be an asset price process following a geometric Brownian motion with stochastic volatility, i.e.*

$$\mathrm{d}S_t = rS_t \,\mathrm{d}t + \sigma_t S_t \,\mathrm{d}W_t.$$

The volatility process $(\sigma_t)_{t \geq 0}$ *can be arbitrary, but has its domain in an interval* $[\sigma_l, \sigma_u]$ *with* $\sigma_u > \sigma_l > 0$ *and should be chosen such that the defining stochastic differential equation has a solution. The bounds* σ_u, σ_l *may be obtained from expert judgements or data like, e.g., available implied Black–Scholes volatilities of liquid options. With these implicitly imposed models, one can calculate the model-free upper and lower bounds for the price of an* $\mathcal{F}_T$*-measurable derivative* X *with maturity* T *by a PDE approach (cf. (Avellaneda et al., 1995)) for the price of some derivatives.*

For pricing purposes, the worst-case approach of (Cont, 2006) might be too conservative in practice (this is discussed later): Using $\Gamma^u(X)$ may add a too large charge for model risk. In case of parameter uncertainty, if $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$ and $\Theta \subset \mathbb{R}^m$ is a compact set and the map $\theta \mapsto \mathbb{E}_{Q_\theta}[X]$ is continuous, the supremum and infimum will be attained for certain parameters $\theta_u, \theta_l \in \Theta$, so $\mathbb{E}_{Q_{\theta_u}}[X]$ would be the largest sensible price and $\mathbb{E}_{Q_{\theta_l}}[X]$ would be the lowest sensible price. Hence, a (very conservative) approach to quote a bid-ask price pair for the derivative X could be to quote $\mathbb{E}_{Q_{\theta_l}}[X]$ as a bid price and $\mathbb{E}_{Q_{\theta_u}}[X]$ as an ask price. But, in practice, the market typically does not accept

these conservative prices, e.g. due to competition among different derivatives traders that are willing to take some parameter risk (thus quoting narrower bid-ask prices). Therefore, a worst case approach might not provide good practice in all situations to calculate an additional charge that may be accepted by the market. Furthermore, in case of an available distribution R on $\mathcal{Q}$, the information contained in this distribution R is disregarded. This immediately leads us to introduce a special situation of model uncertainty where more information is given.

Definition 3.1.7 (Model risk)
Let $(\Omega, \mathcal{F}, \mathbb{F}, (S_t)_{t\geq 0}, \mathcal{Q})$ be a financial market model exhibiting model uncertainty, i.e. $|\mathcal{Q}| > 1$. If there is a σ-algebra $\mathcal{F}^{\mathcal{Q}}$ on $\mathcal{Q}$ and a probability measure $R : \mathcal{F}^{\mathcal{Q}} \to [0,1]$ assigning a "likelihood of validity" to the models in doubt $\mathcal{Q}$, the financial market model exhibits model risk.

Analoguously to model risk, one defines parameter risk. A comprehensive example where parameter risk arises naturally is in the case of parameter estimation. Given the example of uncertain volatility in the Black–Scholes model provided in Example 3.1.5, we can try to estimate the Black–Scholes volatility from time series with representative data. In this case, the estimator itself has a distribution given by its pushforward measure and we exhibit parameter risk in the sense of Definition 3.1.7.

Example 3.1.8 (Parameter risk from Black–Scholes volatility estimation)
We consider a Black–Scholes setting as given in Example 3.1.5, where the volatility σ is the key parameter for risk-neutral pricing. This parameter is not directly given by the market (different from the spot price S_0 and the risk-free rate r). Hence, the determination of the volatility is a situation where one is exposed to parameter uncertainty (as described in (Avellaneda et al., 1995)). If the stock price process actually follows the Black–Scholes model, it may be a sensible idea to estimate the volatility from time series data. Under the assumption of normally distributed logarithmic returns $x_1, \dots, x_N$, $x_j = \log S_{t_j+\Delta t} - \log S_{t_j}$, $j = 1, \dots, N$, one may choose the classical estimator for the variance (it may be more convenient to estimate the returns's variance), corrected for the frequency of the data Δt, which results in the estimator

$$\hat{\sigma^2}_N = \frac{1}{\Delta t(N-1)} \sum_{j=1}^{N} (x_j - \bar{x})^2, \quad \bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_j$$

for the variance corresponding to the Black–Scholes volatility. Applying general theory from statistics, one obtains that, under the assumption of having independent normally distributed returns and a true variance $\sigma_0^2 > 0$, the distribution of the estimator is a

chi-squared distribution up to some scaling (cf. (Knight, 2000, Proposition 2.11), i.e. $\hat{\sigma^2}_N \sim \frac{\sigma_0^2}{\Delta t(N-1)}\chi^2_{N-1}$*, which is a Gamma distribution with shape parameter* $(N-1)/2$ *and scale parameter* $2\sigma_0^2/(\Delta t(N-1))$*. Hence, the distribution determining the parameter risk arising from the estimation of volatility (resp. variance) is essentially determined by the chi-squared distribution and the parameter risk triplet* $(\Theta, \mathcal{F}^\Theta, R)$ *is given by* $\Theta = \mathbb{R}_{>0}$*,* $\mathcal{F}^\Theta = \mathcal{B}(\mathbb{R}_{>0})$ *with* $\mathcal{B}(\mathbb{R}_{>0})$ *denoting the Borel-*σ*-algebra w.r.t.* $\mathbb{R}_{>0}$ *and*

$$R(\mathrm{d}x) = \frac{(\Delta t(N-1))^{\frac{N-1}{2}}}{\Gamma\left(\frac{N-1}{2}\right)(2\sigma_0^2)^{\frac{N-1}{2}}} x^{\frac{N-3}{2}} \exp\left(-\frac{x\Delta t(N-1)}{2\sigma_0^2}\right) \mathbb{1}_{\{x>0\}}\,\mathrm{d}x.$$

A special situation where a parameter risk situation is used for explaining the volatility smile in the Black–Scholes model was done by (Lindström, 2010), again in a Black–Scholes setting.

Example 3.1.9 (Smile modeling via volatility risk)
(Lindström, 2010) tries to explain the smile and term structure of implied volatility surfaces by substituting fixed parameters of given models by random variables; adding noise to the parameters. In particular, he suggests substituting the volatility of the Black–Scholes model by a symmetrically distributed noise variable. Generally speaking, a parameter θ *is substituted by a random variable* $\tilde{\theta}$ *with support in the parameter space* Θ*. Then, he calculates the option price serially: First, the parameter* $\theta \in \Theta$ *is fixed and* $\mathbb{E}_Q[X|\tilde{\theta} = \theta]$ *is calculated for each* $\theta \in \Theta$*. Afterwards, the function* $\theta \mapsto \mathbb{E}_Q[X|\tilde{\theta} = \theta]$ *is integrated w.r.t. the distribution of the random variable* $\tilde{\theta}$*, i.e. the price is calculated as*

$$\Gamma(X) = \int_\Theta \mathbb{E}_Q[X|\tilde{\theta} = \theta]\, F_{\tilde{\theta}}(\mathrm{d}\theta).$$

By defining $Q_\theta(A) := \mathbb{E}_Q[\mathbb{1}_A|\tilde{\theta} = \theta]$*, considering the model family* $(Q_\theta)_{\theta\in\Theta}$*, and define the distribution* R *on the parameter space* Θ *to be the pushforward distribution of the random variable* $\tilde{\theta}$*, we can embed this approach into our notion of parameter risk.*

The ansatz from (Lindström, 2010) leads to fairly successful results for smile explanation, but does not account for aversion towards parameter uncertainty, since the risk-neutral price is still calculated. Similar approaches have been advocated (in a discrete setting) by (Bunnin et al., 2000) and (Branger and Schlag, 2004). The ansatz of (Lindström, 2010) relies simply on integrating out the random parameter. Consequently, no additional charge for parameter risk is stipulated. Beyond that, Jensen's inequality shows that for concave functions $\theta \mapsto \mathbb{E}_Q[X|\tilde{\theta} = \theta]$, adding a symmetric noise to the *true* parameter can even lower the price. Furthermore, (Lindström, 2010) assumes that the distribution of

the volatility is normal (which does not even match from the support - the Black–Scholes volatility is assumed to be nonnegative, while the normal distribution has support all over the real line). Hence, the ansatz of (Lindström, 2010) can be criticized from several perspectives.

3.2 Capturing model and parameter risk by using convex risk measures

In this section, we introduce our concept of model (resp. parameter) risk-capturing functionals based on convex risk measures, extending the idea of using convex risk measures for pricing in an incomplete markets setting[2] described in (Carr et al., 2001; Xu, 2006; Cherny and Madan, 2010) and similar to the ideas of measuring model risk by (Branger and Schlag, 2004; Gupta, 2009; Gupta et al., 2010). Although we mainly focus on parameter risk in this thesis, we introduce it in the more general setting of model risk. Furthermore, we show that the pricing proposals from (Cont, 2006) and (Lindström, 2010) are special cases of pricing with certain risk-capturing functionals, e.g. induced by the well-known Average-Value-at-Risk and limit cases of risk-capturing functionals induced by the entropic risk measure.

First, we state required properties of functionals to qualify for capturing model (resp. parameter) risk. These properties can be related to properties of convex risk measures, which we therefore apply to the current situation.

Properties 3.2.1

A functional Γ incorporating model uncertainty w.r.t. the model family $\mathcal{Q}$ into prices should fulfill the following properties:

1. *Order preservation: If there exists a model-free order, it should be preserved when incorporating model uncertainty, i.e.*

$$\forall X, Y \in \bigcap_{Q \in \mathcal{Q}} L^1(Q) : X(\omega) \geq Y(\omega) \text{ for all } \omega \in \Omega \Rightarrow \Gamma(X) \geq \Gamma(Y).$$

2. *Diversification: Diversification of model uncertainty should not be penalized, i.e. a convex combination of two positions facing model uncertainty should not have a higher price than the convex combination of the single prices:*

$$\forall X, Y \in \bigcap_{Q \in \mathcal{Q}} L^1(Q) \, \forall \lambda \in [0,1] : \Gamma(\lambda X + (1-\lambda)Y) \leq \lambda \Gamma(X) + (1-\lambda)\Gamma(Y).$$

[2]Incomplete markets are a special case of model uncertainty, as mentioned in Section 3.1.

3. *Model independence consistency: If a payoff is consistently priced under all models (resp. parameters), no model uncertainty is present and the model uncertainty captured price agrees with the risk-neutral price, i.e. no charge for model risk is added to the risk-neutral price:*

$$\forall X \in \bigcap_{Q \in \mathcal{Q}} L^1(Q) : Q \mapsto \mathbb{E}_Q[X] \text{ is constant on } \mathcal{Q} \Rightarrow \Gamma(X) = \mathbb{E}_Q[X].$$

We now establish the notion of a (model) risk-capturing functional. We define it by treating the more general type of model risk, but it can equally be applied in case of parameter risk.

Definition 3.2.2 (Model risk-capturing functional)
Let $\mathcal{Q}$ be a family of models and let R be a probability measure on $\mathcal{Q}$. Let $\mathcal{A} \subset \mathcal{L}^0(R)$ be a vector space of measureable functions containing the constants and denote

$$\mathcal{C}^{\mathcal{A}} := \left\{ X \in \bigcap_{Q \in \mathcal{Q}} L^1(Q) : Q \mapsto \mathbb{E}_Q[X] \in \mathcal{A} \right\}$$

as the vector space of all $\mathcal{A}$-regular claims being available for all models in the model family $\mathcal{Q}$. Let furthermore $\rho : \mathcal{A} \to \mathbb{R}$ be a normalized, law invariant convex risk measure. Then the mapping $\Gamma : \mathcal{C}^{\mathcal{A}} \to \mathbb{R}$ defined by

$$\Gamma(X) := \rho(Q \mapsto \mathbb{E}_Q[X]) \tag{3.1}$$

is called a (model) risk-capturing functional *on the set of claims $\mathcal{C}^{\mathcal{A}}$ w.r.t. the distribution R. ρ is called the* generator of Γ *and $\Gamma(X)$ is called the* risk-captured (ask) price *of X given the functional Γ.*

The idea behind Definition 3.2.2 is that a derivatives trader, facing model risk, should choose ask prices that are high enough that some buffer prevents her from losses due to model risk, which may occur when selling too cheap. Conversely, one should account for model risk when buying derivatives by setting bid prices low enough in order to prevent oneself from losses due to model risk. A consistent strategy is to regard the dual $-\Gamma(-X)$ as the risk-captured bid price. Due to the dual nature of the bid and ask price and for brevity, we will use the name "risk-captured price" for the risk-captured ask price in the further, since the discussed properties transfer to the bid price in a dual manner. Since

$$0 = \Gamma(0) = \Gamma((X + (-X))/2) \leq (\Gamma(X) + \Gamma(-X))/2$$

holds for every $X \in \mathcal{C}^{\mathcal{A}}$ (cf. (Bion-Nadal, 2009)), the risk-captured bid price is always lower or equal than the risk-captured ask price.

Canonical choices for $\mathcal{A}$ are, e.g., the Lebesgue spaces $L^p(R)$ for $p \in [1, \infty]$, depending on the domain of ρ. The assumption that constants are included in $\mathcal{A}$ is important for the valuation of model-invariant payoffs. Furthermore, ρ being normalized is also natural from an economic point of view, since the risk-captured price of a zero payoff should equal 0.

Obviously, parameter risk can also be incorporated into this framework: Given a set of equivalent martingale measures $(Q_\theta)_{\theta \in \Theta}$ and a distribution R on Θ, the bijection $\iota : \Theta \to \mathcal{Q}$ with $\mathcal{Q} := \{Q_\theta : \theta \in \Theta\}$ defines a distribution on $\mathcal{Q}$. When $\mathcal{Q}$ can be parameterized by some parameter space Θ, we speak about parameter risk-capturing functionals. In case of parameter risk w.r.t. a parameter space Θ, we abbreviate $\mathbb{E}_\theta[X] := \mathbb{E}_{Q_\theta}[X]$ for $\theta \in \Theta$.

Similar ideas, in a less technical context, were first provided by (Branger and Schlag, 2004) in a discrete setting employing spectral risk measures. (Gupta, 2009; Gupta et al., 2010) suggest measuring model risk with convex risk measures, omitting to require law invariance and normalization and without the interpretation of a risk-captured price.

The following proposition subsumes some basic properties of (model) risk-capturing functionals and the relationship with their generators.

Proposition 3.2.3 (Properties of risk-capturing functionals)
If Γ is a risk-capturing functional as in Definition 3.2.2 with domain $\mathcal{C}^{\mathcal{A}}$ and generator ρ, then the following holds:

1. *Γ is a convex risk measure on $\mathcal{C}^{\mathcal{A}}$ and is translation invariant on the subspace of model-invariant contingent claims $\mathcal{X}_0 := \{X \in \mathcal{C}^{\mathcal{A}} : Q \mapsto \mathbb{E}_Q[X] \text{ is constant}\}$ w.r.t. the (well-defined) linear form $\pi(X) := \mathbb{E}_Q[X], X \in \mathcal{X}_0$ (in particular, Γ fulfills the desirable properties described in 3.2.1).*
2. *If ρ is a coherent risk measure, then Γ is also coherent.*
3. *If ρ carries the Fatou property, then Γ also carries the Fatou property.*
4. *If $\mathcal{A} = L^p(R)$, $p \in [1, \infty]$, or $\mathcal{A} = \mathcal{C}(\mathcal{Q})$ with a compact space of possible probability measures $\mathcal{Q}$, then the Fatou property holds automatically.*
5. *Γ can be represented in terms of positive linear forms extending π, i.e.*

$$\Gamma(X) = \max_{\lambda \in (\mathcal{C}^{\mathcal{A}})_+^{\star \pi}} \lambda(X) - \alpha(\lambda)$$

holds with $(\mathcal{C}^{\mathcal{A}})_+^{\star\pi}$ *denoting the set of all positive linear forms extending* π *and* $\alpha : (\mathcal{C}^{\mathcal{A}})_+^{\star\pi} \to (-\infty, \infty]$ *being a suitable penalty function.*

Proof

1. *Let* $X, Y \in \mathcal{C}^{\mathcal{A}}$*. If* $X \leq Y$*, then the monotonicity of* ρ *and of expected values immediately yield*

$$\Gamma(X) = \rho(\mathbb{E}_\cdot[X]) \leq \rho(\mathbb{E}_\cdot[Y])) = \Gamma(Y).$$

Let furthermore $\lambda \in [0,1]$*, then the convexity of* ρ *and linearity of expected values yield*

$$\begin{aligned}\Gamma(\lambda X + (1-\lambda)Y) &= \rho(\mathbb{E}_\cdot[\lambda X + (1-\lambda)Y]) = \rho(\lambda \mathbb{E}_\cdot[X] + (1-\lambda)\mathbb{E}_\cdot[Y]) \\ &\leq \lambda\rho(\mathbb{E}_\cdot[X]) + (1-\lambda)\rho(\mathbb{E}_\cdot[Y]) = \lambda\Gamma(X) + (1-\lambda)\Gamma(Y).\end{aligned}$$

Let now $Y \in \mathcal{X}_0$*. Then the mapping* $Q \mapsto \mathbb{E}_Q[Y] =: c$ *is constant. Hence, the cash invariance of* ρ *delivers*

$$\Gamma(X+Y) = \rho(\mathbb{E}_\cdot[X+Y]) = \rho(\mathbb{E}_\cdot[X] + c) = \rho(\mathbb{E}_\cdot[X]) + c = \Gamma(X) + c.$$

2. *Let* $c > 0$*,* $X \in \mathcal{C}^{\mathcal{A}}$*, and* ρ *be coherent. Then positive homogeneity of* ρ *and the expected value immediately yield*

$$\Gamma(cX) = \rho(\mathbb{E}_\cdot[cX]) = \rho(c\mathbb{E}_\cdot[X]) = c\rho(\mathbb{E}_\cdot[X]) = c\Gamma(X).$$

3. *Let* $(X_N)_{N\in\mathbb{N}}$ *be a pointwise convergent, uniformly bounded sequence of contingent claims from* $\mathcal{C}^{\mathcal{A}}$ *with* $X := \lim_{N\to\infty} X_N$ *and let* ρ *carry the Fatou property. Since* $(X_N)_{N\in\mathbb{N}}$ *is uniformly bounded, there is a* $C > 0$ *such that* $|X_N| \leq C$ *for all* $N \in \mathbb{N}$*. Hence, Lebesgue's dominated convergence theorem delivers* $\mathbb{E}_Q[X] = \lim_{N\to\infty} \mathbb{E}_Q[X_N]$ *for every* $Q \in \mathcal{Q}$*. Thus, the functions* $\nu_N : \mathcal{Q} \to \mathbb{R}$*,* $\nu_N(Q) := \mathbb{E}_Q[X_N]$*, converge pointwise to the function* $\nu : \mathcal{Q} \to \mathbb{R}$ *defined by* $\nu(Q) := \mathbb{E}_Q[X]$*. Furthermore, the sequence* $(\nu_N)_{N\in\mathbb{N}}$ *is uniformly bounded, since*

$$|\nu_N(Q)| = |\mathbb{E}_Q[X_N]| \leq \mathbb{E}_Q[|X_N|] \leq \mathbb{E}_Q[C] = C$$

hold for every $Q \in \mathcal{Q}$ *and for every* $N \in \mathbb{N}$*. Now the Fatou property of* ρ *immediately yields*

$$\Gamma(X) = \rho(\nu) \leq \liminf_{N\in\mathbb{N}} \rho(\nu_N) = \liminf_{N\in\mathbb{N}} \Gamma(X_N),$$

which is exactly the Fatou property of Γ*.*

4. *If all mappings $\nu_X : Q \mapsto \mathbb{E}_Q[X]$ are p-integrable for $p \in [1, \infty)$, and ρ is a convex risk measure on $L^p(R)$, ρ automatically carries the Fatou property, since the dual space of $L^p(R)$ can be identified with a set of signed measures $\{\tilde{R} \ll R\}$ and the argument presented in (Krätschmer, 2006). Similar, in case of $\mathrm{dom}(\rho) = \mathcal{C}(\mathcal{Q})$ for compact $\mathcal{Q}$, the dual space of $\mathcal{C}(\mathcal{Q})$ can be identified by a vector space of signed Borel measures on $\mathcal{Q}$ via the Riesz representation theorem (cf. (Werner, 2011, Theorem II.2.5)), thus ρ automatically carries the Fatou property. For $\mathrm{dom}(\rho) = \mathcal{L}^\infty(R)$, the results from (Jouini et al., 2006) yield that ρ has the Fatou property due to its law-invariance. Hence, (iii) delivers the Fatou property for the risk-capturing functional Γ.*

5. *Follows directly from the representation in Theorem 2.1.3.* □

If Γ carries the Fatou property, the practical interpretation is as follows: If the risk-captured price $\Gamma(X)$ of a contingent claim X is difficult to calculated (e.g. due to high computational effort), but may be (pointwise) approximated by a (uniformly bounded) sequence of contingent claims $(X_N)_{N\in\mathbb{N}}$ where the risk-captured price $\Gamma(X_N)$ can be calculated easier, the limit inferior of the risk-captured prices $\Gamma(X_N)_{N\in\mathbb{N}}$ can be used as a conservative estimate; never understating the risk-captured price for X. This is actually a desirable property, since an approximation of a claim X by some sequence $(X_N)_{N\in\mathbb{N}}$ will not lead to risk-captured pricing understating the real risk.

An example is the approximation of a continuous arithmetic Asian option by a sequence of Asian options with discrete observation, where the times of observation increase.

Example 3.2.4 (Approximation of a continuous arithmetic Asian option) *Let $(S_t)_{t\geq 0}$ be an asset price process and X be an arithmetic Asian call option with continuous observation, i.e.*

$$X = e^{-rT}\left(\frac{1}{T}\int_0^T S_t \,\mathrm{d}t - K\right)^+$$

for some maturity $T > 0$, some risk-free rate r, and some strike $K > 0$. In practice, continuous observation is impossible. Thus, for valuation purposes (which is typically done by means of Monte Carlo methods), one often approximates continuous Asian options by Asian options with discrete observations, i.e.

$$X_N = e^{-rT}\left(\frac{1}{N}\sum_{j=1}^{N} S_{t_j} - K\right)^+, \quad N \in \mathbb{N},$$

on a fine grid $(t_1, \ldots, t_N)$ *with* $t_N = T$*. It has been shown that* $X_N \to X$ *(if the density of the grid* $(t_1, \ldots, t_N)$ *tends to zero) in a pointwise sense for good-natured models (cf. (Kemna and Vorst, 1990)). Furthermore, from a practical point of view, a maximum payoff of some* $C > 0$ *is realistic in practice and a common assumption for numerical valuation of options (as, e.g., in option valuation via numerical PDE solving). Hence, the Fatou property guarantees that for a continuous Asian option* X *approximated with discrete Asian options* $(X_N)_N \in \mathbb{N}$*, provided that the number of observations* N *is large enough, the risk-captured price of the approximation* $\Gamma(X_N)$ *is not lower than the risk-captured true price* $\Gamma(X)$ *due to* $\Gamma(X) \leq \liminf_{N\in\mathbb{N}} \Gamma(X_N)$*.*

The choice of the convex risk measure ρ (and therefore implicitly its domain $\mathcal{A}$) completely determines the risk-capturing functional Γ. This allows us to exploit the rich pool of convex risk measures fulfilling law-invariance and normalization to calculate risk-captured prices.

3.3 Examples: AVaR- and entropic-driven risk-captured prices

The Average-Value-at-Risk, which was introduced in Example 2.1.2, is a coherent law-invariant risk measure and therefore fulfills the properties to serve as a generator of a risk-capturing functional. Hence, we introduce the AVaR-driven risk-capturing functional.

Example 3.3.1 (Average-Value-at-Risk-induced risk-capturing functional) *Let* $\mathcal{Q}$ *be a family of martingale measures inducing model risk and let* R *be a distribution on* $\mathcal{Q}$*. Consider the* $\mathcal{L}^1(R)$*-regular claims and* $R * \mathrm{AVaR}_\alpha : \mathcal{C}^{\mathcal{L}^1(R)} \to \mathbb{R}$ *to be the risk-capturing functional generated by the coherent risk measure* $\mathrm{AVaR}_\alpha : \mathcal{L}^1(R) \to \mathbb{R}$ *for a given confidence level* $\alpha \in (0, 1]$*, so*

$$R * \mathrm{AVaR}_\alpha(X) := \mathrm{AVaR}_\alpha(Q \mapsto \mathbb{E}_Q[X]).$$

We call $R * \mathrm{AVaR}_\alpha$ *the R-Average-Value-at-Risk, since it captures the risk arising from the uncertainty on the model family* $\mathcal{Q}$ *by the distribution* R *isolated from risk within a specific model* $Q \in \mathcal{Q}$*.* $R * \mathrm{AVaR}_\alpha$ *deals with the risk from the upper* α*-tail of the distribution of the prices by averaging over the tail prices dependent on* α*. It should* not *be confused with the regular Average-Value-at-Risk of a single model, which captures the risk being described in a specific model.*

The R-Average-Value-at-Risk is a generalization of both (Cont, 2006) and (Lindström, 2010): First start with the approach discussed in (Lindström, 2010) to deal with parameter uncertainty. The existence of a single martingale measure Q is assumed. Furthermore, the parameter θ bearing uncertainty follows a distribution Q^θ. In our terminology, (Lindström, 2010) proposes to calculate $Q^\theta * \mathrm{AVaR}_1$, which can be easily seen by

$$\begin{aligned}\int_\Theta \mathbb{E}_Q[X|\theta=\theta_0]\,Q^\theta(\mathrm{d}\theta_0) &= \int_\Theta \mathbb{E}_{Q_{\theta_0}}[X]Q^\theta(\mathrm{d}\theta_0) = \int_0^1 q^{Q^\theta}_{\mathbb{E}_{Q.}[X]}(\beta)\,\mathrm{d}\beta \\ &= \int_0^1 \mathrm{VaR}_\beta(\mathbb{E}_{Q.}[X])\,\mathrm{d}\beta = \mathrm{AVaR}_1(\mathbb{E}_{Q.}[X]) = Q^\theta * \mathrm{AVaR}_1(X).\end{aligned}$$

The approach of (Cont, 2006) is to calculate the supremum of the expectations w.r.t. a given family of models $\mathcal{Q}$. In (Cont, 2006) the availability of a distribution is not assumed, but in case of an at most countable $\mathcal{Q}$, we can create a dummy distribution R on the power set of $\mathcal{Q}$ with $R(Q) > 0$ for each $Q \in \mathcal{Q}$. Furthermore, as described in Section 2.1, one can interpret AVaR_0 *as the essential supremum. By choosing a discrete dummy distribution R on $\mathcal{Q}$, the essential supremum turns out to become a regular supremum and therefore $R * \mathrm{AVaR}_0(X) = \sup_{Q\in\mathcal{Q}} \mathbb{E}_Q[X]$ holds. Therefore, when there is uncertainty about at most countably many models, the suggested worst-case ansatz from (Cont, 2006) agrees with $R * \mathrm{AVaR}_0$.*

This shows that in (Cont, 2006) and (Lindström, 2010) it is suggested to use the extreme points *of $R*\mathrm{AVaR}$, so the notion $R*\mathrm{AVaR}$ is a natural extension to interpolate between these approaches and therefore provides prices with extra charges for uncertainty, being more conservative than an expected value and less conservative than a supremum.*

Remark 3.3.2 (Quantile property of Average-Value-at-Risk)
In Section 2.1, it has been shown that the Average-Value-at-Risk dominates the Value-at-Risk. Translating this into the notion of R-Average-Value-at-Risks, it provides that the risk-captured price induced by the R-Average-Value-at-Risk to a given significance level $\alpha \in [0,1]$ does not understate the fair price with probability $1-\alpha$, since it is more conservative than the upper α-quantile of $\mathbb{E}.[X]$ w.r.t. the distribution R.

In presence of a distribution R on $\mathcal{Q}$, the $R*\mathrm{AVaR}_\alpha$ for a given significance level provides an *objective* risk-captured price, since the only preference will be set by specifying the significance level. A risk-neutral preference would be the expected value suggested in (Lindström, 2010), the most conservative choice is the (essential) supremum as in (Cont, 2006). The next example provides a more subjective view on capturing model risk.

Example 3.3.3 (Entropic-induced risk-capturing functional)
Another famous example for a convex risk measure is the entropic risk measure: If $(\Omega, \mathcal{F}, P)$ *is a probability space and* $X \in \mathcal{L}^\infty(P)$*, the* entropic risk measure *with* risk aversion parameter $\lambda \in (0, \infty)$ *is defined as*

$$\rho_\lambda^{\text{ent}}(X) := \frac{1}{\lambda} \log(\mathbb{E}_P[\exp(\lambda X)]).$$

It is well known that the entropic risk measure is normalized and law-invariant (but not coherent), see the treatment in (Föllmer and Schied, 2004, Example 4.33). We examine (in presence of a distribution R *on* $\mathcal{Q}$*) the resulting* R*-entropic risk-capturing functional generated by the entropic risk measure and denote it by* $\Gamma_\lambda^{\text{ent}}$ *for* $\lambda \in (0, \infty)$*. To ensure that* $\Gamma_\lambda^{\text{ent}}$ *is well-defined, we restrict the set of evaluable claims to* $\mathcal{C}^{\mathcal{L}^\infty}$*.*

Remark 3.3.4 (Properties of the entropic-induced risk-capturing functional)

- *Since for* $\lambda \in (0, \infty)$*,* $x \mapsto \exp(\lambda x)$ *is a convex function, Jensen's inequality provides*

$$\begin{aligned}\Gamma_\lambda^{\text{ent}}(X) = \rho^{\text{ent}}(\mathbb{E}_\cdot[X]) &= \frac{1}{\lambda} \log(\mathbb{E}_R[\exp(\lambda \mathbb{E}_\cdot[X])]) \\ &\geq \frac{1}{\lambda} \log(\exp(\lambda \mathbb{E}_R[\mathbb{E}_\cdot[X]])) = \mathbb{E}_R[\mathbb{E}_\cdot[X]],\end{aligned}$$

  and, therefore, we obtain a more conservative risk-capturing functional compared to the expected value w.r.t. R*, which is the approach of (Lindström, 2010). It is known (see below) that the extremal cases for* $\lambda \searrow 0$ *and* $\lambda \nearrow \infty$ *agree again with the risk-capturing functionals generated by the expected value and the essential supremum.*

- *Jensen's inequality also sheds some light on the role of* λ*: Heuristically speaking, the higher* $\lambda \in (0, \infty)$*, the "more convex" the function* $x \mapsto \exp(\lambda x)$ *and therefore the inequality in the above computation becomes "sharper". Thus, the higher the parameter* λ *is chosen, the more conservative the bid-ask pricing will be.*

- *An interesting feature of the entropic risk measure is that it is not positively homogeneous (unlike* AVaR*). Therefore, it may account for the risk being associated with large trades, reflecting that large trades may bear additional risk due to liquidity effects, risk management issues, or external regulatory constraints.*

Similar to the Average-Value-at-Risk, it holds for the entropic risk measure that its extremal points are the essential supremum and the expected value. In a general probability

space $(\Omega, \mathcal{F}, P)$, the entropic risk measure fulfills

$$\begin{aligned}\rho_\lambda^{\text{ent}}(X) &= \frac{1}{\lambda}\log(\mathbb{E}_P[\exp(\lambda X)]) \\ &= \log\left(\left(\mathbb{E}_P[\exp(X)^\lambda]\right)^{\frac{1}{\lambda}}\right) = \log\|\exp(X)\|_\lambda\end{aligned}$$

for $\lambda > 0$ and $X \in \mathcal{L}^\infty(P)$ with $\|X\|_\lambda = \mathbb{E}_P[X^\lambda]^{1/\lambda}$. Thus, classical convergence theorems (e.g. (Hardy et al., 1934)) yield

$$\begin{aligned}\lim_{\lambda\to 0}\rho_\lambda^{\text{ent}}(X) &= \lim_{\lambda\to 0}\log\|\exp(X)\|_\lambda \\ &= \mathbb{E}_P[\log(\exp(X))] = \mathbb{E}_P[X]\end{aligned}$$

and

$$\begin{aligned}\lim_{\lambda\to\infty}\rho_\lambda^{\text{ent}}(X) &= \lim_{\lambda\to\infty}\log\|\exp(X)\|_\lambda \\ &= \log(\operatorname{ess\,sup}(\exp(X))) = \operatorname{ess\,sup} X.\end{aligned}$$

Therefore, the entropic-driven risk-capturing functional is an alternative generalization of the ansatzes of (Cont, 2006) and (Lindström, 2010).

Parameter risk-captured prices as bid-ask prices

The fundamental idea for the application of parameter risk-capturing functionals to derive bid-ask prices is that higher parameter risk (i.e. a distribution on the parameters yielding a more dispersed price distribution) should result in a wider bid-ask spread. A trader being unsure about the true value of a parameter, or having problems with a position matching his book, usually adds a premium to the given theoretical model price (the systematics behind this are often called *edge rules*). In case of parameter risk, i.e. the quantification of the degree of parameter uncertainty by a distribution, the risk-captured price provides a systematic approach for traders to account for parameter risk. In case of stationary market conditions and the estimation of model parameters being necessary (for example because liquid market data for a calibration is not available), the estimator's distribution delivers useful information how wide a reasonable bid-ask spread could be. The choice of risk-capturing functional (as examples, the AVaR- and the entropic-risk-capturing functional were presented) reflects the subjective preferences of the trader (e.g. parameter risk aversion, aversion to large trades). When $\Gamma : \mathcal{C} \to \mathbb{R}$ denotes a risk-capturing functional on a suitable vector space of contingent claims $\mathcal{C}$, for $X \in \mathcal{C}$ we suggest to treat $\Gamma(X)$ as the ask price and $\bar{\Gamma}(X) := -\Gamma(-X)$ as the bid price for the contingent claim X, as described in Section 3.2.

3.4 Market-implied distributions

In case of liquid markets, some parameters are readily given by market quotes (e.g. liquid stocks), so these parameters are not exposed to uncertainty. Hence, bid-ask spreads are observable and can therefore be considered as market-implied. With available data from liquid markets, we derive a market-implied distribution on the parameter set and recover bid-ask prices via risk-capturing functionals, given this "synthetic parameter risk" situation.

We start with a simple example, showing that the parameter risk-captured pricing approach of explaining bid-ask spreads is compatible to plug-in bid-ask spreads from liquid quote-driven markets.

Example 3.4.1 (Bid-ask pricing from quoted parameters)
If the parameter is quoted in a liquid price-driven market (like spot prices, volatilities of liquid stocks, or exchange rates), a bid-ask quote is readily available. From our framework's point of view, this bid-ask quote can be considered as a two-point distribution on the parameter space: If $\Theta \subset \mathbb{R}$ is a one-dimensional parameter space and we have a quoted pair $(\theta_{\text{bid}}, \theta_{\text{ask}}) \in \Theta \times \Theta$, we define the implied parameter distribution on Θ by $R := 0.5\delta_{\theta_{\text{bid}}} + 0.5\delta_{\theta_{\text{ask}}}$, denoting by δ_θ the Dirac measure w.r.t. $\theta \in \Theta$. If X is a contingent claim, denoting the risk-neutral price w.r.t. $\theta \in \Theta$ by $f(\theta) := E_\theta[X]$, the quantile function of f w.r.t. the probability measure R can be easily calculated (without loss of generality we assume $f(\theta_{\text{bid}}) < f(\theta_{\text{ask}})$)

$$q_f(\alpha) = \begin{cases} f(\theta_{\text{bid}}), & \text{if } \alpha \leq 0.5, \\ f(\theta_{\text{ask}}), & \text{otherwise,} \end{cases}$$

and, therefore, $R * \text{AVaR}_\alpha(X)$ coincides for $\alpha \leq 0.5$ with the simple plug-in ask price. The expected value does not represent the plug-in mid price (defining the mid parameter as the arithmetic mean of bid and ask), but the mean of the plug-in bid and ask prices, i.e. $R * \text{AVaR}_\alpha$ for $\alpha \in (0.5, 1)$ leads to a weighted average of the plug-in bid and ask prices, weighting the plug-in ask price higher than the bid price. The entropic-driven risk measure, being interpreted as more subjective*, results in an ask price between the plug-in prices as well, weighted by its utility.*

Another, slightly more complicated application for the calculation of bid-ask spreads is an order-driven market. If the parameter θ is not directly quoted with bid and ask price, but is traded in a liquid order-driven market (as many stocks do, e.g. in Xetra trading), bid and ask can be recovered by risk-capturing functionals, defining the market-implied distribution by the order book.

Example 3.4.2 (Recovering bid-ask spreads from order-driven markets)
Let $\Theta \subset \mathbb{R}$ be a one-dimensional parameter space and let $(\theta_1, A_1), \ldots, (\theta_M, A_M) \in \Theta \times \mathbb{R}^+$ for $M \in \mathbb{N}$ be a snapshot of the order book (e.g. the top M positions from bid and ask) with parameter values $\theta_1, \ldots, \theta_M$ and associated quantities $A_1, \ldots, A_M$. If we assume that the order book gives a useful prediction for a bid-ask price (which does not exist naturally in an order-driven market), we could define a discrete probability measure on Θ by setting $\bar{A} := \sum_{i=1}^{M} A_i$ and

$$R := \frac{\sum_{i=1}^{M} A_i \delta_{\theta_i}}{\bar{A}}$$

to be the distribution based on which the risk-captured price has to be calculated.

4 Convergence properties of risk-captured prices

In Section 3, we defined a risk-capturing functional given a distribution R on the set of possible models $\mathcal{Q}$ (resp. on the set of possible parameters Θ). To specify the distribution R, there may be numerous possibilities (e.g. the pushforward measure of an estimator), but in some cases, one might only be able to approximate a distribution R with some sequence of distributions $(R_N)_{N\in\mathbb{N}}$. One desirable property of a risk-capturing functional is that the risk-captured price w.r.t. the sequence of distributions $(R_N)_{N\in\mathbb{N}}$ eventually converges to the risk-captured price w.r.t. the limit distribution R. In this section, we scrutinize convergence properties and provide convergence results for some classes of risk-capturing functionals.

4.1 Convergence results of risk-capturing functionals

We postulate the following convergence property for risk-capturing functionals based on weak convergence of probability measures.

Definition 4.1.1 (Convergence property (CP))
Let Θ be a parameter space and $R_N \to R_0$, $N \to \infty$, be a sequence of weakly convergent distributions on Θ. Let $\mathcal{A} \subset \mathcal{L}^0(\Theta)$ be a vector space of measurable functions and $(\rho_N)_{N\in\mathbb{N}}$ be a sequence of convex risk measures on $\mathcal{A}$ such that ρ_N is R_N-law invariant. The sequence of risk-capturing functionals $\Gamma_N : \mathcal{C}^{\mathcal{A}} \to \mathbb{R}$ with

$$\Gamma_N(X) := \rho_N(\theta \mapsto \mathbb{E}_\theta[X])$$

is said to have the convergence property (CP) *on $\mathcal{A}$, iff*

$$\lim_{N\to\infty} \Gamma_N(X) =: \Gamma_0(X) = \rho_0(\theta \mapsto \mathbb{E}_\theta[X])$$

converges for every $X \in \mathcal{C}^{\mathcal{A}}$ and Γ_0 is a risk-capturing functional for the distribution R_0 with generator ρ_0.

Intuitively, the convergence property (CP) describes that a sequence of risk-capturing functionals is consistent with approximations of the parameter's distribution in the sense of weak convergence. Typically, one employs the same convex risk measure $\rho \equiv \rho_N$ (e.g. the AVaR), while varying the parameter distributions $(R_N)_{N\in\mathbb{N}}$, eventually converging to a limit distribution R_0.

At first glance, the convergence property (CP) seems to be abstract and technical, but the special case when the limit distribution is a Dirac distribution (as in case of, e.g., consistent estimators, which will be discussed in Chapter 5) sheds some light on its usefulness.

Proposition 4.1.2 (Convergence to plug-in price)
Let $(R_N)_{N\in\mathbb{N}}$ be a sequence of probability distributions on the parameter set Θ converging to a limit distribution R_0 in the weak sense and $(\Gamma_N)_{N\in\mathbb{N}}$ be a sequence of risk-capturing functionals. If the sequence $(\Gamma_N)_{N\in\mathbb{N}}$ carries the convergence property (CP) on $\mathcal{A}$ and $R_0 = \delta_{\theta_0}$, i.e. R_0 is a Dirac distribution w.r.t. a parameter θ_0, then $\Gamma_N(X) \to \mathbb{E}_{\theta_0}[X]$ holds for every $X \in \mathcal{C}^{\mathcal{A}}$, $N \to \infty$.

Proof
We calculate the risk-captured price given a Dirac distribution w.r.t. $\theta_0 \in \Theta$. Under the Dirac measure δ_{θ_0}, the law of $\theta \mapsto \mathbb{E}_\theta[X]$ is equal to the law of the constant function $\theta \mapsto \mathbb{E}_{\theta_0}[X]$. Since we require the limit risk-capturing functional Γ_0, $\Gamma_0(X) := \lim_{N\to\infty} \Gamma_N(X)$, $X \in \mathcal{C}^{\mathcal{A}}$, to be law-invariant w.r.t. the Dirac measure δ_{θ_0}, the claim is established. □

In particular, this means that if a distribution carries lots of information about the trustworthyness of parameters and clusters around a parameter θ_0, the risk-capturing price – provided the class of risk-capturing functionals carries the convergence property (CP) – does not differ too much from the plug-in price w.r.t. the parameter θ_0 where it clusters.

In the further, we investigate which classes of risk-capturing functionals (or, equivalently, which classes of convex risk measures generating the risk-capturing functionals) fulfill the convergence property (CP) on which domain. We start with the AVaR-generated risk-capturing functionals, which fulfill the convergence property provided the function $\theta \mapsto \mathbb{E}_\theta[X]$ is bounded and continuous.

Proposition 4.1.3 (AVaR-induced risk-capturing functionals fulfill (CP))
Let $\Theta \subset \mathbb{R}^m$ be a Euclidean parameter space. Let $R_N \to R_0$ be a weakly convergent sequence of probability distributions on Θ and $\alpha \in (0, 1]$. Then $R_N * \mathrm{AVaR}_\alpha(X) \to$

$R_0 * \mathrm{AVaR}_\alpha(X)$, $N \to \infty$, for all X such that $\theta \mapsto \mathbb{E}_\theta[X]$ is continuous and bounded, so the sequence $(R_N * \mathrm{AVaR}_\alpha)_{N\in\mathbb{N}}$ fulfills the convergence property (CP) on $\mathcal{C}^b(\Theta)$.

Proof

Let $X \in \mathcal{C}^{\mathcal{C}^b(\Theta)}$ and $\alpha \in (0,1]$ be arbitrary, we abbreviate $f(\theta) := \mathbb{E}_\theta[X]$ for ease of notation. We note that a well-known property of weak convergence is that it transfers to pushforward measures of continuous functions (cf. (Bartoszynski, 1961)), so for every continuous function g and $x \in \mathbb{R}$, $F_{g,R_N}(x) \to F_{g,R_0}(x)$, $N \to \infty$, with $F_{g,S}(x) := S(g \leq x)$ denoting the distribution function of g w.r.t. a probability measure S. Furthermore, since the quantile function is the quasi-inverse of the distribution function, it follows $q_f^{R_N}(\beta) \to q_f^{R_0}(\beta)$ for Lebesgue-almost every $\beta \in (0,1)$ (cf. (Denneberg, 1994, p. 46)). Applying Lebesgue's dominated convergence theorem immediately yields

$$\begin{aligned} R_N * \mathrm{AVaR}_\alpha(X) &= \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta^{R_N}(\mathbb{E}_\cdot[X])\,\mathrm{d}\beta = \frac{1}{\alpha}\int_0^\alpha q^{R_N}_{\mathbb{E}_\cdot[X]}(1-\beta)\,\mathrm{d}\beta \\ &\xrightarrow{N\to\infty} \frac{1}{\alpha}\int_0^\alpha q^{R_0}_{\mathbb{E}_\cdot[X]}(1-\beta)\,\mathrm{d}\beta = R_0 * \mathrm{AVaR}_\alpha(X). \qquad \square \end{aligned}$$

This theorem is rather basic and only states that the convergence property (CP) holds for a very specialized class of risk-capturing functionals. But since the family of convex risk measures $(\mathrm{AVaR}_\alpha)_{\alpha\in(0,1]}$ "constructs" the much broader class of spectral risk measures by a relatively straight-forward construction (cf. Proposition 2.1.6), the result in Proposition 4.1.3 can easily be extended to all spectral risk measures.

Corollary 4.1.4 (Spectral-induced risk-capturing functionals fulfill (CP))

Let $\Theta \subset \mathbb{R}^m$ be a Euclidean parameter space. Let $R_N \to R_0$ be a weakly convergent sequence of probability distributions on Θ. Then for every spectrum $\phi : [0,1] \to \mathbb{R}_{\geq 0}$ satisfying $\phi(0) = \lim_{y\downarrow 0}\phi(y)$, the spectral risk measure-induced risk-capturing functionals fulfill the convergence property (CP) on $\mathcal{C}^b(\Theta)$, i.e. if we define

$$\Gamma_\phi^S(X) := \int_0^1 \mathrm{VaR}_\alpha^S(\mathbb{E}_\cdot[X])\phi(\alpha)\,\mathrm{d}\alpha$$

for a probability distribution S on Θ, we obtain $\Gamma_\phi^{R_N}(X) \to \Gamma_\phi^{R_0}(X)$ for all $X \in \mathcal{C}^{\mathcal{C}^b(\Theta)}$, $N \to \infty$.

Proof

Applying Proposition 2.1.6, the spectral risk measure ρ_ϕ can be represented by a Borel measure μ on $(0,1]$ with

$$\rho_\phi(X) = \int_0^1 \mathrm{AVaR}_\alpha(X)\,\mu(\mathrm{d}\alpha).$$

Hence, it follows by Proposition 4.1.3 and dominated convergence

$$\begin{aligned}\Gamma_{\phi}^{R_N}(X) &= \int_0^1 R_N * \mathrm{AVaR}_\alpha(X)\,\mu(\mathrm{d}\alpha) \\ &\xrightarrow{N\to\infty} \int_0^1 R_0 * \mathrm{AVaR}_\alpha(X)\,\mu(\mathrm{d}\alpha) = \Gamma_{\phi}^{R_0}(X). \qquad \square\end{aligned}$$

As an immediate corollary we obtain that in case of a compact parameter space, the assumption of boundedness can be dropped.

Corollary 4.1.5
Let the assumptions of Corollary 4.1.4 hold and Θ be additionally compact. Then the spectral risk measure-induced risk-capturing functionals fulfill the convergence property (CP) on $\mathcal{C}(\Theta)$.

Proof
Since Θ is compact, $\mathcal{C}(\Theta) = \mathcal{C}^b(\Theta)$. Hence, Corollary 4.1.4 yields the assertion. $\square$

Scrutinizing the proofs above, our results can be regarded in the light of the so-called portmanteau theorem (cf. (Billingsley, 2009, Theorem 2.1)) and we obtain a new characterization of weak convergence.

Corollary 4.1.6 (Portmanteau theorem, spectral risk measure version)
Let Ω be a separable metrizable space and denote the Borel σ-algebra on Ω by $\mathcal{B}$. Furthermore, denote by $\mathcal{P}(\mathcal{B})$ the set of all probability measures on $\mathcal{B}$. Let $(P_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}$. Now the following equivalence holds:

1. *$(P_n)_{n\in\mathbb{N}}$ converges weakly to some limit $P \in \mathcal{P}$.*
2. *For all bounded, continuous random variables X, $\mathbb{E}_{P_n}[X] \to \mathbb{E}_P[X]$ for $n \to \infty$.*
3. *For all bounded, continuous random variables X and spectral risk measures $\rho_n^\phi : \mathcal{L}^\infty(P_n) \to \mathbb{R}$ defined by*

$$\rho_n^\phi(X) := \int_0^1 \mathrm{VaR}_\alpha(X; P_n)\phi(\alpha)\,\mathrm{d}\alpha,$$

$\rho_n^\phi(X) \to \rho^\phi(X)$ for $n \to \infty$, while ρ denotes the spectral risk measure on $\mathcal{L}^\infty(P)$ defined by

$$\rho_\phi(X) := \int_0^1 \mathrm{VaR}_\alpha(X)\phi(\alpha)\,\mathrm{d}\alpha.$$

Proof

1. ⇒ 3. was shown implicitly in Corollary 4.1.4. Since the expectation is the spectral risk measure with spectrum $\phi(y) := 1$, $y \in [0,1]$, 3. ⇒ 2. is trivial. 2. ⇔ 1. is provided by the classical Portmanteau theorem (cf. (Billingsley, 2009, Theorem 2.1)). □

The conditions in the former theorems are not too strict for practical use in mathematical finance. Typically, pricing formulas of contingent claims are continuous w.r.t. the model's parameters (since the derivation of sensitivities is crucial, even differentiability often holds). Furthermore, the additional assumption of boundedness is not really critical, since it is also heavily used when it comes to numerical problems in mathematical finance, as, e.g., numerical option pricing using finite element methods, where options are typically computed as barrier options with out-of-scope barriers.

The class of risk-capturing functionals being generated by entropic risk measures fulfills the convergence property for consistent estimators as well, if the price of the contingent claim is a continuous and bounded function of the parameters. Since the entropic-induced risk-capturing functional is not coherent, it is not covered by the above propositions.

Proposition 4.1.7 (Entropic-induced risk-capturing functions fulfill (CP))
Let $\Theta \subset \mathbb{R}^m$ be a Euclidean parameter space. Let $(R_N)_{N\in\mathbb{N}}$ be a sequence of probability distributions on Θ converging weakly to a probability distribution R_0. Let X be a contingent claim such that the mapping $\theta \mapsto \mathbb{E}_\theta[X]$ is bounded and continuous, so $X \in \mathcal{C}^{\mathcal{C}^b(\Theta)}$. Denote the entropic parameter risk-capturing functional w.r.t. R_N by $\Gamma_\lambda^{\mathrm{ent},N}$ and the one w.r.t. R_0 by $\Gamma_\lambda^{\mathrm{ent},0}$.

It then follows $\Gamma_\lambda^{\mathrm{ent},N}(X) \xrightarrow{N\to\infty} \Gamma_\lambda^{\mathrm{ent},0}(X)$ for all fixed $\lambda \in (0,\infty)$.

Proof

Let $\lambda \in (0,\infty)$ be arbitrary but fix. Since $f(\theta) := \mathbb{E}_\theta[X]$ is assumed to be continuous and bounded, $u_\lambda \circ f$ is continuous and bounded as well for $u_\lambda(x) := \exp(\lambda x)$. Since the expectation of $u_\lambda \circ f$ w.r.t. the measure R^N is exactly the AVaR_1 of $u_\lambda \circ f$, we obtain by Proposition 4.1.3 the convergence

$$\int_\Theta u_\lambda(\mathbb{E}_\theta[X]) R_N(\mathrm{d}\theta) \xrightarrow{N\to\infty} \int_\Theta u_\lambda(\mathbb{E}_\theta[X]) R_0(\mathrm{d}\theta)$$

and since $u_\lambda^{-1}(y) = \lambda^{-1}\log(y)$ is continuous, also

$$u_\lambda^{-1}\left(\int_\Theta u_\lambda(\mathbb{E}_\theta[X])\, R_N(\mathrm{d}\theta)\right) = \Gamma_\lambda^{\mathrm{ent},N}(X) \xrightarrow{N\to\infty} \Gamma_\lambda^{\mathrm{ent},0}(X)$$

follows. □

4.2 Counterexamples

Up to now, we have shown that property (CP) holds for some popular risk-capturing functionals. However, property (CP) is not fulfilled by every risk-capturing functional. It can easily be seen that the ess sup-driven risk-capturing functional does not provide convergence for bounded and continuous functions $\theta \mapsto \mathbb{E}_\theta[X]$.

Example 4.2.1 (Essential supremum does not fulfill (CP))
Consider the parameter space $\Theta = \mathbb{R}$ and the sequence of distributions $R_N \sim \mathcal{N}(0, 1/N)$. Let $f \in \mathcal{C}^b(\mathbb{R})$. Obviously, for $N \to \infty$, $R_N \to \delta_0$ weakly, but on the other hand we obtain $\operatorname{ess\,sup}_{R_N} f = \sup_{x\in\mathbb{R}} f(x)$ for all $N \in \mathbb{N}$, since for every $x \in \mathbb{R}$ and every $\varepsilon > 0$ the environment $B_\varepsilon(x) := \{y \in \mathbb{R} : |x - y| < \varepsilon\}$ has positive measure under the normal distribution, i.e. for all $N \in \mathbb{N}$ $R_N(B_\varepsilon(x)) > 0$ holds. Thus, $\lim_{N\to\infty} \operatorname{ess\,sup}_{R_N} f = \sup_{x\in\mathbb{R}} f(x)$ which, in general, does not coincide with $\operatorname{ess\,sup}_{\delta_0} f = f(0)$.

Furthermore, one can show that boundedness of the plug-in function is a crucial assumption for the validity of the convergence property, even for the Average-Value-at-Risk w.r.t. an arbitrary level $\alpha \in (0, 1]$.

Example 4.2.2 (cf. (Stahl et al., 2012))
Let the parameter space $\Theta = \mathbb{R}$ be the real axis and regard the distributions $(R_N)_{N\in\mathbb{N}}$ on $\mathbb{R}$ defined as convex combinations of uniform and Dirac measures via

$$R_N := \frac{N-1}{N} U[0,1] + \frac{1}{N}\delta_{N^2},$$

denoting by $U[0,1]$ the uniform distribution on the unit interval. Obviously, $R_N \to U[0,1]$ weakly for $N \to \infty$. Now take the (unbounded and continuous) identity function $f(x) = x$ and calculate the AVaR_α-induced risk-captured prices for $\alpha \in (0, 1]$. Obviously, for the limit distribution, $\mathrm{AVaR}_\alpha(f; U[0,1]) = 1 - \alpha$, since f is the identity. On the other hand, we have

$$\begin{aligned}\mathrm{AVaR}_\alpha(f; R_N) &\geq \mathrm{AVaR}_1(f; R_N)\\ &= \mathbb{E}_{R_N}(f) = \frac{N-1}{2N} + N > N,\end{aligned}$$

thus $\mathrm{AVaR}_\alpha(f; R_N) \to \infty$ for $N \to \infty$.

4.3 Outlook

Some further investigations about convergence of (law-invariant) convex risk measures have been made in (Krätschmer et al., 2012), but in a more general environment. The

notion of weak convergence of measures, which is used in this chapter, is a too weak postulate in case the function $\theta \mapsto \mathbb{E}_\theta[X]$ is not bounded or not continuous, e.g. $\theta \mapsto \mathbb{E}_\theta[X] \in L^p(\Theta)$ for $p \in [1, \infty]$. They introduce a topology on the set of distributions (the so-called ψ-weak topology w.r.t. some function ψ as in (Weber, 2006)) that can be strictly stronger than the weak topology, yielding similar convergence results in case of dropping continuity and/or boundedness. Using these results, one may equally prove the theorems presented in this section.

5 Application: Bid-ask prices implied by estimation risk

In numerous cases in mathematical finance and particularly econometrics, one operates with time series data and processes these to estimate financial market models for, e.g., derivatives pricing and hedging, risk management, or technical analysis. If available, most financial market professionals try to rely on forward-looking data as, e.g., market prices of options, since these instruments carry the information of *market-implied* price dynamics. But, unfortunately, in numerous cases, forward-looking data is not available (e.g. stocks with little market turnover), while there is an urgent need to estimate parameters in such a situation. Hence, historical estimation of parameters from time series of financial market data remains an important tool. In this chapter, we translate the results obtained for risk-capturing functionals exhibiting property (CP) to consistent estimators. Furthermore, we state asymptotic results in case that the estimator's distribution is not known, but one knows the asymptotic distribution, which is often the case in practical applications. Finally, we treat estimation risk in two extensive numerical case studies, one concerning multivariate equity markets, the other concerning the joint modeling of electricity and commodity markets. Sections 5.1-5.3 are based on the paper (Bannör and Scherer, 2013a), while Section 5.4 is based on the paper (Bannör et al., 2013).

5.1 Estimation risk-captured prices for consistent estimators

In the previous chapter, we have investigated continuity properties of risk-captured prices in presence of weak convergence of the distribution of the parameters. This can now be translated into the case of consistent estimators.

Corollary 5.1.1 (Property (CP) for consistent estimators)
Let $(\hat{\theta}_N)_{N\in\mathbb{N}}$ be a consistent estimator sequence for some true parameter $\theta_0 \in \Theta$, i.e. $\hat{\theta}_N = \hat{\theta}(x_1, \dots, x_N)$ for some sample $x_1, \dots, x_N$, $N \in \mathbb{N}$, with support in the parameter space Θ and define the distributions quantifying the estimation risk as the estimator's distributions, i.e. $R_N := P^{N\hat{\theta}_N}$, the pushforward measure of the random variable $\hat{\theta}_N$ w.r.t. the product measure P^N. Let furthermore $(\Gamma_N)_{N\in\mathbb{N}}$ be a sequence of risk-capturing functionals that fulfill the convergence property (CP), i.e. each generator ρ_N is R_N-law-invariant and weak convergence is preserved. Then the sequence of estimation risk-captured prices converges to the plug-in price w.r.t. the true parameter, i.e. $\Gamma_N(X) \to \mathbb{E}_{\theta_0}[X]$, $N \to \infty$.

Proof
Since $(\hat{\theta}_N)_{N\in\mathbb{N}}$ is a consistent estimator, $\hat{\theta}_N \to \theta_0$ P-stochastically. In particular, we obtain $\hat{\theta}_N \to \theta_0$ in P-distribution, which exactly describes weak convergence of the push-forward measures R_N to the Dirac measure δ_{θ_0}. Hence, Proposition 4.1.2 yields the assertion. □

Having proved that convergence of risk-captured prices prices holds in Corollary 5.1.1, we apply the results to the Black–Scholes example from Example 3.1.8.

Example 5.1.2
Continuing the Black–Scholes example from Example 3.1.8, it is a well-known fact that the estimator $\hat{\sigma^2}_N$ for the variance, dependent on the sample size $N \in \mathbb{N}$, is consistent (i.e. converges in probability to the true variance parameter σ_0^2) and follows a Gamma distribution. Unfortunately, the function assigning to every volatility σ the Black–Scholes prices $BS(\sigma)$ for some option is continuous, but not bounded. Hence, one cannot directly apply the above results. But, if we either restrict the available volatilities to some compact set $\Sigma \subset \mathbb{R}_{\geq 0}$ or cut the Black–Scholes price at some $C > 0$, i.e. regarding the function $BS^{cut}(\sigma) := C \wedge BS(\sigma)$, we obtain a bounded and continuous function in σ. In practice, similar restrictions are often done (e.g. in the uncertain volatility model of (Avellaneda et al., 1995) or for finite element PDE pricing). Hence, when applying these mild restrictions, the estimation risk-captured Black–Scholes prices converge to the Black–Scholes price w.r.t. the true volatility parameter σ_0.

In case of the entropic-induced risk-capturing functional, we can furthermore show that in case of an unbiased estimator and convex relationship between parameter and price, the risk-captured price always overstates the plug-in price w.r.t. the true parameter.

Remark 5.1.3
Let $\hat{\theta}$ be an unbiased estimator for the true parameter $\theta_0 \in \Theta$ and denote for $\lambda \in (0, \infty)$ the entropic risk-capturing functional w.r.t. the estimator's distribution by $\Gamma_\lambda^{\mathrm{ent}}(X)$. Let $X \in \mathcal{C}^{\mathcal{L}^\infty(\Theta)}$. If $\theta \mapsto \mathbb{E}_\theta[X]$ is convex, $\Gamma_\lambda^{\mathrm{ent}}(X) \geq \mathbb{E}_{\theta_0}[X]$ holds.

Proof
Denote the estimator's distibution by R and let $\lambda \in (0, \infty)$ be arbitrary. Defining the convex function $u_\lambda(x) := \exp(\lambda x)$ and applying Jensen's inequality twice establishes

$$\begin{aligned}\Gamma_\lambda^{\mathrm{ent}}(X) &= u_\lambda^{-1}\left(\int_\Theta u_\lambda(\mathbb{E}_\theta[X])\, R(\mathrm{d}\theta)\right) \\ &\geq \int_\Theta \mathbb{E}_\theta[X]\, R(\mathrm{d}\theta) \geq \mathbb{E}_{\int_\Theta \theta\, R(\mathrm{d}\theta)}[X] = \mathbb{E}_{\theta_0}[X].\end{aligned}$$ □

5.2 Asymptotics of risk-capturing functionals

Up to now, we have treated estimation risk-captured prices where the parameter's distribution is given by the pushforward measure of an estimator. But, in many cases, the calculation of estimation risk-captured prices bears a substantial obstacle: For many estimators, the distribution is not known or not available in closed form. In this case, it may be recovered by resampling methods such as bootstrapping, which is often computationally expensive. Fortunately, there are situations where this problem can be circumvented by substituting the estimator's actual distribution by the asymptotic distribution, provided that the sample size is large enough. As an important example, broad classes of estimators (as, e.g., Maximum Likelihood estimators under mild technical conditions, cf. (van der Vaart, 2000, Theorem 5.39)) share the feature of asymptotic normality (like, e.g., the arithmetic mean as the expectation estimator which is guaranteed by the central limit theorem). In this case, the delta method (see (van der Vaart, 2000, Theorem 3.1)) provides convenient approximations for the distribution quantifying the parameter risk of the price, provided that the sample size is large enough.

Proposition 5.2.1 (Asymptotics for parameter risk, general version)
Let $(\hat{\theta}_N)_{N\in\mathbb{N}}$ be an asymptotically normal sequence of estimators for the true parameter $\theta_0 \in \Theta \subset \mathbb{R}^m$ with asymptotic positive definite covariance matrix Σ, i.e. $\sqrt{N}(\hat{\theta}_N - \theta_0) \to \mathcal{N}_m(0, \Sigma)$ weakly. Let furthermore $X \in \mathcal{C}^{\mathcal{L}^\infty(\Theta)}$. If the plug-in parameter price $\theta \mapsto \mathbb{E}_\theta[X]$ is continuously differentiable in θ_0 and the gradient $\nabla\mathbb{E}_{\theta_0}[X] \neq 0$, then

$$\sqrt{N}(\mathbb{E}_{\hat{\theta}_N}[X] - \mathbb{E}_{\theta_0}[X]) \xrightarrow{N\to\infty} \mathcal{N}\left(0, (\nabla\mathbb{E}_{\theta_0}[X])' \cdot \Sigma \cdot \nabla\mathbb{E}_{\theta_0}[X]\right)$$

weakly.

Proof
Follows directly from applying the delta method as described in (van der Vaart, 2000, Theorem 3.1). □

As an immediate corollary, we obtain an approximation for the "parameter risk distribution" w.r.t. a Maximum Likelihood estimator (in case of mild technical requirements on the parameter estimation), since this class of estimators fulfill asymptotic normality.

Corollary 5.2.2 (Asymptotics for parameter risk, ML estimators)
Let $(\hat{\theta}_N)_{N\in\mathbb{N}}$ be a Maximum Likelihood estimator sequence for the true parameter θ_0 and let the conditions for asymptotic normality hold for the Maximum Likelihood estimator. Furthermore, denote by $\mathcal{I}(\theta)$ the Fisher information matrix w.r.t. $\theta \in \Theta$. Let furthermore $X \in \mathcal{C}^{\mathcal{L}^\infty(\Theta)}$. If the plug-in parameter price $\theta \mapsto \mathbb{E}_\theta[X]$ is continuously differentiable in θ_0 and the gradient $\nabla\mathbb{E}_{\theta_0}[X] \neq 0$, then

$$\sqrt{N}(\mathbb{E}_{\hat{\theta}_N}[X] - \mathbb{E}_{\theta_0}[X]) \xrightarrow{N\to\infty} \mathcal{N}\left(0, (\nabla\mathbb{E}_{\theta_0}[X])' \cdot \mathcal{I}^{-1}(\theta_0) \cdot \nabla\mathbb{E}_{\theta_0}[X]\right)$$

weakly.

Proof
If the Maximum Likelihood estimator is asymptotically normal, its asymptotic covariance matrix is the inverse Fisher information matrix $\mathcal{I}^{-1}(\theta_0)$ w.r.t. the true parameter θ_0. Applying Proposition 5.2.1 yields the assertion. □

With this approximation at hand, the calculation of the AVaR- and entropic-induced risk-capturing functionals can be reduced to a simple symmetric interval around the evaluation at the true parameter since they preserve weak convergence of the distributions, provided that some assumptions hold.

We start with the entropic risk-capturing functional.

Proposition 5.2.3 (Approximation of entropic-induced risk-captured prices)
Let $\Gamma_\lambda^{\mathrm{ent}}$ be the risk-capturing functional induced by the entropic risk measure with risk aversion coefficient $\lambda > 0$. If $(\hat{\theta}_N)_{N\in\mathbb{N}}$ is an asymptotically normal estimator for the true parameter θ_0 with asymptotic covariance matrix Σ and X a contingent claim such that the plug-in price function $\theta \mapsto \mathbb{E}_\theta[X]$ is continuous and bounded, then we can approximate the entropic-induced risk-captured price $\Gamma^{\mathrm{ent}}(X; R_N)$ via

$$\Gamma_\lambda^{\mathrm{ent},N}(X) := \Gamma^{\mathrm{ent}}(X; R_N) \approx \mathbb{E}_{\theta_0}[X] + \frac{\lambda}{2N}(\nabla\mathbb{E}_{\theta_0}[X])' \cdot \Sigma \cdot \nabla\mathbb{E}_{\theta_0}[X],$$

denoting the distribution w.r.t. $\hat{\theta}_N$ by R_N.

Proof

Due to the asymptotic normality of $(\hat{\theta}_N)_{N\in\mathbb{N}}$*, we approximate* $\mathbb{E}_{\hat{\theta}_N}[X]$ *by a normal random variable* Y *with*

$$Y \sim \mathcal{N}\left(\mathbb{E}_{\theta_0}[X], \frac{(\nabla\mathbb{E}_{\theta_0}[X])' \cdot \Sigma \cdot \nabla\mathbb{E}_{\theta_0}[X]}{N}\right).$$

Using the weak continuity of the entropic-induced risk-captured price, pricing with Y *instead with* $\mathbb{E}_{\hat{\theta}_N}[X]$ *is reasonable for large* $N \in \mathbb{N}$*. With this approximation at hand, we have* $\Gamma_\lambda^{\mathrm{ent},N}(X) \approx \log\mathbb{E}[\exp(\lambda Y)]/\lambda$*, which is (up to the scaling factor* λ*) the cumulant-generating function of* Y *evaluated at* λ*. Hence, we readily obtain*

$$\Gamma_\lambda^{\mathrm{ent},N}(X) \approx \frac{1}{\lambda}\log\mathbb{E}[\exp(\lambda Y)] = \mathbb{E}_{\theta_0}[X] + \frac{\lambda}{2N}(\nabla\mathbb{E}_{\theta_0}[X])' \cdot \Sigma \cdot \nabla\mathbb{E}_{\theta_0}[X]. \qquad \square$$

A similar closed-form approximation can be found in case of the AVaR-induced risk-capturing functional, since in the normal case, the AVaR can be calculated in closed form.

Proposition 5.2.4 (Approximation of* AVaR*-induced risk-captured prices)

Let $(\hat{\theta}_N)_{N\in\mathbb{N}}$ *be an asymptotically normal estimator for the true parameter* θ_0 *with asymptotic covariance matrix* Σ *and* X *a contingent claim such that the plug-in price function* $\theta \mapsto \mathbb{E}_\theta[X]$ *is continuous and bounded. Let furthermore* $\alpha \in (0,1]$ *and denote the* AVaR_α*-induced risk-capturing functional w.r.t. the distribution of* $\hat{\theta}_N$*,* $N \in \mathbb{N}$*, by* $\hat{\theta}_N * \mathrm{AVaR}_\alpha$*. Then the risk-captured price of* X *can be approximated via*

$$\hat{\theta}_N * \mathrm{AVaR}_\alpha(X) \approx \mathbb{E}_{\theta_0}[X] + \frac{\varphi(\Phi^{-1}(1-\alpha))}{\alpha\sqrt{N}}\sqrt{(\nabla\mathbb{E}_{\theta_0}[X])' \cdot \Sigma \cdot \nabla\mathbb{E}_{\theta_0}[X]},$$

denoting by φ *the density and by* Φ *the distribution function of the standard normal law.*

Proof

Applying (2.1) for the AVaR *of a normal distributed random variable, we immediately obtain*

$$\theta_N * \mathrm{AVaR}_\alpha(X) \approx \mathbb{E}_{\theta_0}[X] + \frac{\varphi(\Phi^{-1}(1-\alpha))}{\alpha\sqrt{N}}\sqrt{(\nabla\mathbb{E}_{\theta_0}[X])' \cdot \Sigma \cdot \nabla\mathbb{E}_{\theta_0}[X]}.$$

The normal approximation sheds some light on the convergence rates for the entropic and the AVaR risk measure and immediately yield some questions.

Remark 5.2.5 (Speed of convergence of risk-captured prices)

As discussed in Section 3.3, both the entropic risk measure and the AVaR risk measure deliver some "interpolations" between the essential supremum and the expected value, i.e.

$$\lim_{\lambda\to 0}\rho_\lambda^{\mathrm{ent}}(X) = \mathbb{E}[X] \quad , \quad \lim_{\lambda\to\infty}\rho_\lambda^{\mathrm{ent}}(X) = \operatorname{ess\,sup} X,$$
$$\lim_{\alpha\to 1}\mathrm{AVaR}_\alpha(X) = \mathbb{E}[X] \quad , \quad \lim_{\alpha\to 0}\mathrm{AVaR}_\alpha(X) = \operatorname{ess\,sup} X,$$

and so do their induced risk-captured prices. On the other hand, the formulae derived in the Propositions 5.2.3 and 5.2.4 yield that the entropic-induced risk-captured price converges with $\mathcal{O}(1/N)$ to the risk-neutral price computed with true parameter θ_0, while $\theta_N * \mathrm{AVaR}_\alpha$ converges with $\mathcal{O}(1/\sqrt{N})$. At first glance, this result may look surprising, since the limit points are the same. But, on the one hand, for the expected value, the upper approximation does not hold, since for $\lambda \to 0$ resp. $\alpha \to 1$ the feasible approximation would simply be $\mathbb{E}_{\theta_0}[X]$. Hence, there is "infinite speed of convergence" for both the AVaR and entropic risk-captured prices. On the other hand, property (CP) does not hold for the essential supremum (see Example 4.2.1). Hence, convergence to the risk-neutral price is actually not provided for the essential supremum. Hence, although the AVaR and the entropic risk-captured prices are both reasonable "interpolations" between the extreme cases of the essential supremum (complete risk aversion) and the expectation (no risk aversion), the asymptotic speed of convergence differs considerably.

5.3 Case study: Estimation risk for Margrabe options in a two-dimensional Black-Scholes market

In Section 3.4, we have already stated examples arising from market-implied distributions. These examples are quite simple, since they only deal with discrete distributions and are therefore more of a descriptive nature. We now give a more specific example where the incorporation of parameter estimation may be useful: The correlation between two stocks in a two-dimensional Black-Scholes market can typically not be derived from liquid assets (such as stock prices from cash equity markets or the implied volatility from liquid option markets), which implies that obtaining the correlation via calibration to market prices is often not possible. But under an equivalent change of measure (which is done to change to the suitable risk-neutral measure), the correlation in a two-dimensional Black-Scholes market remains invariant (cf. (Jacod and Shiryaev, 2003, Theorem III.3.24)). Therefore, a feasible method to obtain the correlation is to estimate the correlation historically over a suitable period.

We give a numerical example and a comparison between the different risk-captured prices, dependent on the sample size used to estimate the correlation.

Example 5.3.1 (Exchange option with estimated correlation)
Suppose we have a Black–Scholes model with three assets, the money market account $S^{(0)}$ and two stocks $S^{(1)}$ and $S^{(2)}$. Under the real-world measure P, they follow the dynamics

$$\mathrm{d}S_t^{(0)} = rS_t^{(0)}\,\mathrm{d}t,\ \mathrm{d}S_t^{(i)} = \mu_i S_t^{(i)}\,\mathrm{d}t + \sigma_i S_t^{(i)}\,\mathrm{d}W_t^{(i)},\ i = 1, 2,$$

5.3 Case study: Estimation risk for Margrabe options in a two-dimensional Black-Scholes market

with $(W_t^{(i)})_{t\in[0,T]}$ *being correlated Wiener processes,* $\mu_i \in \mathbb{R}$, $r, \sigma_i \in \mathbb{R}_{>0}$ *for* $i = 1, 2$. *Furthermore we assume the covariation of the Wiener processes to be given by*

$$\mathrm{d}W_t^{(1)}\,\mathrm{d}W_t^{(2)} = \rho\,\mathrm{d}t$$

for an unknown $\rho \in [-1, 1]$. *Therefore the covariation between the stock prices is* $\mathrm{d}S_t^{(1)}\,\mathrm{d}S_t^{(2)} = \rho\sigma_1\sigma_2 S_t^{(1)} S_t^{(2)}\,\mathrm{d}t$. *The risk-neutral dynamics follow*

$$\mathrm{d}S_t^{(0)} = rS_t^{(0)}\,\mathrm{d}t,\ \mathrm{d}S_t^{(i)} = rS_t^{(i)}\,\mathrm{d}t + \sigma_i S_t^{(i)}\,\mathrm{d}W_t^{(i)},\ i = 1, 2,$$

and the covariation between the stocks remains $\mathrm{d}S_t^1\,\mathrm{d}S_t^2 = \rho\sigma_1\sigma_2 S_t^1 S_t^2\,\mathrm{d}t$.

We are interested in the fair value of a European Margrabe option *(also called* exchange option*), giving the holder the right to exchange stock 1 into an equal amount of stock 2 at maturity* T, *so the payoff of the contingent claim with exercise date* $T > 0$ *is* $X = (S_T^{(2)} - S_T^{(1)})^+$. *In (Margrabe, 1975) a closed-form solution is calculated based on a change of numéraire technique, which is given by*

$$BS(S_0^{(1)}, S_0^{(2)}, \sigma_1, \sigma_2, T, \rho) = S_0^{(2)}\Phi(d_1) - S_0^{(1)}\Phi(d_2).$$

with

$$d_1 := \frac{\log(S_0^{(2)}) - \log(S_0^{(1)}) + 0.5T\tilde{\sigma}^2}{\tilde{\sigma}\sqrt{T}}, \quad d_2 := d_1 - \tilde{\sigma}\sqrt{T}, \quad \tilde{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho}.$$

We assume that all parameters except for the correlation ρ *are known (in cases of fairly liquid plain vanilla options on the stocks, this assumption may not be unappropriate). One possible approach to obtain an estimate for the correlation is to estimate it via Pearson's sample correlation coefficient*

$$\rho^{(N)} = \frac{N\sum_{i=1}^N x_i y_i - \left(\sum_{i=1}^N x_i\right)\left(\sum_{i=1}^N y_i\right)}{\sqrt{\sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i\right)^2}\sqrt{\sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i\right)^2}}$$

with $N \in \mathbb{N}$ *being the largest number of feasible samples,* x_i, y_i *denoting the log-returns of stock 1 resp. 2 for* $i = 1, \ldots, N$. *Since a change of measure to the equivalent risk neutral measure does not affect the parameter* ρ, *it is reasonable to derive the correlation* ρ *under the real-world measure.*

Without accounting for the uncertainty of the parameter ρ, *we would simply plug in the resulting point estimate of Pearson's correlation estimator into the pricing formula. To*

calculate a risk-captured price, however, we first have to determine the distribution of the correlation in the parameter space $[-1,1]$ *induced by the estimator* $\rho^{(N)}$*.*

Assuming to live in a Black–Scholes world, log returns are independent over time and follow a bivariate normal distribution with correlation ρ*. Therefore the well-known Fisher transformation (see (Fisher, 1915)) can be applied, i.e. the transformed distribution approximately fulfills*

$$\operatorname{artanh}(\rho^{(N)}) \sim \mathcal{N}\left(\operatorname{artanh}(\rho_0), \frac{1}{N-3}\right), \quad N \gg 3,$$

where ρ_0 *denotes the true correlation parameter. Furthermore, it has to be noted that Pearson's correlation estimator is consistent but only asymptotically unbiased.*[1]

We are now able to calculate risk-captured prices. We start with the calculation of the entropic-driven risk-captured price: We fix $\lambda \in (0,\infty)$ *and obtain*

$$\begin{aligned}\Gamma^{\mathrm{ent}}_{\lambda,N}(X) &= \frac{1}{\lambda}\log\left(\int_{-1}^{1}\exp\left(\lambda \cdot BS(S_0^{(1)}, S_0^{(2)}, \sigma_1, \sigma_2, T, \rho)\right) P^{\rho^{(N)}}(\mathrm{d}\rho)\right)\\ &\approx \frac{1}{\lambda}\log\left(\int_{-\infty}^{\infty}\exp\left(\lambda \cdot BS(S_0^{(1)}, S_0^{(2)}, \sigma_1, \sigma_2, T, \tanh(x\sqrt{N-3}+\rho_0))\right)\varphi(x)\,\mathrm{d}x\right).\end{aligned}$$

This integral has to be evaluated numerically and defines the entropic-induced ask price. In the further, we abbreviate

$$f(\rho) := BS(S_0^{(1)}, S_0^{(2)}, \sigma_1, \sigma_2, T, \rho),$$

fixing the parameters $S_0^{(1)}, S_0^{(2)}, \sigma_1, \sigma_2$*.*

For an efficient calculation of $\rho^{(N)} * \mathrm{AVaR}$*, we use the characterization of* AVaR *as tail conditional expectation from (Föllmer and Schied, 2004, Corollary 4.49) and calculate*

$$\rho^{(N)} * \mathrm{AVaR}_\alpha(X) = \mathbb{E}_{\rho^{(N)}}\left[f | f \geq \mathrm{VaR}_\alpha(f)\right],$$

which is favourable from a computational point of view. The conditional expectation $\mathbb{E}_{\rho^{(N)}}[f|f \geq \mathrm{VaR}_\alpha(f)]$ *can easily be estimated by Monte Carlo simulation given the distribution of* $\rho^{(N)}$*. In our example, we use the Fisher transformation approximation due to its simplicity.*

In our numerical example we assume the following values to be known: $S_0^{(1)} = 120$*,* $S_0^{(2)} = 110$*,* $\sigma_1 = 0.16$*,* $\sigma_2 = 0.32$*, and* $T = 1$*. The correlation* ρ *is supposed to be*

[1] To obtain a truly unbiased estimator, one can use the Olkin–Pratt adjustment of the estimator as described in (Olkin and Pratt, 1958), at the cost of a much more complicated distribution of the estimator.

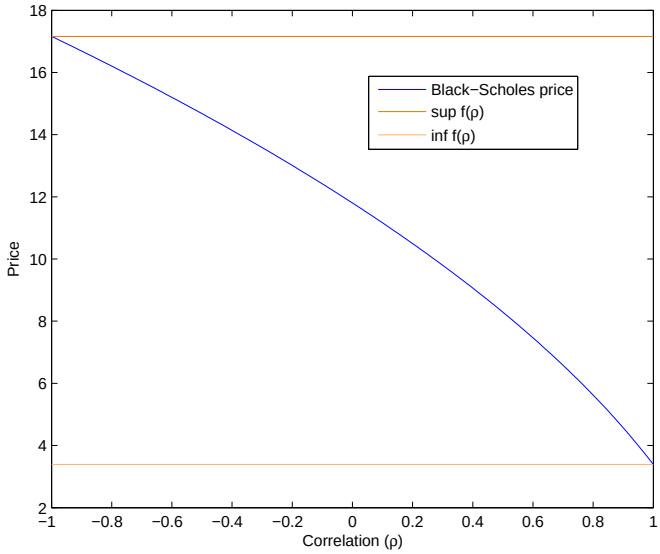


Figure 5.1 Exchange option price as a function of the stocks' correlation, i.e. $\rho \mapsto f(\rho)$. In this case, the option price is decreasing and concave in correlation.

unknown, so we are exposed to parameter uncertainty w.r.t. ρ. As visualized in Figure 5.1, the correlation parameter has massive influence on the price of the exchange option. In case of perfect positive correlation the price is 3.39*, while in case of perfect negative correlation the price is* 17.16*. Furthermore, Figure 5.1 shows that in this case the upper and lower risk-captured price provide very rough estimates for a spread price when a supremum-generated risk-capturing functional to determine risk-captured prices is used, so the supremum/infimum approach may hardly be useful for a trader in this setting to calculate a bid-ask spread that is accepted by the market.*

If the stocks' correlation is estimated, we may have more information about the parameter and can apply the risk-capturing functionals developed in this thesis. Consistent with the bivariate Black-Scholes model, we assume to have estimated the correlation with Pearson's correlation estimator from bivariate normal distributed stock returns, the real correlation is supposed to be $\rho_0 = 0.4$ and obtain the entropic risk-captured price and the AVaR*-induced risk-captured prices dependent on the selection of $\lambda \in (0, \infty)$ resp. $\alpha \in (0, 1]$ and the sample size N. Since f is a continuous function being defined on the compact interval $[-1, 1]$, it fulfills the conditions to apply the convergence results of the previous section. Convergence can also be visualized by plotting our risk-captured prices against the sample size (see Figure 5.2).*

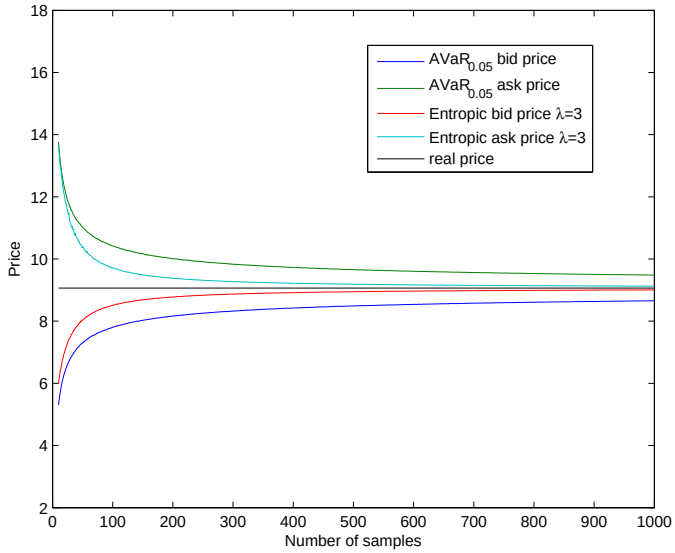


Figure 5.2 Different risk-captured prices as a function of the sample size. One can see the convergence of the risk-captured prices towards the risk-neutral price computed with the true parameter. Note that even for 250 samples (about one year of daily data), the bid-ask spreads induced by parameter risk are still considerably large.

First, we examine the AVaR*-induced risk-captured price. The remaining parameter to specify is the significance level* $\alpha \in (0, 1]$*, steering the level of conservativeness. For* $\alpha = 0$ *we obtain Cont's supremum approach and therefore the ask price* $\rho^{(N)} * \mathrm{AVaR}_0(X) = 17.16$ *- independent of the sample size. The higher (i.e. the less conservative) the significance level* $\alpha \in (0, 1]$ *is chosen, the narrower are bid-ask spreads. For illustration, we have plotted some spreads for different significance levels in Figure 5.3. As one can see, the expectation-induced price converges very fast to the real price (which does not provide a bid-ask spread due to the linearity of the expectation), while convergence velocity shrinks for lower significance levels. When regarding the entropic-induced prices, a higher risk-aversion parameter* λ *leads to wider bid-ask spreads. Since the entropic risk measure is not positively homogeneous, a higher quantity of the same claim leads to a higher price per notional in the same way a higher risk-aversion parameter does, since for a quantity* $a \in \mathbb{R}^+$ *the entropic-induced risk-captured price per notional w.r.t.*

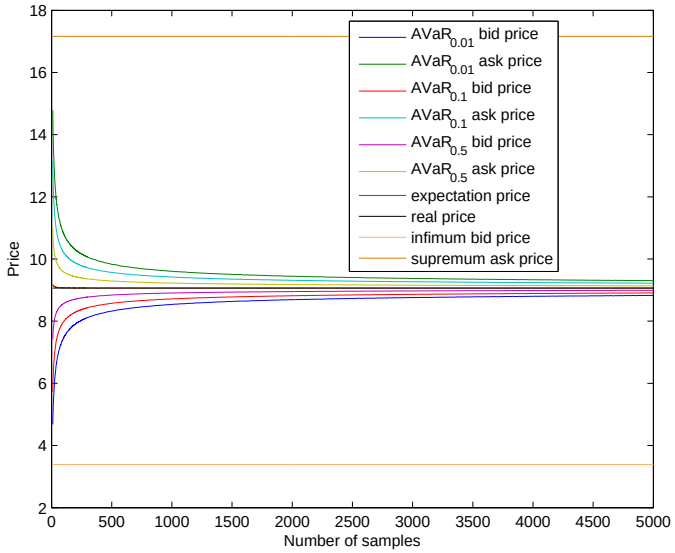


Figure 5.3 The AVaR-induced bid-ask prices for different significance levels. The higher the significance level $\alpha \in (0, 1]$ is chosen, the wider is the bid-ask spread. The extreme case of a supremum-induced ask price (which agrees with $\rho^{(N)} *$ AVaR_0) does not depend on the sample size and leads to a very conservative bid-ask spread which may not comply to bid-ask spreads being quoted in the market. The calculation of the expected value leads to prices being very close to the plug-in price of the true parameter.

a distribution S is calculated as

$$\begin{aligned}\frac{\Gamma_\lambda^{\mathrm{ent}}(aX)}{a} &= \frac{1}{a\lambda}\log(\mathbb{E}_S\left[\exp(\lambda a \mathbb{E}_\cdot[X])\right]) \\ &= \frac{1}{\tilde{\lambda}}\log(\mathbb{E}_S\left[\exp(\tilde{\lambda}\mathbb{E}_\cdot[X])\right]) = \Gamma_{\tilde{\lambda}}^{\mathrm{ent}}(X)\end{aligned}$$

with $\tilde{\lambda} := a\lambda$. Due to this symmetric relationship between quantities and risk-aversion parameters, we have only visualized how the risk-captured prices evolve for different risk-aversion parameters (see Figure 5.4). It can be seen that higher risk-aversion parameters bear numerical problems as well: The calculation of the exponential function w.r.t. a high value (occuring for high risk-aversion parameters) will be less accurate and therefore the results are less stable.

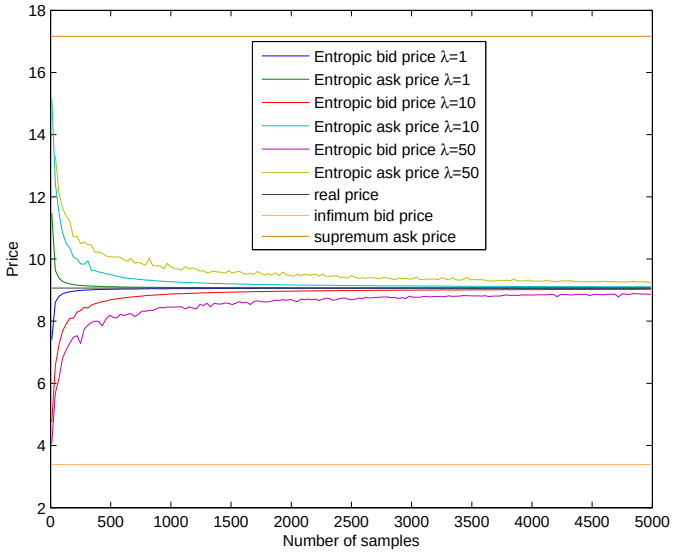


Figure 5.4 The entropic-induced bid-ask prices for different risk-aversion parameters. A higher risk aversion parameter $\lambda \in (0, \infty)$ results in wider bid-ask spreads, leading to supremum resp. infimum induced prices in the limit. The entropic-induced bid-ask prices are more fluctuating with rising risk aversion parameter due to numerical issues.

5.4 The parameter risk-captured valuation of a gas power plant

In energy finance, multi-asset models are used to evaluate derivatives that jointly depend on the price of electricity and some other commodity serving as an energy feedstock. In particular, power plants can be regarded as derivatives on the spread of the energy price and the respective fuel price. Typically, the parameters of these models are estimated from time series. However, since the models often have many different parameters to model the single prices and the price dependency, one is heavily exposed to parameter risk (more precise: estimation risk). In this section, we apply our developed framework to calculate parameter risk-captured bid-ask prices for a gas power plant. From a practical perspective, this topic is highly relevant considering the German "Energiewende", where new capacities of gas power plants have to be build to support renewable energy sources like wind or solar power. Hence, the correct valuation of gas power plants is crucial to ensure sound economic decisions for policymakers and power plant operating

companies.

Power plant valuation via spread options

Since we want to use financial models to evaluate the economic value of a gas power plant, we have to state some simplifying assumptions. Hence, we assume that an operating gas power plant's profit stems from the difference of the produced energy's price and the price of the fired gas that was needed to produce the energy. In particular, we disregard further costs as, e.g., maintenance costs. Additionally, due to the introduction of the European Union Emission Trading Scheme, power production by firing gas is accompanied by costs caused by the need to buy carbon emission certificates. Furthermore, we assume that gas power plants can easily be switched on and off on demand. Thus, the plant will only be switched on when the electricity price is higher than the price of the fired gas and the emission certificates. Hence, we model the daily profit $(V_t)_{t\geq 0}$ of a gas power plant as an optional payoff and obtain

$$V_t = \max\{P_t - h\,G_t - \eta E_t, 0\}, \tag{5.1}$$

where P_t is the power price, G_t is the gas price, E_t is the carbon certificate price, h is the heat rate of the power plant, and η is the CO2 emission rate of the power plant. Here, we assume t to run on a discrete daily scale, i.e. $t \in \{0, 1/365, 2/365, \dots\}$. Hence, the value of a gas power plant can be calculated as the present value of the expected daily profits, i.e.

$$\mathbb{E}\left[\sum_{t=0}^{T} \exp(-rt) V_t\right],$$

denoting by $r \in \mathbb{R}$ the risk-free rate and by $T > 0$ the remaining lifetime of the power plant. Using the terminology of (Burger et al., 2008), a gas power plant can be regarded as a strip of options on the *clean spark spread*, which is defined as the difference between the power price and the sum of the respective gas and emissions certificate prices.

The joint emissions/gas/power price model

We model the emission price $(E_t)_{t\geq 0}$ as a geometric Brownian motion (following the short-term emission price dynamics assumptions of (Yang et al., 2008)), i.e.

$$\mathrm{d}E_t = \alpha^E E_t\,\mathrm{d}t + \sigma^E E_t\,\mathrm{d}W_t^E, \tag{5.2}$$

the gas price $(G_t)_{t\geq 0}$ as a mean-reverting process[2]

$$\begin{aligned} G_t &= e^{g(t)+Z_t}, \\ \mathrm{d}Z_t &= -\alpha^G Z_t \,\mathrm{d}t + \sigma^G \,\mathrm{d}W_t^G, \end{aligned} \tag{5.3}$$

and the power price $(P_t)_{t\geq 0}$ as a sum of two mean-reverting processes[3]

$$\begin{aligned} P_t &= e^{f(t)+X_t+Y_t}, \\ \mathrm{d}X_t &= -\alpha^P X_t \,\mathrm{d}t + \sigma^P \,\mathrm{d}W_t^P, \\ \mathrm{d}Y_t &= -\beta Y_t \,\mathrm{d}t + \mathrm{d}Z_t, \end{aligned} \tag{5.4}$$

where $\alpha^G, \alpha^P, \beta \in \mathbb{R}$ are the respective mean-reversion forces for the gas and power price processes, $\alpha^E \in \mathbb{R}$ is the drift of the emission price process, $\sigma^E, \sigma^G, \sigma^P > 0$ are the respective volatilities for the emissions, gas, and power price processes, W^G, W^E, and W^P are Brownian motions, and Z is a compound Poisson process with intensity $\lambda > 0$ and homogeneous jump size distribution J. For the jump sizes, we consider two different scenarios: First, we suggest to use a non-central Laplace distribution to employ a heavy-tailed jump distribution. Second, for comparison, we employ the Gaussian distribution as already done in (Cartea and Figueroa, 2005). The functions $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ and $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ model the respective seasonal trends for the gas and power prices and are defined as

$$\begin{aligned} f(t) &= a_1 + a_2\, t + a_3 \cos(a_5 + 2\pi t) + a_4 \cos(a_6 + 4\pi t), \\ g(t) &= b_1 + b_2\, t + b_3 \cos(b_5 + 2\pi t) + b_4 \cos(b_6 + 4\pi t), \end{aligned}$$

where a_1 and b_1 are the production expenses, and a_2 and b_2 are the slopes of increase in these costs. The remaining parameters $a_3, a_4, a_5, a_6, b_3, b_4, b_5, b_6 \in \mathbb{R}$ are responsible for modeling seasonal price changes in the respective underlying. Furthermore, we assume that the processes W^E, W^G, Z are mutually independent, but the driving Brownian motions of the power and gas price processes are correlated by some correlation $\rho \in [-1, 1]$, i.e.

$$\mathrm{d}W_t^P \,\mathrm{d}W_t^G = \rho \,\mathrm{d}t. \tag{5.5}$$

As described above, two processes $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are responsible for capturing the power price movements. The process $(X_t)_{t\geq 0}$ is a Brownian-driven Ornstein–Uhlenbeck

[2] See (Lucia and Schwartz, 2002).
[3] See (Hambly et al., 2009).

process and models daily price fluctuations in the power price. On the other hand, the process $(Y_t)_{t\geq 0}$, which is an Ornstein–Uhlenbeck process driven by the compound Poisson process Z, is responsible for modeling sudden shocks (so-called "spikes") in the power price, which may occur due to sudden electricity shortages. A detailed discussion on electricity spikes can be found in (Benth et al., 2011).

Empirical Investigation

The total set of parameters[4] that has to be estimated in the multi-asset model is given by

$$\{\alpha^E, \sigma^E, g, \alpha^G, \sigma^G, f, \alpha^P, \beta, \sigma^P, \lambda, \mathbb{E}[J], \mathbb{E}[J^2], \rho\}.$$

Hence, our model for the clean spark spread has several degrees of freedom. Consequently, the risk of determining parameters wrongly is considerable and it will turn out that even the determination of single parameters may lead to tremendous results for prices obtained in the model. Thus, using the methodology of parameter risk-captured prices, we assess how the value of a gas power plant depends on the parameter risk of the chosen model.

Input data

To estimate the parameters, time series of three different contracts are analyzed[5]: The Phelix Day Base[6] price (EUR/MWh) as a proxy for electricity prices, the daily NCG[7] price (EUR/MWh) as a proxy for gas prices, and the daily emissions price[8] measuresd in EUR/EUA. The observation period covers approx. three years and runs from 05/29/2009 to 06/08/2012. Figure 5.5 depicts the development of the emissions, gas, and power

[4] In the parametric form of the Laplace resp. Gaussian distribution, we do not directly use the second moment as a parameter, but the standard scaling parameters for Laplace resp. normal distribution (variance for the normal distribution, mean absolute deviation from median for the Laplace distribution). Obviously, using the second moment is only an equivalent reparametrization. Furthermore, instead of specifying the respective parameters in the trend functions f and g, we abbreviate this by treating the trend functions f and g themselves as parameters.

[5] All datasets are taken from the European Energy Exchange, `www.eex.com`.

[6] Phelix Day Base is defined as the arithmetic average of the auction prices for the hours 1 to 24 in the market area Germany/Austria disregarding power transmission bottlenecks. Source: European Energy Exchange.

[7] Source: European Energy Exchange.

[8] Source: European Energy Exchange.

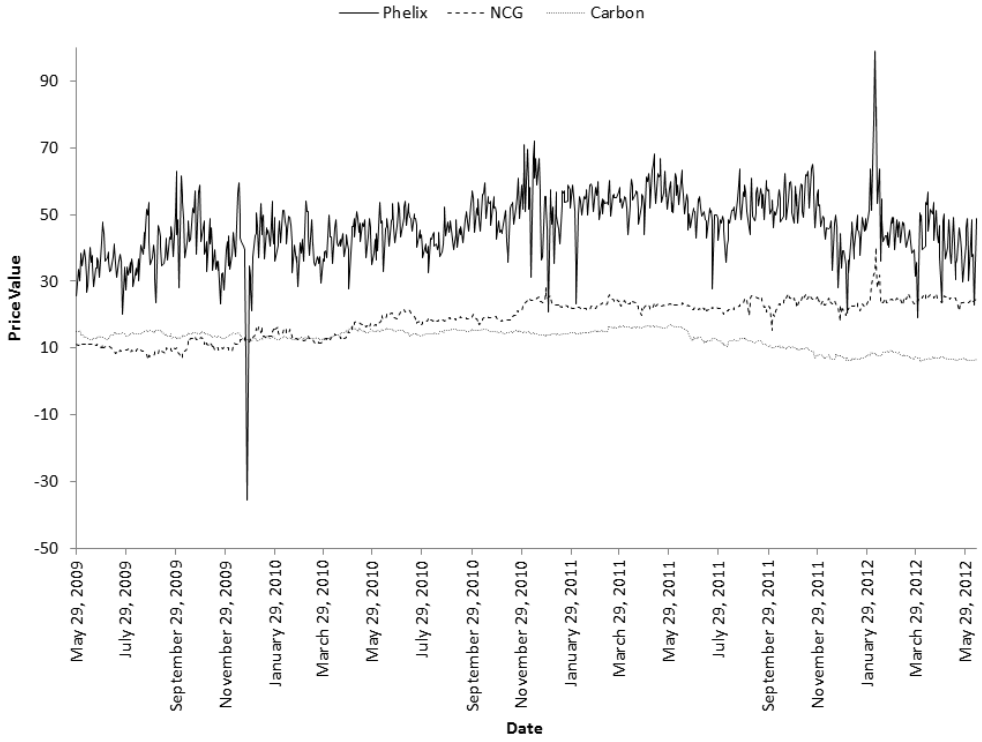


Figure 5.5 Evolution of the power (base load), gas, and carbon prices between 05/29/2009 and 06/08/2012.

prices during the observation period, while Figure 5.6 depicts the development of the spark spread path.

Parameter estimation techniques

We try to rely on maximum likelihood estimators (ML estimators) as far as possible, since maximum likelihood estimators exhibit (under mild technical conditions) asymptotic distributions that are normal and their asymptotic variance is provided by the inverse Fisher information matrix, i.e. for an ML estimator $\hat{\theta}_N$, dependent on the sample size $N \in \mathbb{N}$, one has

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \to \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0))$$

weakly, denoting by θ_0 the true parameter which is subject to be estimated and $\mathcal{I}(\theta_0)$ the Fisher information matrix. In particular, we use the following estimation techniques:

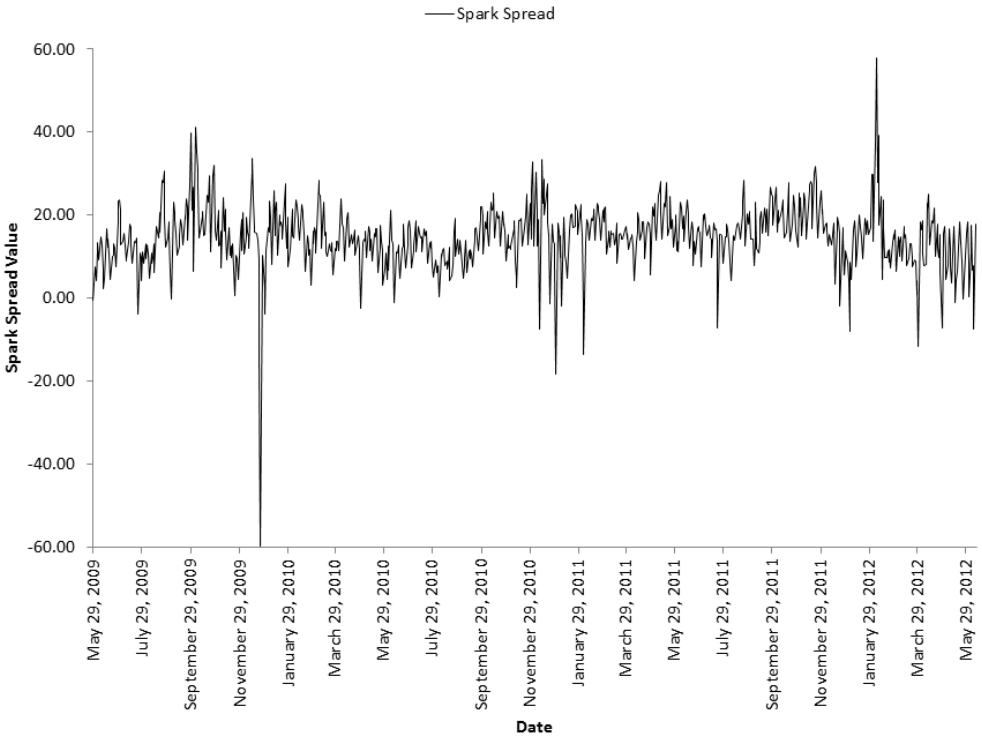


Figure 5.6 Evolution of the clean spark spread between 05/29/2009 and 06/08/2012.

1. *Estimation of the trend and seasonalization parameters* - i.e. estimation of the the parameters a_j, b_j, $j = 1, \ldots, 6$. These parameters represent the deterministic (linear and seasonal) trends of the gas and power price. We obtain the parameters by using linear (for the linear trend) and nonlinear (for the seasonal effects) regression on the functions f and g.

2. *Decomposition of the deseasonalized power price into Brownian part* $(X_t)_{t\geq 0}$ *and jump part* $(Y_t)_{t\geq 0}$ - The deseasonalized power price data has to be decomposed into two data sets: One arising from the Brownian-driven Ornstein–Uhlenbeck process $(X_t)_{t\geq 0}$ and another arising from the mean-reverting jump process $(Y_t)_{t\geq 0}$. We follow the methodology in (Cartea and Figueroa, 2005), where an iterative filtering technique is used. First, all data points with a deviation of more than three standard deviations of the mean are considered to arise from jumps and are taken out of the sample. Afterwards, the standard deviation of the remaining data points is recomputed and again all data points with a deviation of more than three standard deviations of the mean are considered to arise from jumps and are taken out of the sample. This procedure is repeated until all remaining data lies within the range of three standard deviations of the mean. Hence, the taken out data points represent the jump part $(Y_t)_{t\geq 0}$ and the remaining data repesents the continuous part $(X_t)_{t\geq 0}$.[9]

3. *Estimation of the emissions price process* - i.e. estimation of the parameters α^E, σ^E, which is done by the standard estimators for mean and standard deviation on the logarithmic process.

4. *Joint estimation of the base power and gas signals* - i.e. estimation of the parameters $\rho, \alpha^P, \alpha^G, \sigma^P, \sigma^G$. Since the process $(X_t, G_t)_{t\geq 0}$ is a two-dimensional (Brownian) Ornstein–Uhlenbeck process, we can estimate the parameters from the discrete data by estimating the parameters of a first order autoregressive model, i.e. an AR(1) process with normal noise. We use the standard maximum likelihood estimator, cf. (Knittel and Roberts, 2005).

[9] One should note that the threshold filtering technique to filter out the spikes – although often applied in practice – yields that only large jumps are filtered out. When estimating the jump size distribution from this data via ML estimation, one obtains considerably different location parameters for the normal and the Laplace distribution, since the ML estimator for the location parameter of the Laplace distribution is the median, while the ML estimator for the location parameter of the normal distribution is the mean. This methodology can be criticized, but we rely on standard techniques used in energy finance, cf. (Cartea and Figueroa, 2005).

5. *Estimation of the spike signal* $(Y_t)_{t\geq 0}$ - i.e. estimating the intensity λ, the speed of mean reversion β, and the jump distribution parameters. We estimate the intensity λ of the driving compound Poisson process by the spikes frequency, i.e. the number of detected spikes per period. Furthermore, we use the autocorrelation function techniques described in (Barndorff-Nielsen and Shephard, 2001) to obtain the speed of mean reversion β. Furthermore, we estimate the parameters of jump size distribution by applying the according ML estimators on the filtered jump data.

The result of the estimation procedure is given in Table 5.1.

Assessment of estimation risk

Since the distribution of the overall parameter distribution is very complex and difficult to obtain (e.g. due to the filtering technique), we reduce the problem by considering the estimator distributions of certain groups of parameters separately (e.g. the diffusion parameters, the jump size distribution parameters). Hence, we scrutinize the estimation risk w.r.t. selected groups of parameters and assume the remaining parameters to be known, i.e. we disregard the parameter risk arising from estimating the remaining parameters. In particular, all our results only describe *lower bounds* of the overall parameter risk. For parameter groups, we choose the following groups:

- Jump size distribution parameters $(\mathbb{E}[J], \mathbb{E}[J^2])$;
- Diffusion parameters for the emissions, gas, and power processes, i.e. the parameter set $(\alpha^E, \sigma^E, \alpha^G, \sigma^G, \alpha^P, \sigma^P, \rho)$;
- Diffusion parameters for the gas and power processes, i.e. $(\alpha^G, \sigma^G, \alpha^P, \sigma^P, \rho)$.

To incorporate parameter risk into prices, we use the AVaR-induced risk-captured prices w.r.t. different significance levels $\alpha \in (0, 1]$. As derived in Proposition 4.1.3, the AVaR-induced risk-captured prices are continuous w.r.t. the weak topology on the parameter distributions, if the price function $\theta \mapsto \mathbb{E}_\theta[X]$ is continuous and bounded in the parameter $\theta \in \Theta$. Hence, it is justified to approximate the estimator distributions by the respective asymptotic distributions and applying the approximation formula from Proposition 5.2.4. Therefore, we assume the respective plug-in parameter θ_0 as the true parameter and calculate the expected value by a Monte Carlo simulation with $K = 5000$ runs. To estimate the respective derivative $\partial \mathbb{E}_{\theta_0}/\partial\theta$, we use the Monte Carlo central-difference

Estimation Step	Product	Estimates	Method
Geometric Brownian motion	Emissions	$\alpha^E = -0.2843$, $\sigma^E = 0.4079$	MLE
Seasonal trend	Power	$a_1 = 3.6716$, $a_2 = 0.0980$, $a_3 = -0.0274$ $a_4 = 0.0368$, $a_5 = 0.6524$, $a_6 = 0.9530$	OLS
Seasonal trend	Gas	$b_1 = 2.3420$, $b_2 = 0.3503$, $b_3 = 0.0218$ $b_4 = -0.0445$, $b_5 = 0.7829$, $b_6 = 1.6126$	OLS
Filtering	Power		3×Std.Dev rule
Base process	Gas	$\alpha^G = 13.5827$, $\sigma^G = 0.7768$	Multivariate
Base process	Power	$\alpha^P = 121.8684$, $\sigma^P = 2.5943$, $\rho = 0.1247$	normal regression
Spike mean-reversion	Power	$\beta = 243.7240$	$2\times\alpha^P$
Spike intensity	Power	$\lambda = 13.4936$	Annual frequency
Spike size (Laplace)	Power	$\mu_s(median) = 0.3975$, $\sigma_s(scale) = 0.6175$	MLE
Spike size (normal)	Power	$\mu_s(mean) = 0.0863$, $\sigma_s^2(variance) = 0.5857$	MLE
Heat rate	Gas	$h = 2.5$	technical constant
Emission rate	Gas	$\eta = 0.4$	technical constant
Interest rate		$r = 3\%$	market-quoted

Table 5.1 The estimated parameters of the joint model for emissions, gas, and power prices and the respective estimation methods. One should note that the resulting parameters are sometimes quoted to different (i.e. daily) timescales.

estimator (cf. (Glasserman, 2004, p. 378f.)) applying the same paths from the Monte Carlo simulation.

Results and interpretation

Calculating the parameter risk-captured prices using the above methodology, we obtain risk-captured bid and ask prices. To compare the different risk-captured prices, we calculate the relative width of the parameter risk-implied bid-ask spread by

$$\frac{\text{ask price} - \text{bid price}}{\text{mid price}},$$

where the mid price is here given by the plug-in parameter price. A comparison of the relative width is done in Table 5.2.

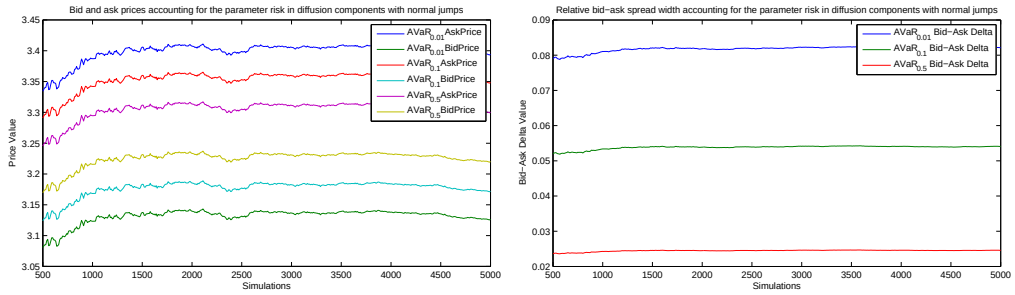
Due to the large sample size of $M = 790$ observations, the estimation of the diffusion signals has pretty high accuracy, provided that the parametric form of the model is correctly chosen. Hence, parameter risk from the estimation of the parameters in the diffusion components of the multi-asset model is very moderate and the parameter risk-captured bid-ask spreads are relatively narrow. A completely different picture results for the estimation of the spike size. Due to the threshold filtering technique, the number of spikes is relatively small ($\tilde{M} = 41$), which naturally enlarges the variance of the asymptotic distribution. Furthermore, the jump size distribution is crucial for ensuring that the real option representing the gas power plant gets deep into the money: When there is a large upward spike in the power price process, the payoff of the real option immediately jumps deeply into the money. Hence, the probability of producing large upward spikes in the power price plays a major role in determining the future value of a gas power plant. In case of Laplace distributed jumps, due to the different estimation of the location parameter and the fatter tails of the Laplace distribution, the option prices are significantly higher than in case of normally distributed jumps. Hence, the relative width of the bid-ask spread is wider for normally distributed jumps when regarding parameter risk arising from the base signals.

To ensure the stability of the Monte Carlo simulation, we additionally examine the bid-ask prices as a function of the Monte Carlo samples, which we exemplarily illustrate in the Figures 5.7a-5.10b. One can see that the numerical stability of the results is generally provided, but is more sensitive in the assessment of parameter risk for the jump size distribution, particularly for the case where a Laplace distribution is assumed for the jump size.

	Jumps size distribution					
	Gaussian			Laplace		
	α_1	α_2	α_3	α_1	α_2	α_3
Jump distribution	111.9%	73.71%	33.51%	163.5%	107.7%	48.96%
Gas, power base, and carbon	8.2%	5.5%	2.6%	3.9%	2.6%	1.5%
Gas and power base	6.5%	4.3%	1.9%	3.1%	2%	0.9%

Table 5.2 Resulting values for the relative width of the AVaR-induced parameter risk-captured bid-ask spread, concerning the parameter risk in the jump size distribution and the diffusion components. The significance levels for the AVaR are $\alpha_1 = 0.01$ (the highest risk-aversion), $\alpha_2 = 0.1$, and $\alpha_3 = 0.5$. One can see that the relative width of the parameter risk-captured bid-ask spread for the jump size parameters is much higher than for the diffusion parameters, even when estimating all diffusion parameters jointly.

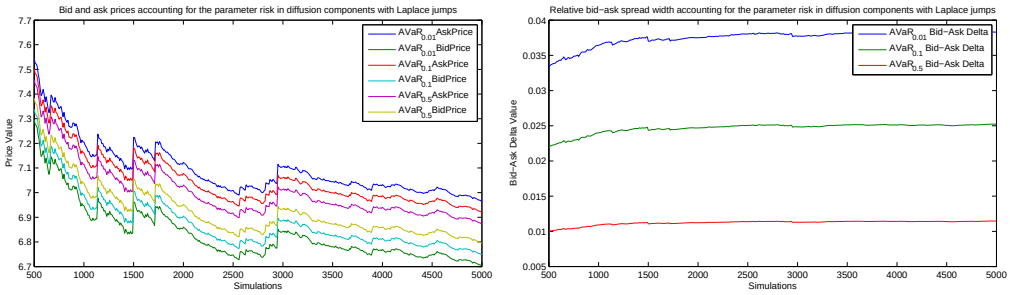
As a concluding result, we obtain that, in the given setting, the correct determination of the jump size parameters primarily drives the estimation risk-captured prices. One has to remark that the presented numbers are actually lower bounds of the assessed estimation risk, since we simplified the setting and disregarded many sources of estimation risk like the filtering procedure and the estimation of the seasonal trend parameters. Thus, the obtained numbers provide a hint which parameters are crucial for the evaluation, but do not exhaust the present parameter risk.



(a) Parameter risk-captured bid and ask prices for different significance levels when changing **diffusion parameter** values, normal jumps.

(b) Relative width of the risk-captured bid-ask spread for different significance levels when changing **diffusion parameter** values, normal jumps.

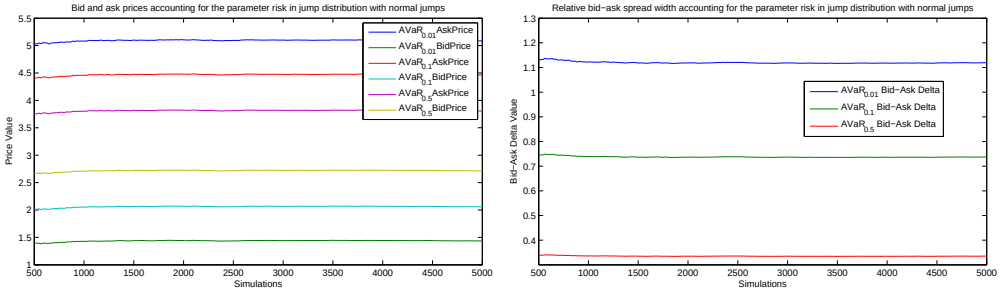
Figure 5.7 Parameter risk-implied bid-ask spread w.r.t. the diffusion components, normal jumps. The parameter risk is moderate and the result is numerically stable.



(a) Parameter risk-captured bid and ask prices for different significance levels when changing **diffusion parameter** values, Laplace jumps.

(b) Relative width of the risk-captured bid-ask spread for different significance values when changing **diffusion parameter** values, Laplace jumps.

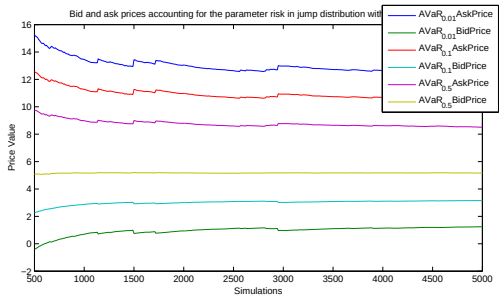
Figure 5.8 Parameter risk-implied bid-ask spread w.r.t. the diffusion components, Laplace jumps. The parameter risk is moderate, but the result is numerically less stable than when employing normal jumps.



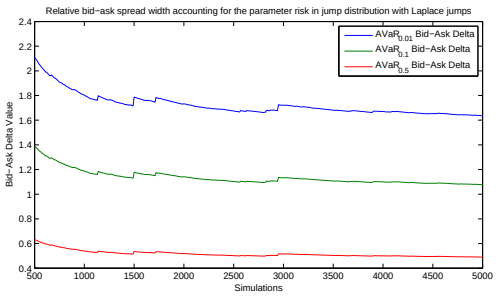
(a) Risk-captured bid and ask prices for different alpha levels when changing the **jump size distribution parameter** values, normal jumps.

(b) Relative width of the risk-captured bid-ask spread for different alpha levels when changing the **jump size distribution parameter** values, normal jumps.

Figure 5.9 One can see that the parameter-risk implied bid-ask spread w.r.t. the jump size distribution is crucial – the correct determination of the jump size distribution is the major driver for the power plant price. The results are numerically stable.



(a) Risk-captured bid and ask prices for different alpha levels when changing the **jump size distribution parameter** values, Laplace jumps.



(b) Relative width of the risk-captured bid-ask spread for different alpha values when changing the **jump size distribution parameter** values, Laplace jumps.

Figure 5.10 One can see that the parameter-risk implied bid-ask spread w.r.t. the jump size distribution is crucial – the correct determination of the jump size distribution is the major driver for the power plant price. Furthermore, convergence is much slower in the Laplace case than in the normal case and the results become much less numerically stable.

6 Application: Parameter risk induced by calibration to market prices

In the previous chapter, we have discussed parameter risk which is induced by historical estimation, i.e. where the risk can be expressed as the pushforward measure of some estimator $\hat{\theta}$. In these cases, there are plenty of situations where the estimator's distribution is (at least asymptotically) available and the knowledge of the estimator's distribution can be exploited according to our suggested ansatz. While historical estimation is a possibility for obtaining parameters, it is usually not the most preferred one by practitioners. First, one can often see that the estimated parameters (e.g. the Black–Scholes volatility from historical estimation) does not coincide with parameters obtained from forward-looking instruments (as, e.g., implied volatilities of liquid options) in reality. Second, practitioners prefer forward-looking instruments since they directly impose a "market-implied model". Third, since the model is supposed to be served for derivatives valuation, one would like to obtain the risk-neutral parameters. For some parameters (in particular: diffusion coefficients like the Black–Scholes volatility, or the correlation in a two-dimensional Black–Scholes model), the coefficients are not affected by some equivalent change of measure. But other parameters (as, e.g., the speed of mean reversion and the long-term variance in the Heston model, jump parameters in jump models) can have different values after an equivalent change of measure. Hence, their historically estimated value may not be credible any more in the situation of risk-neutral pricing. Thus, historical estimation is less popular compared to methods that fit model prices of liquid forward-looking instruments to their given market prices.

These methods are usually subsumed under the phrase "calibration". In these cases, one has to have a liquid market for some forward-looking instruments (in the equity and FX markets, these are usually plain vanilla options), where prices of these instruments are readily available. Then, the model is fitted to the prices of these instruments in an optimization procedure where an error function is minimized and the best-fitting parameters are chosen. Hence, the parameter is given by estimation, but not necessarily by historical estimation.

Unfortunately, the calibration procedure bears several pitfalls. First, there may be many parameters minimizing the error function, leaving us ambiguous which parameter to choose. Second, observed market prices can be noised by illiquidity, quotation time lag, or asymmetric quotes, so incorporating just the minimizing parameter may be overconfident when low errors are observed for several parameters. Third, the exposure to numerical problems in the optimization procedure is pretty large, thus, one might end up in some local minima. Fourth, the choice of error function may influence the calibration results. Finally, the number of liquid instruments might be too small to identify the parameters of a complex model. Thus, using the resulting parameter vector from a standard calibration exposes derivative pricing to a certain type of parameter risk, which we call *calibration risk*.

In contrast to the situation of estimating parameters historically, calibration does not naturally yield a distribution on the parameter space. Furthermore, there is little knowledge about the asymptotic distributions of calibration-type estimators. Hence, in this section, we present a method to construct a distribution on the parameter set that is in line with the calibration's results. We suggest using an error function of model to market prices and to transform it to obtain a pricing error-implied distribution. First, we show some consistency properties the error function has to fulfill. Second, we focus on theoretical solutions for obtaining a continuous distribution on the parameter set (most financial market models have parameter sets having positive Lebesgue measure). Furthermore, we focus on the practical problem how to deal with this from a computational point of view, suggesting an algorithm how to obtain a discrete distribution on feasible parameter subsets. Finally, we exemplarily investigate three popular models: The stochastic volatility models of (Heston, 1993) and (Barndorff-Nielsen and Shephard, 2001) as well as the Variance Gamma model (see (Madan et al., 1998)). This is done by calculating calibration risk-captured prices for three exotic derivatives. This enables us to compare the exposure to parameter risk of the different models and also of the exotic options. There is related works in the literature: (Gupta and Reisinger, 2012) suggest a Bayesian procedure and proves consistency of Bayesian calibration estimators under some assumptions (i.e. independence and normality of noise in prices), while (Schoutens et al., 2004) compare the exotics prices of different models that are calibrated to the same dataset of plain vanilla options. On the other hand, (Detlefsen and Härdle, 2007) scrutinize the effect of using different error functions to measure the distance of model to market prices.

This chapter is written along the lines of Chapter 6 of the paper (Bannör and Scherer, 2013a) and is organized as follows: In Section 6.1, we recall the setting which is required

for calibration to market prices. Section 6.2 discusses how to create a density on the parameter set that is consistent with observed market prices, both from a theoretical and a computational point of view. Finally, in Section 6.3, we apply the discussed tools to a real-data calibration situation and scrutinize the parameter risk of different models as well as parameter risk of different exotics.

6.1 Calibration to market prices

In this section, we briefly summarize the given facts about calibration to market prices. Throughout the chapter, we assume that there exist liquid contingent claims $C_1, \dots, C_M$ (in reality, these claims will often be plain vanilla options), $M \in \mathbb{N}$, with known respective market prices $C_1^\star, \dots, C_M^\star \in \mathbb{R}$ that are used for calibration.

Crucial for calibration to market prices is a function measuring the error of model prices to market prices for some parameter vector θ. Therefore, we state an axiomatization for an error function.

Definition 6.1.1 (Error function)
Let $\Theta \subset \mathbb{R}^k$ be a parameter space. We call $\eta_{C_1,\dots,C_M} : \Theta \to \mathbb{R}_{\geq 0}$ an error function *if there is a componentwise monotone function $\tilde{\eta} : \mathbb{R}^M_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that η can be decomposed via*

$$\eta_{C_1,\dots,C_M}(\theta) = \tilde{\eta}(|\mathbb{E}_\theta[C_1] - C_1^\star|, \dots, |\mathbb{E}_\theta[C_M] - C_M^\star|)$$

with $\tilde{\eta}(0, \dots, 0) = 0$.[1]

There are several different error functions in use, we present some of them in the following example.

Example 6.1.2 (Possible error functions)

1. *A popular error function is the* root mean square error

$$\mathrm{RMSE}_{C_1,\dots,C_M}(\theta) := \left(\frac{1}{M} \sum_{j=1}^{M} (\mathbb{E}_\theta[C_j] - C_j^\star)^2 \right)^{\frac{1}{2}},$$

possibly standardized by some number, e.g. the mean market price $\overline{C^\star}$.

[1] For the sake of notation simplicity, we occasionally omit the dependence on the liquid contingent claims $C_1, \dots, C_M$.

2. *If $C_1 \ldots, C_M$ are European call options, i.e. $C_j = (S_{T_j} - K_j)^+$ for some $T_j, K_j > 0$, $j = 1, \ldots, M$, there is a mapping to implied Black–Scholes volatilities $C_j^\star \mapsto \sigma_j^{\text{impl}}$. For a fixed maturity $T > 0$ and fixed strike $K > 0$, the mapping is strictly monotone and one-to-one. Hence, implied Black–Scholes volatilities provide a method to quantify prices in a "unit-free" manner. Thus, denoting the implied volatility mapping for a strike K and maturity T by $\sigma_{\text{impl}}^{(K,T)} : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{\geq 0}$, $C^\star \mapsto \sigma_{\text{impl}}^{(T,K)}(C^\star)$, one can regard the* mean absolute deviation of implied volatilities

$$\text{MADIV}(\theta) = \frac{1}{M} \sum_{j=1}^{M} |\sigma_{\text{impl}}^{(K_j,T_j)}(\mathbb{E}_\theta[C_j]) - \sigma_{\text{impl}}^{(K_j,T_j)}(C_j^\star)|.$$

If Definition 6.1.1 is slightly extended to transform the measurement of the option prices by their respective implied volatitility function[2], the mean absolute deviation of implied volatilities can also considered as an error function.

Different error functions can be used for different purposes. Among practitioners, using some measure determined on implied volatilities such as the mean absolute deviation of implied volatilities (MADIV, cf. Example 6.1.2) is popular, since one typically quotes vanilla prices in implied volatility rather than in prices. Furthermore, implied volatilities provide a "unit-free measure" for the price of plain vanilla options that is not distorted by the options' moneyness. A detailed discussion of the effect of error function choice on calibration results and exotics pricing can be found in (Detlefsen and Härdle, 2007).

Given an error function, the standard calibration procedure works as follows: A numerical optimization procedure determines the error minimizing parameter

$$\theta_0 = \arg\min_{\theta \in \Theta} \eta_{C_1,\ldots,C_M}(\theta).$$

Afterwards, the parameter θ_0 is used as the model's "true" parameter for pricing exotics. Note that the result of this procedure might be seen as a Dirac distribution on Θ_0 with all probability mass concentrated on the parameter θ_0.

6.2 Parameter risk from calibration to market prices

As described in Chapter 3, risk-capturing functionals enable us to incorporate parameter risk into prices of contingent claims. Required for the application of risk-capturing

[2] Since such error functions are not applied in this thesis, we stick to Definition 6.1.1, where derivatives prices are always measured in monetary units.

functionals is the existence of a distribution on the parameter space. Unfortunately, this distribution is typically not known in a standard calibration setting. In this section, we therefore present possible constructions of a distribution on Θ, being consistent with the calibration result.

A theoretical approach

Since a distribution on Θ does not arise naturally in a calibration framework (opposed to the case of estimation, where the estimator's distribution is often at hand), it has to be constructed from available information. The information used in a calibration stems from the error function η, so a natural starting point is a suitable transformation of the error function to imply a Lebesgue density on the parameter space Θ. Having found such a transformation, the induced density can be used as a distribution on the parameter space that is consistent with the calibration result. Hence, we try to find a function h such that $h \circ \eta$ is a proper Lebesgue density on Θ, reflecting pricing errors in a consistent manner. We formally describe which properties a transformation h of the error function has to fulfill.

Properties 6.2.1 (Transformation function requirements)
Let $\eta : \Theta \to \mathbb{R}_{\geq 0}$ be an error function. A transformation function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, creating a meaningful density from the error function η, has to fulfill the following properties:

1. *h is decreasing: This is important to guarantee that the density's values are in concordance with the error function's results. Parameters with lower aggregate errors to market prices should be attributed with a higher likelihood.*
2. *$\int_\Theta h(\eta(\theta))\,\mathrm{d}\theta = 1$: This normalization assures that $h \circ \eta$ indeed induces a distribution R on Θ via $R(\mathrm{d}\theta) = h(\eta(\theta))\,\mathrm{d}\theta$.*

The second property can easily be obtained by scaling, as long as $\int_\Theta h(\eta(\theta))\,\mathrm{d}\theta < \infty$. Obviously, there may be several decreasing functions $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, but not all of them may establish a finite integral $\int_\Theta h(\eta(\theta))\,\mathrm{d}\theta$.

Transformation of measures yields

$$\int_\Theta h(\eta(\theta))\,\mathrm{d}\theta = \int_0^\infty h(t)\,(\lambda \circ \eta^{-1})(\mathrm{d}t), \tag{6.1}$$

denoting by λ the Lebesgue measure on Θ. Equation (6.1) may be useful to exploit for considerations whether a function h is appropriate to induce a density on Θ. A first

observation reveals that if "too many parameters" (in the sense of Lebesgue measure) with low pricing errors exist, a Lebesgue density may not be established on Θ. Clearly, in this case we are exposed to massive parameter risk and model prices should be considered with uttermost caution.

Remark 6.2.2

Let $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be decreasing and $\delta \geq 0$ such that $h(\delta) > 0$. If the set of parameters having an error less than δ has infinite Lebesgue measure, i.e. $\lambda(\eta^{-1}([0, \delta])) = \infty$ with λ denoting the Lesbesgue measure on $\mathbb{R}^d$, then h cannot fulfill the properties described in 6.2.1 due to $\int_\Theta h(\eta(\theta))\,\mathrm{d}\theta = \infty$.

Proof

Let $\lambda(\eta^{-1}([0, \delta])) = \infty$. Then we immediately obtain

$$\begin{aligned}\int_\Theta h(\eta(\theta))\,\mathrm{d}\theta &= \int_0^\infty h(t)\,(\lambda \circ \eta^{-1})(\mathrm{d}t) \\ &\geq \int_0^\delta h(t)\,(\lambda \circ \eta^{-1})(\mathrm{d}t) \\ &\geq h(\delta) \cdot \lambda(\eta^{-1}([0, \delta])) = \infty.\end{aligned}$$

□

Geometrically speaking, this means that the error to market prices may not be too small for "too large" (in the sense of Lebesgue measure, particularly unbounded) sets of parameters.[3] We consider such situations to be pathological, since we do not want to incorporate this huge amount of parameter risk into prices via a Lebesgue-a.c. distribution. A situation where we have parameter risk to an unbounded set of parameters, a feasible choice which parameter to use and how to weight it does not make sense to evaluate the way we suggest in this work.

Furthermore, Remark 6.2.2 highlights that the choice of transformation function h may be limited. In some cases, the choice of transformation function is much more restricted. So, one can think of criteria which transformation functions may be suitable choices: To ensure that the transformation function h produces a proper density, one may choose

- a function with compact support,
- a function decreasing to zero fast enough.

We now present some functions which may be used as transformation functions and fall in one of the two categories above.

[3] This would typically happen when the parameters are underspecified, e.g. when sophisticated models with several parameters meet few liquid market prices to calibrate to.

Example 6.2.3 (Suitable transformation functions)
Decreasing functions which may ensure small values for parameters with large errors to market prices are the normal transformation function

$$h_\lambda^{\mathcal{N}}(t) := c \cdot \exp\left(-\left(\frac{t}{\lambda}\right)^2\right), \quad t \geq 0,$$

and the triangular transformation function

$$h_\lambda^{\Delta}(t) := c \cdot \mathbb{1}_{\{t \leq \lambda\}}\left(-\frac{t}{\lambda} + 1\right), \quad t \geq 0,$$

each equipped with a scaling parameter $\lambda > 0$ and $c > 0$ chosen such that the function induces a density. The scaling parameter determines the amount of weight which is assigned to parameters with low pricing error.

A particular class of functions, incorporating exactly the desired properties is the class of *Schwartz functions*, which we introduce in the following definition (cf. (Werner, 2011, Definition V.2.3)).

Definition 6.2.4
Let $a, b \in \mathbb{R}$, $a < b$. A function $f : [a,b] \to \mathbb{R}$, $f \in \mathcal{C}^\infty([a,b])$, is called a Schwartz function*, if for every $\alpha > 1$ and every $k \in \mathbb{N}$, the function*

$$\gamma_{\alpha,k}(t) : [a,b] \to \mathbb{R}, \quad \gamma_{\alpha,k}(t) := t^\alpha \cdot \frac{\partial^k f(t)}{\partial t^k}$$

is bounded. Usually, the set of all Schwartz functions defined on the interval $[a,b]$ is denoted by $\mathcal{S}([a,b])$.

The interpretation of Schwartz functions is that they decrease very fast to zero – every derivative of a Schwartz function falls faster than any polynomial. Obviously, if we denote the infinitely differentiable functions with compact support by $\mathcal{C}_c^\infty$, the relationship $\mathcal{C}_c^\infty \subset \mathcal{S}^\infty \subset \mathcal{C}^\infty$ holds. The normal transformation function of Example 6.2.3 (and, if we extend the definition of Schwartz functions to weak derivatives, also the triangular) is a Schwartz function, i.e. it can be shown that for every $\alpha > 1$ and every $k \in \mathbb{N}$, the function $t \mapsto t^\alpha \cdot \partial^k h_\lambda(t)/\partial t^k$ is bounded. Thus, Schwartz functions decrease faster to zero than every polynomial increases to infinity. We denote the set of all Schwartz functions on $[0,\infty)$ by $\mathcal{S}([0,\infty))$. A detailed discussion about Schwartz functions can be found in (Werner, 2011) and (Hörmander, 1990). A sufficient criterion for a Schwartz function to serve as a transformation function for the error function is presented in the following proposition.

Proposition 6.2.5

Let a transformation function $h \in \mathcal{S}([0,\infty))$ be decreasing and the "acceptable error area map" $t \mapsto (\lambda \circ \eta^{-1})[0,t]$, mapping an error level t to the Lebesgue measure of the parameter area where all parameters have an aggregate error to market prices of at most t, be differentiable. If the "error area" grows at most polynomial, i.e.

$$\frac{\partial}{\partial t}(\lambda \circ \eta^{-1})[0,t] \in \mathcal{O}(t^{\alpha}),\ \alpha \geq 0,$$

then $\int_{\Theta} h(\eta(\theta))\,\mathrm{d}\theta < \infty$ and h qualifies for inducing a density on Θ.

Proof

Transformation of the integral yields

$$\int_{\Theta} h(\eta(\theta))\,\mathrm{d}\theta = \int_{0}^{\infty} h(t)\,(\lambda \circ \eta^{-1})(\mathrm{d}t).$$

By differentiability of the Lebesgue–Stieltjes measure $\lambda \circ \eta^{-1}$, it follows

$$= \int_{0}^{\infty} h(t) \cdot \frac{\partial}{\partial t}(\lambda \circ \eta^{-1})[0,t]\,\mathrm{d}t.$$

The finiteness of the integral is equivalent to $h \mapsto \int_0^{\infty} h(t)\,(\lambda \circ \eta^{-1})(\mathrm{d}t)$ being a continuous linear form, thus the measure $\lambda \circ \eta^{-1}$ is in the topological dual space $\mathcal{S}'([0,\infty))$. These linear forms are known as tempered distributions. In particular, if $t \mapsto \frac{\partial}{\partial t}(\lambda \circ \eta^{-1})[0,t]$ grows at most polynomial, the functional

$$h \mapsto \int_{0}^{\infty} h(t) \cdot \frac{\partial}{\partial t}(\lambda \circ \eta^{-1})[0,t]\,\mathrm{d}t = \int_{\Theta} h(\eta(\theta))\,\mathrm{d}\theta$$

is a tempered distribution. Thus, the integral is finite. □

Interpreting the mapping $t \mapsto \frac{\partial}{\partial t}(\lambda \circ \eta^{-1})[0,t]$ geometrically, it means that the marginal error area may not grow faster (with growth meaning growth of Lebesgue measure) than polynomial. Furthermore, when considering the proof of Proposition 6.2.5, the assumptions can be weakened. Differentiability does not have to hold, the only crucial condition is that the Lebesgue-Stietjes measure $\lambda \circ \eta^{-1}$ on $[0,\infty)$ induces a tempered distribution.

Computational issues

In practice, however, the function $t \mapsto \frac{\partial}{\partial t}(\lambda \circ \eta^{-1})[0,t]$ may not be easy to calculate, since the error function η may be cumbersome and usually involves complicated evaluations, e.g. for the Heston model, the semi-closed form price of vanilla options involves the Fourier transform of a quite complicated characteristic funtion. Thus, the error function may only be evaluated on selected parameters $\theta_1, \dots, \theta_N \in \Theta$, $N \in \mathbb{N}$, on a specified grid. A further convenient advantage of a discretization of the parameter space is that finiteness of the integral in (6.1) is always assured. Although discretizing may not lead to large deviations from the integral in a sufficiently small neighborhood of $\theta_1, \dots, \theta_N$ when discretizing fine enough, it may not cover the integral over Θ if Θ is unbounded. Thus, we have to select a compact $\Theta_0 \subset \Theta$ and discretize Θ_0.

Algorithm 1 (Discrete density on a compact parameter space)
Given market prices $C_1^\star, \dots, C_M^\star$ of liquid securities $C_1, \dots, C_M$, a continuous error function η, and a continuous decreasing function h, the discrete density on a compact parameter space Θ_0 can be calculated in the following way:

1. *Apply a standard calibration algorithm to determine $\theta_0 = \arg\min_{\theta \in \Theta_0} \eta(\theta)$.*
2. *Discretize Θ_0 by choosing $\theta_1, \dots, \theta_N \in \Theta_0$ such that there is a $\delta > 0$ such that for every $\theta \in \Theta_0$ there is an $i \in \{0, \dots, N\}$ with $\|\theta - \theta_i\| < \delta$ (which is possible due to the compactness of Θ).*
3. *Calculate $h(\eta(\theta_i))$, $i = 0, \dots, N$.*
4. *Define $f(\theta_i) = h(\eta(\theta_i)) / \sum_{j=1}^N h(\eta(\theta_j))$ as the (discrete) density on Θ_0.*

The last steps in Algorithm 1 may be computationally expensive, especially if the error function η involves time-consuming computations (as it often does). Fortunately, after specifying and discretizing Θ_0, it is completely parallelizable. Thus, by distributing it to several kernels, the performance may easily be improved.

Obviously, the choice of a compact set $\Theta_0 \subset \Theta$ bears a lot of discretion. Experienced traders with sufficient information about calibrated parameters in the past may evaluate the parameters on a suitable cuboid $\Theta_0 = [\theta_1^{\min}, \theta_1^{\max}] \times \cdots \times [\theta_k^{\min}, \theta_k^{\max}]$. Others may choose Θ_0 as a regular ball centered at θ_0 w.r.t. a norm on $\mathbb{R}^k$, so $\Theta_0 = \overline{B_\delta(\theta_0)} \cap \Theta$ for some $\delta > 0$. From an economic point of view, incorporating existing bid-ask spreads from liquid options can be a reasonable choice. So, if the error function η represents a mean relative deviation from market prices, choosing $\Theta_0 = \eta^{-1}[0, \delta]$ for some $\delta > 0$ makes sense.

Unfortunately, chosen this way, it is not guaranteed that Θ_0 is bounded. Furthermore, the determination of a preimage of a sophisticated function may be challenging. Finally, Θ_0 may be disconnected, situations can arise where there may be many feasible "islands" in the parameter space. However, we have developed an algorithm determining the connected component containing θ_0 of $\eta^{-1}[0, \delta]$.

Algorithm 2 (Computation of a discretized preimage)
Let η be a continuous error function representing mean relative deviation from market prices, $\delta > 0$, and θ_0 the calibration result parameter with $\eta(\theta_0) < \delta$. We furthermore assume that $\eta^{-1}[0, \delta]$ is bounded and denote the connected component of $\eta^{-1}[0, \delta]$ containing θ_0 by Θ_0. Then the following algorithm returns a discretization of Θ_0.

1. *Apply a standard calibration algorithm to determine $\theta_0 = \arg\min_{\theta \in \Theta_0} \eta(\theta)$.*

2. *Choose a discretization step width $s \in \mathbb{R}^k_{\geq 0}$ (different directions might have different step widths).*

3. `pointsToSurround` := $\{\theta_0\}$
   `pointsToCalculate` := $\emptyset$
   `evalPairs` := $\{(\theta_0, \eta(\theta_0))\}$
   `surroundedPoints` := $\emptyset$
   `insidePoints` := $\{\theta_0\}$
   `outsidePoints` := $\emptyset$

4. `while (pointsToSurround` $\neq \emptyset$`) do {`
   `for each` $\theta \in$ `pointsToSurround {`
   `pointsToCalculate := pointsToCalculate` $\cup$ $\{(\theta_1 + \{-1, 0, 1\} \cdot s_1, \ldots, \theta_k + \{-1, 0, 1\} \cdot s_k)\}$
   `}`
   `pointsToCalculate := pointsToCalculate` $\cap$ Θ
   `pointsToCalculate := (pointsToCalculate - insidePoints) - outsidePoints`[4]
   `for each` $\theta \in$ `pointsToCalculate {`
   $y := \eta(\theta)$
   `if (`$y \leq \delta$`) then {`
   `insidePoints.Add(`θ`)`
   `evalPairs.Add(`(θ, y)`)`
   `}`

[4]We denote the set difference operator by a regular minus sign.

```
      else outsidePoints.Add(θ)
    }
    surroundedPoints := surroundedPoints ∪ pointsToSurround
    pointsToSurround := insidePoints − surroundedPoints
  }
```

The set `evalPairs` *contains the discretized set* Θ_0 *and the evaluations of the error function* $\eta(\Theta_0)$.

Unfortunately, this algorithm is only partially parallelizable, since the iterated sets have to be changed in the outer loop. However, the inner loop evaluating the error function can be parallelized if concurrency can be avoided when writing on the sets (e.g. by storing them in a database).

6.3 Case study: Comparing parameter risk of different models and exotic options

In this section, we apply the techniques described above and calculate bid-ask spreads of different exotic options induced by parameter risk, calculated by risk-capturing functionals as proposed in Chapter 3. In the present work, however, we induce a distribution on the respective parameter set by evaluating the errors to market prices of a Heston model, a Barndorff-Nielsen–Shephard model, and a Variance Gamma model. This enables us to calculate parameter risk-captured prices for different exotic options (similar to (Schoutens et al., 2004) and (Jessen and Poulsen, 2012)), to compare parameter risk of the three models, and to profile exotics w.r.t. their parameter risk.

Our universe of liquid securities consists of $M = 887$ DAX plain vanilla call options with different strikes and maturities as of February 26, 2009, at a spot $S_0 = 3942$. We use the popular FFT method (see (Carr and Madan, 1999)) to calculate vanilla prices for a whole strike grid simultaneously. We apply Algorithm 2 to calculate the connected component of a preimage w.r.t. a specified error function. Afterwards, we evaluate different exotic options (discrete barrier option, Asian option, discrete lookback option) via Monte Carlo simulation and calculate $\mathrm{AVaR}_{0.05}$-induced bid-ask prices implied by parameter risk for different densities.

We choose the relative root mean squared error

$$\eta(\theta) = \frac{\mathrm{RMSE}_{C_1,\dots,C_M}(\theta)}{\bar{C}^\star},$$

i.e. the mean deviation from mean market prices of the securities $C_1, \dots, C_M$ with market prices $C_1^\star, \dots, C_M^\star$ as error function. This error function is comfortable to interpret and calibration is equivalent to a calibration with RMSE, since it is just a strictly monotone transformation of RMSE. We transform η by the normal transformation function

$$h_\lambda^{\mathcal{N}}(t) = c \cdot \exp\left(-\left(\frac{t_0 - t}{\lambda}\right)^2\right), \quad t \geq t_0,$$

to a density with $t_0 := \eta(\theta_0)$, denoting by θ_0 the calibration parameter, scaling parameter $\lambda > 0$ and $c > 0$ matching such that a density is obtained. We correct the centered transformation functions for the minimal encountered error to market prices t_0. First, this corrects for advantages in calibration performance of one model compared to another. Second, the correction ensures that the induced distributions $R_\lambda(\mathrm{d}\theta) = h_\lambda^{\mathcal{N}}(\eta(\theta))\,\mathrm{d}\theta / \int_\Theta h_\lambda^{\mathcal{N}}(\eta(\theta))\,\mathrm{d}\theta$ converges weakly to δ_{θ_0} for $\lambda \to 0$. Since the AVaR-induced risk-capturing functionals fulfill the convergence property (CP) (cf. Proposition 4.1.3), we obtain the calibration plug-in price in the limit $\lambda \to 0$.

The Heston model

Heston introduced his popular model in (Heston, 1993), modifying the classical Black–Scholes model by modeling the square of Black–Scholes volatility using a Cox–Ingersoll–Ross process. The risk-neutral dynamics of the *Heston model* follow the SDEs

$$\begin{aligned} \mathrm{d}S_t &= rS_t\,\mathrm{d}t + \sigma_t S_t\,\mathrm{d}W_t^{(1)}, \\ \mathrm{d}\sigma_t^2 &= \kappa(\sigma_{\text{long}}^2 - \sigma_t^2)\,\mathrm{d}t + \xi\sigma_t\,\mathrm{d}W_t^{(2)}, \\ \mathrm{d}W_t^{(1)}\mathrm{d}W_t^{(2)} &= \rho\,\mathrm{d}t, \end{aligned}$$

$(W_t^{(i)})_{t\geq 0}$, $i = 1, 2$, being correlated Brownian motions and $r, S_0, \sigma_0^2, \kappa, \sigma_{\text{long}}^2, \xi > 0$, $\rho \in [-1, 1]$. Assuming that the spot S_0 and risk-free rate r are quoted in the market, parameter uncertainty arises from the quintuple $(\sigma_0^2, \kappa, \sigma_{\text{long}}^2, \xi, \rho)$ that has to be calibrated. Furthermore, to avoid the occurence of negative volatilities, we require our parameters to fulfill the so-called *Feller condition* $\xi^2 \leq 2\kappa\sigma_{\text{long}}^2$, which guarantees that the process $(\sigma_t^2)_{t\geq 0}$ is strictly positive (cf. (Feller, 1951)). Hence, our parameter space for the Heston model is

$$\Theta_{\text{Heston}} = \{(\sigma_0^2, \kappa, \sigma_{\text{long}}^2, \xi, \rho) \in \mathbb{R}_{>0}^4 \times [-1, 1] : \xi^2 \leq 2\kappa\sigma_{\text{long}}^2\}.$$

For FFT pricing à la (Carr and Madan, 1999) we need the characteristic function of the log-price process $X_t := \log S_t$. It is given as in (Schoutens et al., 2004) by

$$\begin{aligned}\log \phi_{X_t}^{\text{Heston}}(u) = &iu(\log S_0 + rt) + \frac{\sigma_0^2(a-c)(1-\exp(-ct))}{\xi^2(1-g\exp(-ct))} \\ &+ \frac{\kappa\sigma_{\text{long}}^2}{\xi^2}\left((a-c)t - 2\log\left(\frac{1-g\exp(-ct)}{1-g}\right)\right)\end{aligned}$$

with $a := \kappa - \rho\xi ui$, $c := \sqrt{a^2 - \xi^2(-ui-u^2)}$, and $g := (a-c)/(a+c)$. It may be noted that there are two specifications for the Heston characteristic function as pointed out in (Albrecher et al., 2007), we have chosen the (numerically) less pathological one.

The Barndorff-Nielsen–Shephard model

Barndorff-Nielsen and Shephard developed a stochastic volatility model in which the variance process is a subordinator-driven Ornstein–Uhlenbeck process (cf. (Barndorff-Nielsen and Shephard, 2001; Nicolato and Venardos, 2003)). One of the most common specifications for the subordinator is a compound Poisson process with exponentially distributed jump sizes as in (Schoutens et al., 2004), resulting in the Γ-*OU-Barndorff-Nielsen–Shephard model*, which will be used in the further. To account for the leverage effect, the stock price has a negative jump whenever the volatility has an upward jump. The dynamics of the log-price process $X_t := \log S_t$ are governed by the following SDEs:

$$\begin{aligned}\mathrm{d}X_t &= \left(r - \frac{\sigma_t^2}{2} + \frac{\lambda c\rho}{\alpha+\rho}\right)\mathrm{d}t + \sigma_t\,\mathrm{d}W_t - \rho\,\mathrm{d}Z_{\lambda t}, \\ \mathrm{d}\sigma_t^2 &= -\lambda\sigma_t^2\,\mathrm{d}t + \mathrm{d}Z_{\lambda t},\end{aligned}$$

with parameters $r, S_0, \sigma_0^2, \lambda, c, \rho, \alpha > 0$, $(W_t)_{t\geq 0}$ is a Brownian motion and $(Z_t)_{t\geq 0}$ is a compound Poisson process with exponentially distributed jump size, i.e. $Z_t = \sum_{j=1}^{N_t} U_j$ with a Poisson process $(N_t)_{t\geq 0}$ with intensity $c > 0$ and $(U_j)_{j\in\mathbb{N}}$ are exponentially i.i.d. with parameter $\alpha > 0$, $(W_t)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ are independent. Given the observable spot price S_0 and risk-free rate r, the unspecified parameters with exposure to uncertainty is the quintuple $(\sigma_0^2, c, \alpha, \lambda, \rho)$. Thus, without any further restrictions, our parameter space for the Γ−OU-Barndorff-Nielsen–Shephard model[5] is

$$\Theta_{\text{BNS}} = \{(\sigma_0^2, c, \alpha, \lambda, \rho) : \sigma_0^2, c, \alpha, \lambda, \rho \in \mathbb{R}_{>0}\} = \mathbb{R}_{>0}^5.$$

[5] We abbreviate this by BNS model.

The characteristic function of the log-price process is (cf. (Nicolato and Venardos, 2003; Schoutens et al., 2004; Cont and Tankov, 2004))

$$\log \phi_{X_t}^{\mathrm{BNS}}(u) = iu\left(\log S_0 + \left(r + \frac{c\lambda\rho}{\alpha+\rho}\right)t\right) - gh\sigma_0^2 + c\frac{\alpha \log \frac{\alpha - f_1}{\alpha - \rho u i} + f_2 \lambda t}{\alpha - f_2}$$

with

$$g := \frac{u^2 + ui}{2}, \quad h := \frac{1 - \exp(-\lambda t)}{\lambda}, \quad f_1 := \rho u i - gh, \quad f_2 := \rho u i - \frac{g}{\lambda}.$$

The Variance-Gamma model

The *variance gamma model* is an exponential Lévy model without a diffusion component.[6] In this pure jump model, the log-returns are following the variance gamma process. This model was introduced in (Madan and Senata, 1990) for modeling returns, (Madan et al., 1998) introduced an option pricing model based on the variance gamma process. The variance gamma process is a special case of a time-changed Brownian motion with drift

$$Z_t = \vartheta \Lambda_t + \sigma W_{\Lambda_t}$$

with $\vartheta \in \mathbb{R}$, $\sigma > 0$, $(W_t)_{t\geq 0}$ is a Brownian motion and $(\Lambda_t)_{t\geq 0}$ is a Lévy subordinator with $\Lambda_0 = 0$. In case of the variance gamma process, the subordinator $(\Lambda_t)_{t\geq 0}$ is a Gamma process, i.e. for $h > 0$ and $t \geq 0$ the increments $\Lambda_{t+h} - \Lambda_t$ follow a $\Gamma(h/\kappa, \kappa)$-distribution. Thus, under a risk-neutral measure, the log-price process X_t is given (compare (Madan et al., 1998)) by

$$\mathrm{d}X_t = \left(r + \frac{1}{\kappa}\log\left(1 - \vartheta\kappa - \frac{\sigma^2\kappa}{2}\right)\right)\mathrm{d}t + \mathrm{d}Z_t.$$

Like the other two models, the variance gamma model can be regarded as an extension of the classical Black–Scholes model. Instead of making volatility stochastic, time is made stochastic, as it was first proposed in (Clark, 1973). The characteristic function of the log-price process is calculated in (Madan et al., 1998) as

$$\phi_{X_t}^{\mathrm{VG}}(u) = \exp\left(iu\left(\log(S_0) + t\left(r + \frac{\log\left(1 - \vartheta\kappa - \frac{\sigma^2\kappa}{2}\right)}{\kappa}\right)\right)\right)\left(1 - i\vartheta\kappa u + \frac{\sigma^2\kappa u^2}{2}\right)^{-\frac{t}{\kappa}}.$$

[6] For abbreviation, we will frequently use the short name VG model.

Error function calculation

Since the characteristic function of the log-prices of all three models is known, we use the Fourier transform pricing method as described in (Carr and Madan, 1999). Provided that the βth moment, $\beta > 0$, of the log-price exists, it is shown in (Carr and Madan, 1999) that the price of a plain vanilla call option $C(K)$ at maturity $T > 0$ with strike $K > 0$ is given by

$$C(K) = \frac{1}{\pi K^\beta} \int_0^\infty \exp(-\log(K)vi)\psi_T(v)\,\mathrm{d}v$$

with

$$\psi_T(v) = \frac{\exp(-rT)\phi_T(v-(\beta+1)i)}{\beta^2+\beta-v^2+(2\beta+1)vi},$$

denoting by ϕ_T the characteristic function of the log-price at time T. We choose (as in (Schoutens et al., 2004)) $\beta = 0.75$. The usage of the Fast Fourier Transform (FFT) allows us to calculate call prices on a large strike grid simultaneously (cf. (Carr and Madan, 1999)). We interpolate between the given strikes to obtain prices for the strikes from our data.

A case study on parameter risk

We use Algorithm 2 to obtain a dense grid of parameters, all of them having a pricing error smaller than the smallest pricing error t_0 plus 2%. This enables us to compare the pure parameter risk effect w.r.t. the same transformation function. For the Heston and Barndorff-Nielsen–Shephard models we have chosen not to observe parameter risk arising from the short-term volatility σ_0^2 to reduce the parameter space to four dimensions (cf. (Guillaume and Schoutens, 2011)). The following table gathers the result of the algorithm and the minimal relative RMSE to market prices, denoting by t_0 the minimal error to market prices.

On the obtained parameter set, we evaluate three exotic options: An ITM discrete barrier call option, an arithmetic Asian call option, and a discrete lookback call option. An ITM discrete barrier call option with maturity $T > 0$, strike $K > 0$, barrier $B > 0$, and observation points $0 < t_1 < \cdots < t_L =: T$, $L \in \mathbb{N}$, is given by the payoff

$$Z_{\text{Barrier}} = \mathbb{1}_{\bigcap_{l=1}^{L}\{S_{t_l}<B\}}(S_T - K)^+.$$

Model	Discretization step vector	♯ of feasible parameters	t_0
Heston model	$\begin{pmatrix} 0 & 0.07 & 0.1 & 0.07 & 0.02 \end{pmatrix}$	5 868	1.44%
BNS model	$\begin{pmatrix} 0 & 0.1 & 0.5 & 0.15 & 0.2 \end{pmatrix}$	9 536	1.79%
VG model	$\begin{pmatrix} 0.04 & 0.04 & 0.02 \end{pmatrix}$	6 705	5.41%

Table 6.1 The calibration environment of the different models to compare with the according discretization step vectors, the resulting number of feasible parameters, and the calibration error. One can see that the calibration performance of the Variance Gamma model is much lower than the calibration performance of the two stochastic volatility models.

An arithmetic Asian call option with maturity $T > 0$, strike $K > 0$, and observation points $0 < t_1 < \cdots < t_L =: T$, $L \in \mathbb{N}$, is given by the payoff

$$Z_{\text{Asian}} = \left(\frac{1}{L} \sum_{l=1}^{L} S_{t_l} - K \right)^+ .$$

Finally, the payoff of a discrete lookback call option with maturity $T > 0$, strike $K > 0$, and observation points $0 < t_1 < \cdots < t_L =: T$, $L \in \mathbb{N}$, is given by

$$Z_{\text{Lookback}} = \bigvee_{l=1}^{L} (S_{t_l} - K)^+ .$$

The payoff always takes place at maturity. In our case, for all three options we choose as maturity $T = 1$, $K = 4\,000$ as strike, $L = 24$ equidistant observation points and, in case of the barrier option, a barrier at $B = 5\,000$. For simplicity, we disregard dividends and assume the risk-free rate to be $r = 0.03$.

The contingent claims described above are path-dependent, we evaluate them by Monte Carlo simulation. Therefore, we draw accordingly distributed random numbers with the same seed and rescale them with the different parameters. This prevents us from mixing parameter risk with noise risk due to differently drawn random numbers. In our Monte Carlo simulation, we evaluate using 10 000 sample paths and 24 observation points per year.

6.3.1 Parameter risk in different models

Since we have evaluated the above mentioned contingent claims in different models, we can empirically evaluate the amount of model specific parameter risk. Parameter risk

Asian option	Heston model	BNS model	VG model
Modus (Calibration plug-in value)	311.8868	316.1214	274.9964
Expected value	311.4427	316.0965	274.2531
(integrating out the distribution on Θ)			
Coefficient of variation	0.0125	0.0152	0.0348
Skewness	0.2707	−0.3451	−0.2017

Table 6.2 Overall, the coefficient of variation is smallest for the Asian price distributions. In particular, the dispersion of the BNS model distribution is not much larger than the dispersion of the Heston model distribution. The VG model implies much different option prices (due to bad fit to ATM vanilla options) and considerable higher dispersion. Furthermore, the jump model price distributions are slightly skewed to the left, while the Heston model prices have a slight skew to the right.

is reflected in the width of bid and ask prices calculated according to Chapter 3. Our methodology to create a distribution on the parameter set allows us to scrutinize the pushforward distributions of the prices. The probability distributions can be visualized in scatterplots. For a proper visual model comparison, we calculate and compare the resulting cumulative distribution functions of the option prices and compare the $\text{AVaR}_{0.05}$-induced bid-ask spreads for the different models when varying the scaling parameter λ. The cumulative distribution functions of the different option price distributions are shown in Figures 6.1, 6.3, and 6.5. The parameter risk-captured bid-ask prices are depicted together with scatterplots of the price distributions for the respective models in the Figures 6.2, 6.4, and 6.6. Furthermore, we calculate statistical properties of the distributions and compare these in Tables 6.2-6.4. We calculate the modus (which essentially is the plug-in calibration price), the expected value, the coefficient of variation, and the skewness. We have decided to apply the coefficient of variation for measuring dispersion due to its invariance w.r.t. multiplicative factors. Since we compare prices with different levels, the coefficient of variation enables us to directly compare their dispersion. The expected value references to the AVaR_1-induced parameter risk-captured price, which was suggested for incorporating parameter risk by (Lindström, 2010).

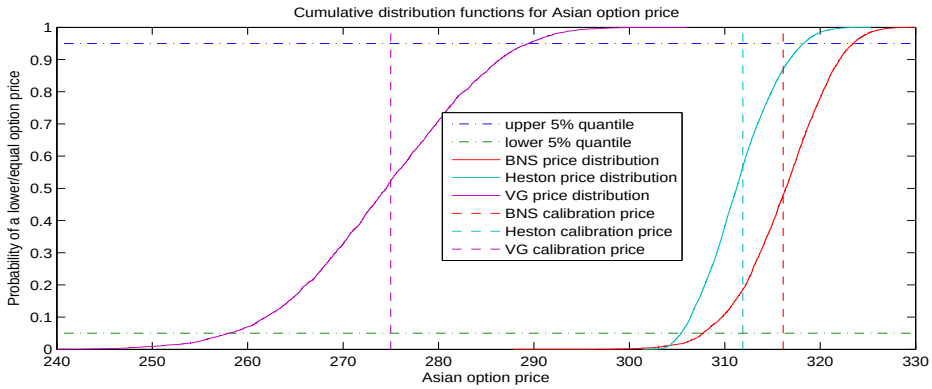


Figure 6.1 The cumulative distribution functions of Asian option prices induced by the normal transformation function with scaling parameter $\lambda = 0.008$. One can see that the Heston price distribution observes slightly lower dispersion than the BNS price distribution. Furthermore, the BNS price distribution is moderately skewed to the left. Overall, the Heston model is less exposed to parameter risk in this case. The VG price distribution is a lot more dispersed and the peak of the distribution is quite different. This is due to the large calibration error of the VG model, in particular when fitting to ATM prices.

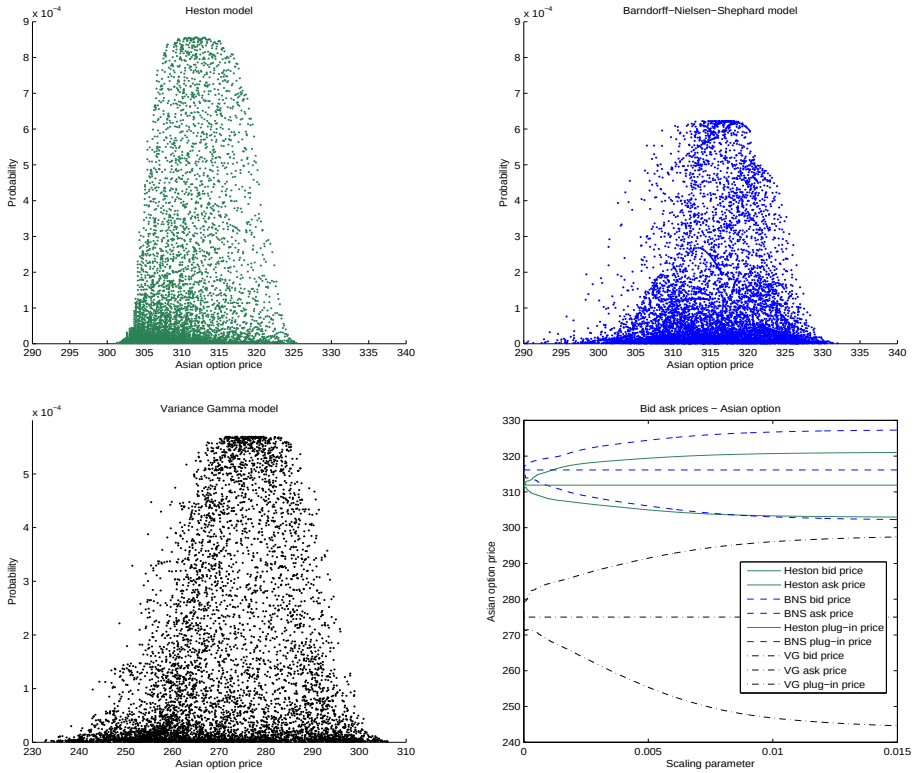


Figure 6.2 The scatterplots of the price distributions for the Asian option, employing a scaling parameter $\lambda = 0.008$, and Asian option parameter risk-captured bid-ask prices as a function of the scaling parameter. The bid-ask prices of the BNS and Heston models are quite close, while the VG model bid-ask spread is broader and has a different location.

Barrier option	Heston model	BNS model	VG model
Modus (Calibration plug-in value)	77.5529	97.9619	82.5580
Expected value	75.0691	92.1595	98.8515
(integrating out the distribution on Θ)			
Coefficient of variation	0.0420	0.1257	0.2650
Skewness	−0.0156	0.2889	1.6040

Table 6.3 The dispersion of the barrier price distribution is (regardless of the model) a lot higher than for the Asian and lookback option due to higher sensitivity. In particular, the coefficients of variation are a lot higher in the BNS model than in Heston model and the dispersion of the VG model is by far the highest among all models. While the Heston model price distribution observes almost no skewness, the BNS model price distribution is slightly skewed to the right and the VG model price distribution has very strong skewness, compared to the other price distributions.

The Heston model

The Heston model shows narrow-peaked price distributions for all three options with little variance. Overall, the price distributions are quite symmetric (skewness is close to zero). Compared to the price distributions arising in the Barndorff-Nielsen–Shephard model, the distribution has lower coefficients of variation for all three exotics. This may be partly attributed to the better calibration performance, so values with worse calibration performance are incorporated in the BNS model as well. Furthermore, the Heston model, being completely driven by multivariate Brownian motion, has continuous paths, while the Barndorff-Nielsen–Shephard model (in our specification) is also driven by a compound Poisson process and allows for discontinuities in the stock price process.

The BNS model

In contrast to the Heston model, the Barndorff-Nielsen–Shephard model delivers distributions for all three options without a clear single peak and considerably larger dispersion. This means that many parameters deliver an equally good fit to market prices as the parameter obtained from the standard calibration, but differ substantially in the calculated prices for exotics. We attribute this result to the strong focus of the jump

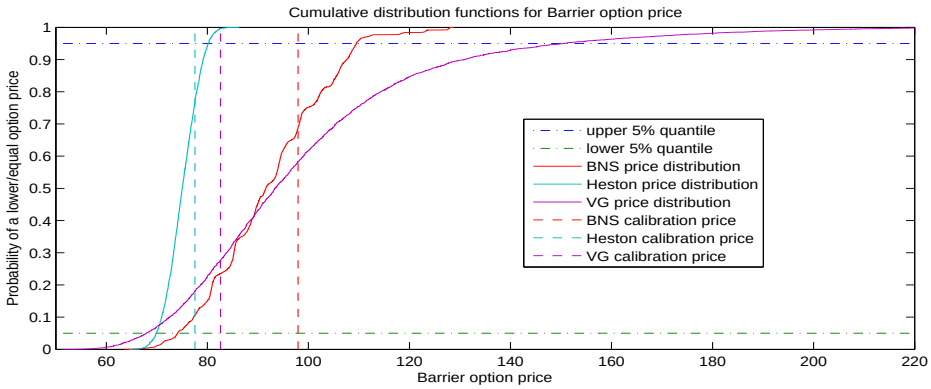


Figure 6.3 The cumulative distribution functions of ITM barrier option prices induced by the normal transformation function with scaling parameter $\lambda = 0.008$. Dispersion for ITM Barrier options is a lot higher for the BNS model than for the Heston model, thus parameter risk for barrier options is higher in the BNS model than in the Heston model. Furthermore, due to the discontinuity of the payoff function, prices in the Barndorff-Nielsen–Shephard model jump significantly. Furthermore, BNS model prices are considerably higher than Heston model prices. This may be due to downward jumps, making stock prices not breaking the barrier. Jumps are also very likely to be the reason for the snatchy shape of the BNS distribution function. The VG model prices are heavily skewed to the right and have much higher variability, due to their poor calibration performance.

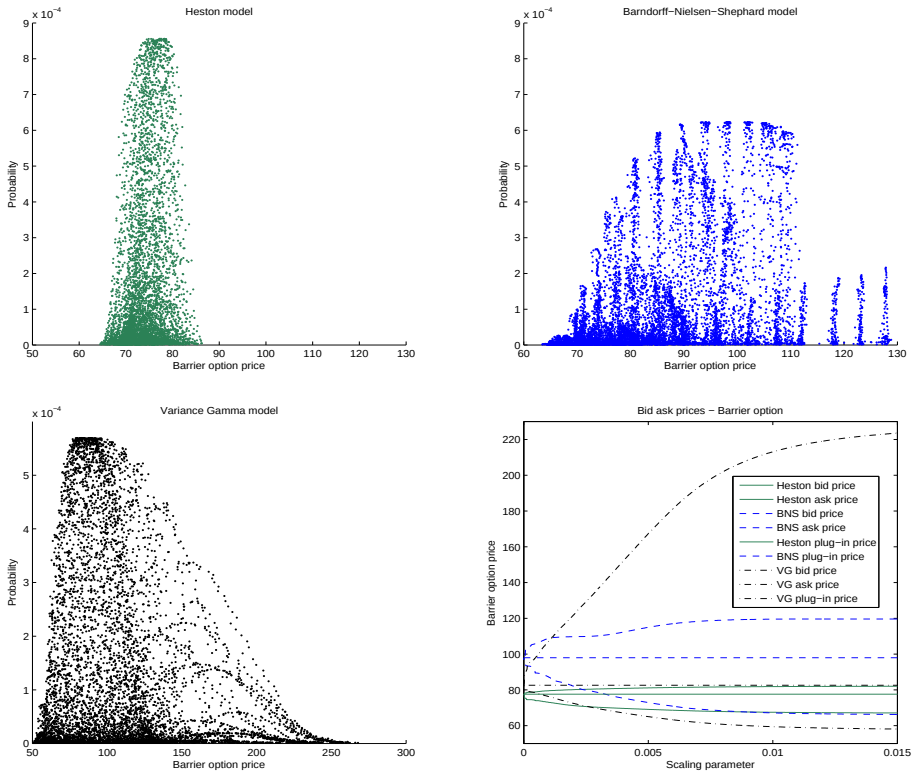


Figure 6.4 The scatterplots of the price distributions for the ITM barrier option, employing a scaling parameter $\lambda = 0.008$, and ITM barrier option parameter risk-captured bid-ask prices as a function of the scaling parameter. The bid-ask prices of the VG model are clearly lower/higher than the induced bid-ask prices of the Heston and BNS models. The methodology of calculating the preimage results in a special structure in the distribution scatterplots of the BNS and VG models, causing certain parameter combinations to be feasible, while other paths are filtered out due to too high errors to market prices.

Lookback option	Heston model	BNS model	VG model
Modus (Calibration plug-in value)	847.9149	837.7727	840.3980
Expected value	852.1520	856.3299	824.4740
(integrating out the distribution on Θ)			
Coefficient of variation	0.0127	0.0380	0.0638
Skewness	0.2621	−0.1539	−0.2031

Table 6.4 The coefficient of variation of the BNS model distribution is a lot higher than in the Heston model. Indeed, in the Heston model, the lookback price distribution has comparable dispersion to the Asian price distribution, while variability is higher in the jump models. Furthermore, the sign of skewness in the Heston model is positive, while in both jump models skewness is negative (probably due to jumping down). The VG model shows by far the highest coefficient of variation.

component to replicate smile and term structure of given market prices. Only slightly different parameters concerning the jump component may result in considerable price differences, especially concerning more sensitive contingent claims like lookback or barrier options. It is observed that the Barndorff-Nielsen–Shephard model - considered isolated - bears substantial parameter risk, even if only parameters with very low differences are incorporated. Cont and Tankov observe in (Cont and Tankov, 2004, pp. 488–490) that the Barndorff-Nielsen–Shephard model has problems to fit to several smiles, thus it has some problems with calibration performance. This is verified by our observations. Furthermore, the calibration result seems to be quite unstable in the sense that there are many parameters matching equally well. From a computational point of view, this is a drawback of the Barndorff-Nielsen–Shephard model. From an economic point of view, the BNS model contains a higher degree of parameter risk than the Heston model.

The most striking differences occur for the ITM barrier option. Not only is the price of the ITM barrier option significantly higher in the BNS model than in the Heston model, which may be due to the downward pressure of the downward-directed jumps. Furthermore, observed parameter risk in the BNS model is huge. Besides larger parameter risk, there are significant differences regarding the lookback option: In the Barndorff-Nielsen–Shephard model, the distribution is clearly skewed to the left. This may be the result of the jump component, allowing jumps only to move stock prices downwards, while in the Heston model, the distribution looks quite symmetric.

The Variance Gamma model

The VG model's calibration performance is a lot weaker than the calibration performance of the Heston and the BNS model. This is consistent with the results in (Jessen and Poulsen, 2012) which state that the calibration performance of the VG model is inferior to that of the Heston model and ties approximately with the classical Black–Scholes model. From this point of view, it plays in a different league than the Heston and the Barndorff-Nielsen–Shephard model. The mediocre calibration performance is reflected in the high degree of parameter risk, as one can see. The price distributions for all three options are tremendously wider than the price distributions for the BNS and Heston model. Many observed parameter constellations that match the criterion for the VG model have quite poor performance in matching market prices, thus, in case of lookback and ITM barrier options, this mismatch may produce extreme results. In case of the barrier option price, it is heavily skewed to the right. Since the VG model did not match ATM options well, it is no surprise that the price of the Asian option deviates a lot from the prices in the BNS and the Heston model.

6.3.2 Parameter risk profiles of different exotics

Comparing the parameter risk w.r.t. the different types of options, we mainly expect the parameter risk profiles of exotics to follow their sensitivities: Contingent claims with high deltas and high Black–Scholes vegas are supposed to show greater parameter risk, since stock price and volatility are supposed to be captured by the models. Figure 6.7 compares the relative deviation of $\mathrm{AVaR}_{0.05}$ induced bid-ask prices to the plug-in prices for different options within the Heston model, the BNS model, and the VG model.

It can be easily identified that the parameter risk of the Asian option is quite small in all models. This may be attributed to the averaging taking place in Asian payoff profiles, so calculating prices with slightly different parameters does not lead to much different results. On the other hand, the risk profile of an Asian option is close to the risk profile of a plain vanilla option, delta and vega tend to be lower on an absolute basis. Thus, since all parameter constellations on average match vanilla market prices quite good, too large deviations from parameters are not expected.

For lookback options, the Heston model does not bear substantially higher parameter risk than for Asian options. Parameter risk is more skewed to the ask price, this can be explained by higher maximums being covered with the risk-captured prices. This is

mainly due to the sensitivity of the maximum of the stock price to parameter changes. For the BNS model, the lookback option suffers from more parameter risk, especially on the ask side. The parameter risk of the VG model is skewed a lot more. Bid prices deviate significantly more than the ones in the BNS model, while ask prices do not.

The ITM barrier option shows large variability in the prices, mainly due to the knock-out feature. For some parameter constellations, the price becomes quite sensitive, since a high probability of stock prices hitting the barrier cause the option price to fall. Overall, we have an option with highly varying delta and vega. Compared to the other options, the parameter risk is substantially higher for both the Heston and the BNS model. In the BNS model, incorporating parameter risk delivers a bid price of more than 30% less than the plug-in price. Even worse deviations can be observed for the VG model. While the bid price deviation can be compared to the one from the BNS model, the ask price has enormous parameter risk - there are values that are 150% higher than the plug-in price.

In a nutshell, our observations meet our expectations concerning the degree of parameter risk. Options with low deltas and Black–Scholes vegas (e.g. Asian options) seem to be quite robust w.r.t. parameter risk. On the other hand, setting the price for lookback and barrier options classically may bear large amounts of parameter risk. Calculating risk-captured prices may give useful hints for setting bid-ask spreads, they should be narrow for good-natured exotics like Asians and wide for ITM barrier options.

6.4 Outlook

In this chapter, we have presented a method to induce a distribution on the parameter space by employing a transformation of the error function of a calibration to market prices. To select a transformation is highly subjective, roughly spoken, the transformation of the error function controls the “trustworthiness” of a given error to market prices. One may argue that the error to market prices measured by the error function is not sufficient to describe whether a parameter (vector) is trustworthy in the sense of assigning it a certain likelihood. One possibility could be that, e.g., a minimum fit of all options (in terms of relative price deviation or implied volatility) is regarded as a necessary condition to assign any weight to some parameter vector. Actually, this practice would even go beyond the typical calibration techniques that are used in industry and would enhance some “constrained calibration” technique, where an error function has to

be minimized additionally fulfilling some constraints (which may be given e.g. on the single prices).

Furthermore, one may develop a numerical scheme yielding a truly continuous distribution on the parameter space. In our case, a discretization scheme is employed, which does not capture the parameters that do not lie on the predefined vector grid. One way to obtain a continuous distribution on the parameter could be realized by employing density kernels on the grid. This could lead to several problems as, e.g., the support of the density kernels have to be contained in the parameter space, which may be challenging in many cases (e.g. the correlation in the Heston model has support in $[-1, 1]$, the Feller condition in the Heston model implies a nonlinear inequality condition on the parameters).

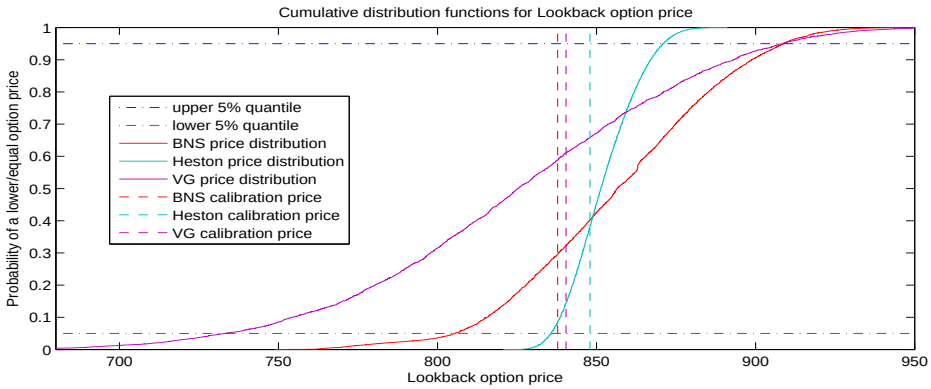


Figure 6.5 The cumulative distribution functions of lookback option prices induced by the normal transformation function with scaling parameter $\lambda = 0.008$. For lookback options, the BNS model also shows considerably higher dispersion than the Heston model. The VG model observes much more variability compared to the other models. This may be directly related to the results of the barrier option, since little stock price maximums make barriers not to break.

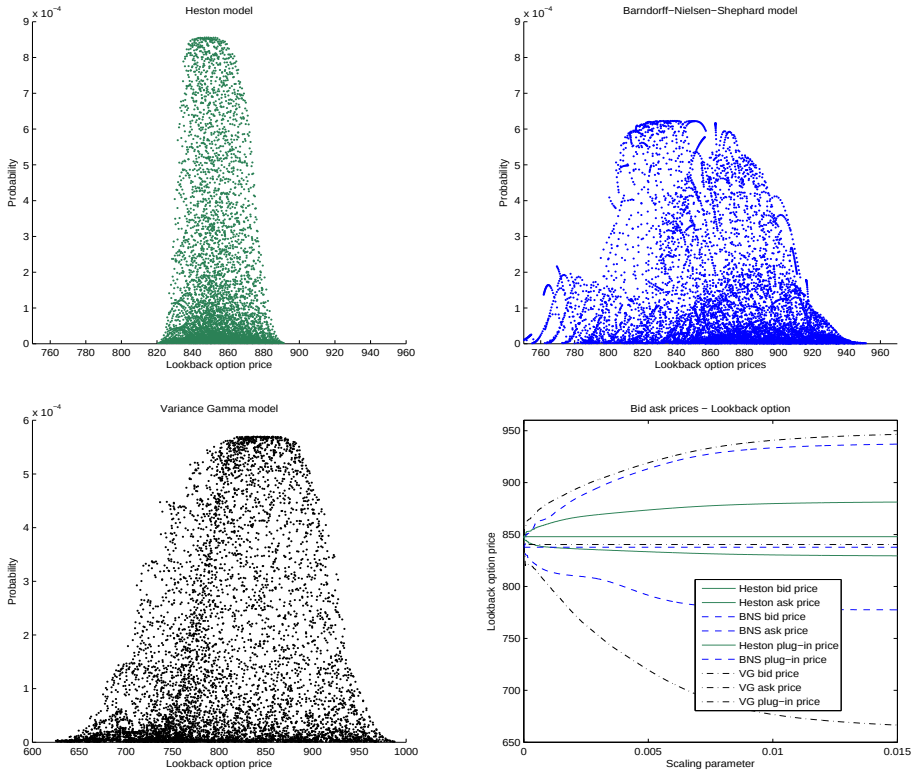


Figure 6.6 The scatterplots of the price distributions for the lookback option, employing a scaling parameter $\lambda = 0.008$, and lookback option parameter risk-captured bid-ask prices as a function of the scaling parameter. The bid-ask prices of the VG model clearly dominate the induced bid-ask prices of the Heston and BNS models, although their plug-in prices are very close.

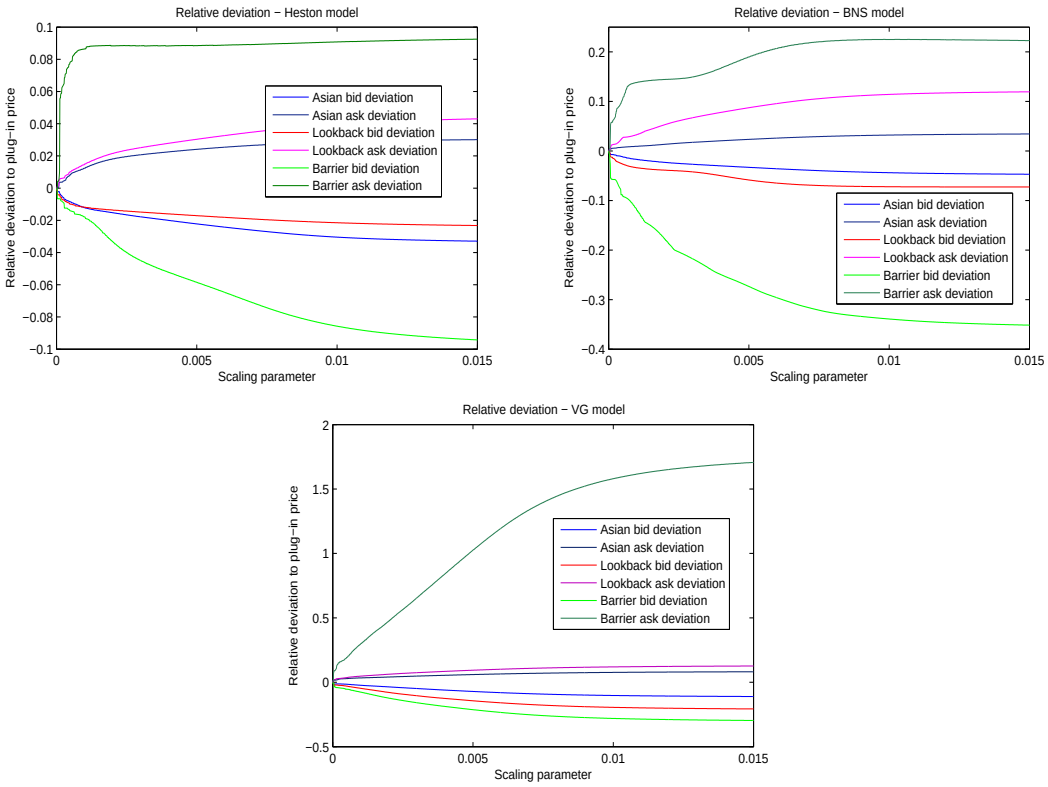


Figure 6.7 Relative deviations from the plug-in price compared among different option types for the Heston, Barndorff-Nielsen–Shephard, and Variance Gamma models. In the Heston model, both Asian and lookback options observe low parameter risk. For lookback options, parameter risk is skewed to the ask side. In the BNS model, Asian options observe low parameter risk, but lookbacks have considerable parameter risk. For both models, parameter risk is largest for the ITM barrier option, the BNS model observes large deviations ($> 20\%$) from the plug-in price. The Variance Gamma model has higher deviations from the plug-in price. An extreme deviation can be observed for the ITM barrier option ask price. Here we have deviations of $> 150\%$ of the original price.

7 Calibration of risk-captured model prices to bid-ask market prices

7.1 Introduction

Several authors (see, e.g., (Carr et al., 2001; Xu, 2006; Bion-Nadal, 2009; Cherny and Madan, 2010)) have suggested modeling bid-ask prices with convex risk measures, so does this thesis in Chapter 3. All approaches have in common that a certain convex risk measure $\Gamma := \Gamma^{\text{ask}}$ is employed for the calculation of ask prices. Moreover, the dual functional Γ^{bid}, defined by $\Gamma^{\text{bid}}(X) := -\Gamma^{\text{ask}}(-X)$, is usually used for calculating bid prices. In our case, we stick to this kind of "bid-ask symmetry".

(Carr et al., 2001) and (Cherny and Madan, 2010) generalize the classical results from incomplete markets, suggesting pricing based on coherent risk measures. In incomplete markets with real-world measure P, a set of "stress-test measures" $\mathcal{Q}$ is selected from the set of all equivalent martingale measures and the ask price of a contingent claim X is determined as the supremum of the expectations w.r.t. the stress-test measures[1]:

$$\Gamma(X) := \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[X].$$

In particular, (Cherny and Madan, 2010) suggest choosing the set of stress test measures by an acceptability index. Their result consists of using coherent risk measures induced by a parametric family of concave distortion functions $(\gamma_\lambda)_{\lambda \in \Lambda}$ with

$$\Gamma_\lambda(X) := \int X \, \mathrm{d}(\gamma_\lambda \circ P).$$

This approach is described in (Cherny and Madan, 2010) as "conic finance" and we will adopt this name.

[1] (Cont, 2006) discuss a similar supremum-based approach not restricting to an incomplete markets setting.

A different approach to bid-ask prices, also relying on convex risk measures, is presented in Chapter 3: Here, bid-ask spreads are explained in the context of parameter risk of a parametric family of martingale measures $(Q_\theta)_{\theta\in\Theta}$. In presence of a distribution R on the parameter space Θ, one can define for contingent claims X and law-invariant, normalized convex risk measures ρ the parameter risk-captured price, which may serve as an ask price, by

$$\Gamma(X) := \rho(\theta \mapsto \mathbb{E}_\theta[X]),$$

denoting the risk-neutral price w.r.t. a fixed parameter $\tilde{\theta} \in \Theta$ by $\mathbb{E}_{\tilde{\theta}}[X]$.

Obviously, we can restrict the choice of risk measures to distortion risk measures, since they are law-invariant and normalized. Thus, using a concave distortion γ, the risk-captured ask price turns into

$$\int_\Theta \mathbb{E}_\theta[X]\,(\gamma \circ R)(\mathrm{d}\theta).$$

In the following, we discuss the calibration to bid-ask prices that are created in a distortion risk measure-driven environment. The conic finance approach of (Cherny and Madan, 2010) can be embedded into the parameter risk framework, since it restricts parameter risk to the market price of risk. To cover both ideas in a unified way, we formulate the framework in a general manner and write $f(X)$ (f being a linear map w.r.t. some matching vector spaces) as a symbol for the integrand under the Choquet integral w.r.t. the distorted probability and R as symbol for the probability measure. Hence, the generic ask pricing formula is given by

$$\int f(X)\,\mathrm{d}(\gamma \circ R). \tag{7.1}$$

In the parameter risk-capturing framework described in Chapter 3, $f(X) = \mathbb{E}_\theta[X]$ and R is the probability measure on the parameter space Θ. In the conic finance environment as described in (Cherny and Madan, 2009), we identify $f(X) = X$ and R is a (previously specified) risk-neutral measure Q.[2]

7.2 The bid-ask calibration problem

For classical linear pricing systems, the calibration of liquid instruments $C_1, \ldots, C_M$ to their quoted mid prices $C_1^{\star\,\mathrm{mid}}, \ldots, C_M^{\star\,\mathrm{mid}}$ is a common way of obtaining the model's unobservable parameters; thus obtaining a market-implied distribution (from a previously

[2] Due to the general formulation, we omit the area of integration in the generic ask pricing formula (7.1).

selected family) for the expectation. Consequently, when extending to non-linear pricing systems as briefly summarized in Section 7.1, it is natural to think about obtaining a market-implied convex risk measure from quoted bid-ask prices. As a starting point, (Cherny and Madan, 2010) develop semi-closed-form solutions for the calculation of bid and ask prices in their conic finance approach. Using these semi-closed-form representations, they calibrate to given bid-ask prices employing, e.g., the *minmaxvar*-family (cf. Example 2.2.10) of distortion functions.

In this section, we restate the bid-ask calibration problem in a formal way, using the generic language of (7.1), and show that there exists a solution in a mildly restricted class of distortion functions, not restricted to a fixed parametric shape of the distortion function. As a corollary, we show the existence of a solution of the bid-ask calibration problem in the AVaR- and *minmaxvar*-family of distortion functions.

We start with a general formulation of the bid-ask calibration problem.

Problem 7.2.1 (Bid-ask calibration problem)
Let $C_1, \dots, C_M$ be contingent claims (e.g. plain vanilla options) with given market bid-ask quotes $(C_1^{\star\text{bid}}, C_1^{\star\text{ask}}), \dots, (C_M^{\star\ \text{bid}}, C_M^{\star\ \text{ask}})$. Let furthermore $\eta : \mathbb{R}_{\geq 0}^{2M} \to \mathbb{R}_{\geq 0}$ be an error function[3] measuring the pricing error between model prices and market quotes.

A convex risk measure $\tilde{\Gamma}$ solves the (symmetric) bid-ask calibration problem *on a domain $\mathcal{G}$, if $\tilde{\Gamma}$ minimizes the function*

$$\begin{aligned}\Gamma \mapsto \eta\big(&|-\Gamma(-C_1) - C_1^{\star\text{bid}}|, \dots, |-\Gamma(-C_M) - C_M^{\star\ \text{bid}}|,\\ &|\Gamma(C_1) - C_1^{\star\text{ask}}|, \dots, |\Gamma(C_M) - C_M^{\star\ \text{ask}}|\big)\end{aligned}$$

over the set of admissible functionals $\mathcal{G}$.

The formulation of the bid-ask calibration in Problem 7.2.1 is quite general. Analogously to the calibration to mid prices, we want the model bid-ask prices to fit market quoted bid-ask prices as good as possible. Therefore, we rely on minimizing an error function, but now aggregate both the errors of the model bid and ask prices to the respective market prices to ensure a consistent calibration to bid and ask prices.

Since the class of convex risk measure functionals can be very large and a numerical solution may be difficult to obtain, we restrict ourselves to the set of distortion risk measures. As described in Chapter 2, they represent a very tractable and rich class of convex risk measures. The tractability of distortion risk measures stems from the

[3] Recall the definition of an error function from Definition 6.1.1.

convenient representation as a Choquet integral w.r.t. a distorted probability (cf. 2.2.6). We formulate our results in the generic language of (7.1).

The first theorem provides the existence of a best fit to market prices in broadly defined subclasses of distortion risk measures. In short, Problem 7.2.1 is solvable.

Theorem 7.2.2 (Existence of a solution to the bid-ask calibration problem) *Let $K > 0$, η be a continuous error function, $f(X)$ be R-a.s. bounded and define*

$$G_K := \{\gamma : [0,1] \to [0,1] : \gamma \text{ Lipschitz continuous, concave distortion function with Lipschitz constant } K\}.$$

Furthermore, write $\Gamma_\gamma(X) := \int f(X)\,\mathrm{d}(\gamma \circ R)$ for $\gamma \in \mathcal{G}_K$. Then the bid-ask calibration problem has a solution in G_K, i.e. there exists a minimizing distortion function $\tilde{\gamma} \in G_K$ such that the function

$$\tilde{\eta} : \gamma \mapsto \eta\big(|-\Gamma_\gamma(-C_1) - C_1^{\star\,\mathrm{bid}}|, \ldots, |-\Gamma_\gamma(-C_M) - C_M^{\star\,\mathrm{bid}}|, |\Gamma_\gamma(C_1) - C_1^{\star\,\mathrm{ask}}|, \ldots, |\Gamma_\gamma(C_M) - C_M^{\star\,\mathrm{ask}}|\big)$$

is minimized in $\tilde{\gamma}$ in the following sense:

$$\tilde{\eta}(\tilde{\gamma}) \le \tilde{\eta}(\gamma) \qquad \text{for all } \gamma \in G_K.$$

Proof

Step 1: G_K is compact in the uniform topology.

We start by showing that G_K is $\|\cdot\|_\infty$-closed: Let $(\gamma_N)_{N\in\mathbb{N}}$ be a uniformly convergent sequence in G_K with (uniform) limit $\gamma := \lim_{N\to\infty}\gamma_N$. Since uniform convergence implies pointwise convergence, for each $x, y \in [0,1]$, $x \le y$, $\gamma(x) = \lim_{N\to\infty}\gamma_N(x) \le \lim_{N\to\infty}\gamma_N(y) = \gamma(y)$ holds. Furthermore, $\gamma(0) = \lim_{N\to\infty}\gamma_N(0) = 0$ and $\gamma(1) = \lim_{N\to\infty}\gamma_N(1) = 1$ hold. Hence, γ is a distortion function. Furthermore, it can be shown in a straightforward manner that γ is concave: Let $x, y \in [0,1]$ and $\lambda \in [0,1]$. Now

$$\begin{aligned}\gamma(\lambda x + (1-\lambda)y) &= \lim_{N\to\infty}\gamma_N(\lambda x + (1-\lambda)y)\\ &\ge \lim_{N\to\infty}\lambda\gamma_N(x) + (1-\lambda)\gamma_N(y) = \lambda\gamma(x) + (1-\lambda)\gamma(y)\end{aligned}$$

yields the concavity of γ.

It remains to show that γ is also Lipschitz continuous with Lipschitz constant $K > 0$: Let $\varepsilon > 0$ be arbitrary. Due to uniform convergence, there exists some $N \in \mathbb{N}$ such that

$\|\gamma - \gamma_N\|_\infty < \varepsilon/2$. *Thus, applying* $\gamma_N \in G_K$ *and the triangular inequality,*

$$|\gamma(x) - \gamma(y)| \leq |\gamma(x) - \gamma_N(x)| + |\gamma_N(x) - \gamma_N(y)| + |\gamma_N(y) - \gamma(y)| < K|x - y| + \varepsilon$$

for all $x, y \in [0,1]$. *Hence,* $G_K \subset \mathcal{C}[0,1]$ *is a closed set in the uniform topology.*

Trivially, $\|\gamma\|_\infty = 1$ *for all* $\gamma \in G_K$, *thus,* G_K *is uniformly bounded. Moreover, since all functions* $\gamma \in G_K$ *have the joint Lipschitz constant* K *by definition,* G_K *is also an equicontinuous set. Hence,* G_K *being a uniformly closed, uniformly bounded and equicontinuous set, the Arzelà–Ascoli theorem (e.g. (Werner, 2011, Theorem II.3.4)) yields that* G_K *is compact in the uniform topology on* $\mathcal{C}([0,1])$.

Step 2: The map $I_X : G_K \to \mathbb{R}$ *defined by* $\gamma \mapsto I_X(\gamma) = \int f(X)\,\mathrm{d}(\gamma \circ R)$ *is* $\|\cdot\|_\infty$*-continuous.*

Let $\gamma_N \to \gamma$ *and recall the definition of the Choquet integral from Definition 2.2.2. Dominated convergence (applicable due to the boundedness of* $f(X)$*) immediately delivers*

$$\begin{aligned}\int f(X)\,\mathrm{d}(\gamma_N \circ R) &= \int_{-\infty}^{0} \gamma_N(R(f(X) > s)) - 1\,\mathrm{d}s + \int_0^\infty \gamma_N(R(f(X) > s))\,\mathrm{d}s \\ &\to \int_{-\infty}^{0} \gamma(R(f(X) > s)) - 1\,\mathrm{d}s + \int_0^\infty \gamma(R(f(X) > s))\,\mathrm{d}s \\ &= \int f(X)\,\mathrm{d}(\gamma \circ R),\end{aligned}$$

hence $\gamma \mapsto \int f(X)\,\mathrm{d}(\gamma \circ R)$ *is continuous w.r.t. the uniform topology on* $\mathcal{C}([0,1])$.

Since η *is a continuous error function (thus* $\eta \circ I_X$ *is* $\|\cdot\|_\infty$*-continuous) and* G_K *is compact, there exists an element* $\tilde{\gamma} \in G_K$ *which minimizes the function* $\eta \circ (I_{C_1}, \ldots, I_{C_M})$ *for given contigent claims* $C_1, \ldots, C_M$. *Hence,* $\tilde{\gamma}$ *minimizes the error measured by* η *to given market bid-ask prices* $(C_1^{\star\,\mathrm{bid}}, C_1^{\star\,\mathrm{ask}}), \ldots, (C_M^{\star\,\mathrm{bid}}, C_M^{\star\,\mathrm{ask}})$, *i.e.* $\eta(I(\tilde{\gamma})) \leq \eta(I(\gamma))$ *holds for all* $\gamma \in G_K$. □

Restricting to the treated type of sets G_K from Theorem 7.2.2, $K > 0$, deals with the assumption that all considered distortion functions have a common maximum Lipschitz constant K. At first glance, this may look technical and one could question this restriction, but in practice, this assumption is not too restrictive and distortion functions behave smoothly on the majority of its domain: Every concave distortion function on $[0,1]$ is Lipschitz continuous on $[\varepsilon, 1]$ with constant $1/\varepsilon$ for every $\varepsilon \in (0,1)$ (furthermore,

it can only have a jump at zero and is continuous on $(0,1]$). Hence, for a numerical discussion and practical implementation, it is sufficient and rather natural to focus on sets of distortion functions which have a common Lipschitz constant.

Remark 7.2.3

1. *We will show that the* AVaR- *and* minmaxvar-*families described in Examples 2.2.9 and 2.2.10 have a common Lipschitz constant when the allowed significance level resp. the slope parameter are mildly restricted. Thus, these popular examples are captured by our calibration framework.*

2. *If we scrutinize the proof of Theorem 7.2.2, in the second part we have only used that $\gamma_N \to \gamma$ converges pointwise. In particular, we have shown that $\gamma \mapsto \int X \, \mathrm{d}(\gamma \circ R)$ is sequentially continuous w.r.t. the pointwise topology.*

From a practitioner's point of view, the class G_K is often too large to consider and one tries to solve the bid-ask calibration problem in smaller domains, which may be easier to parameterize (e.g. (Cherny and Madan, 2010) employ a model which is based on *minmaxvar* (cf. Example 2.2.10) and other families and calibrate it to market bid-ask prices). Therefore, the following immediate corollary provides a helpful sufficient criterion that the bid-ask calibration problem is solved in smaller classes.

Corollary 7.2.4

Let $G \subset G_K$ for some $K > 0$ (G_K as above) be $\|\cdot\|_\infty$-closed and η be a continuous error function. Then, the bid-ask calibration problem has a solution in G.

Proof

Since G is $\|\cdot\|_\infty$-closed and G_K is $\|\cdot\|_\infty$-compact, G is $\|\cdot\|_\infty$-compact. Thus, the same arguments as in Theorem 7.2.2 provide the existence of a minimizing element $\tilde{\gamma} \in G$. $\square$

As already mentioned, Corollary 7.2.4 can be applied to ensure the existence of a solution of the bid-ask calibration problem in some parametric family. As an example, we show that – under mild technical restrictions – the bid-ask calibration problem has a solution in the popular AVaR- and *minmaxvar*-families.

Example 7.2.5

1. *Let $\varepsilon > 0$ and define the ε-bounded* AVaR-*family*

$$G_\varepsilon := \left\{ \gamma : [0,1] \to [0,1] : \exists \alpha \in [\varepsilon, 1] : \gamma_{|[0,\alpha]}(y) = \frac{y}{\alpha} \text{ and } \gamma_{|[\alpha,1]}(y) = 1 \right\}.$$

From a statistical point of view, we disregard the AVaR-*distortions focusing only on the upper ε-tail of the price distribution.*

If we now select a convergent sequence $(\gamma_N)_{N\in\mathbb{N}}$, $\gamma_N \to \gamma := \lim_{N\to\infty}\gamma_N$ *and denote the adjacent significance levels by* $(\alpha_N)_{N\in\mathbb{N}}$ *resp.* α, *then obviously* $\alpha_N \to \alpha$ *holds and* $\alpha \geq \varepsilon$. *Summarizing,* $\gamma_N \in G_\varepsilon$ *for all* $N \in \mathbb{N}$ *implies* $\gamma \in G_\varepsilon$, *thus* G_ε *is closed. Applying Corollary 7.2.4 now immediately yields that the bid-ask calibration problem has a solution in* G_ε.

2. *Let* $K \in [0,\infty)$ *and define*

$$G_K := \left\{\gamma : [0,1] \to [0,1] : \exists L \in [0,K] : \gamma(y) = 1 - (1 - y^{\frac{1}{L+1}})^{L+1}\right\},$$

i.e. $\gamma \in G_K$ *is a* minmaxvar*-type distortion function with parameter* $L \in [0,K]$. *If* $(\gamma_N)_{N\in\mathbb{N}}$ *is a sequence in* G_K *that converges uniformly, it is easy to show that* $L_N \to L$, *in particular,* $\lim_{N\to\infty}\gamma_N \in G_K$. *Thus,* G_K *is* $\|\cdot\|_\infty$*-closed, applying Corollary 7.2.4 shows that the bid-ask calibration problem has a solution in* G_K.

7.3 A non-parametric calibration scheme for bid-ask price

In this section, we present a tractable and highly flexible alternative to the bid-ask calibration of parametric families à la (Cherny and Madan, 2010). We estimate the distortion function in a non-parametric way, using a piecewise linear approximation. Therefore, we remark that the set of piecewise linear distortion function is $\|\cdot\|_\infty$-dense in the set of continuous distortion functions. Furthermore, we derive convenient evaluation formulae for bid-ask pricing with piecewise linear distortion functions, which eventually reduce the bid-ask calibration problem to a constrained optimization problem on the unit cuboid. We argue that a non-parametric calibration scheme may be useful to obtain the shape of a market-implied distortion function: Empirical results on market-implied distortion functions are very rare, yet. Thus, a parametric approach may not capture the distortions appropriately that are quoted on the market. Furthermore, our approach can be used to find and justify some parametric shape of the market-implied distortion function from empirical observations.

7.3.1 General results

We start with the following lemma, ensuring that a piecewise linear approximation of a distortion function is a sensible approximation and does not produce much different distorted expectations.

Lemma 7.3.1 (Piecewise linear approximation)
Let $\gamma : [0,1] \to [0,1]$ be a continuous concave distortion function. Let $N \in \mathbb{N}$ and choose $0 =: y_0 < y_1 < \cdots < y_N := 1$. Denote $n(x) := \max\{n \in \{0,\dots,N\} : x \geq y_n\}$ and define

$$\gamma_N(x) := \begin{cases} \gamma(y_{n(x)}) + \frac{\gamma(y_{n(x)+1}) - \gamma(y_{n(x)})}{y_{n(x)+1} - y_{n(x)}}(x - y_{n(x)}) & , x \in [0,1) \\ 1 & , x = 1. \end{cases} \tag{7.2}$$

Then γ_N is a piecewise linear concave distortion function *and $\gamma_N \to \gamma$ uniformly, if the mesh of the decompositions $\mathfrak{Y}_N := \{y_0,\dots,y_N\}$ converge to zero, i.e. if $\max_{n\in\{1,\dots,N\}}\{y_n - y_{n-1}\} \to 0$ for $N \to \infty$.*

Proof
Let $N \in \mathbb{N}$. Obviously, γ_N is piecewise linear, monotone, and $\gamma_N(0) = 0$, $\gamma_N(1) = 1$. Thus, γ_N is a piecewise linear distortion function. Furthermore, γ_N is concave, since the a.e. defined derivative is decreasing. The convergence property follows from the proof that the set of piecewise linear functions is uniformly dense in the space of continuous functions on the unit interval $\mathcal{C}[0,1]$ (cf. (Shilov, 1996, Section 1.23)). □

As an immediate corollary, we obtain that for all $K > 0$, the set of piecewise linear distortion functions with maximum elevation K, is dense in the set G_K from Theorem 7.2.2. Thus, solving the bid-ask calibration problem in G_K^{lin} approximates a solution in G_K arbitrarily precise.

Corollary 7.3.2 (Piecewise linear approximation of calibration solutions)
Let $K > 0$ and G_K be as in Theorem 7.2.2, let $\tilde{\gamma}$ be a solution of the bid-ask calibration problem provided by Theorem 7.2.2, and define the set of piecewise linear distortion functions with maximum elevation K, denoted by

$$G_K^{\text{lin}} := \{\gamma \in G_K : \gamma \text{ is piecewise linear}\}.$$

Then the solution of the bid-ask calibration problem in G_K can uniformly be approximated in G_K^{lin}, i.e. there exists a sequence $(\gamma_N)_{N\in\mathbb{N}}$ in G_K^{lin} such that $\gamma_N \to \tilde{\gamma}$ uniformly.

Proof
Applying Lemma 7.3.1, we can construct a sequence $(\gamma_N)_{N\in\mathbb{N}}$ in G_K^{lin} converging to $\tilde{\gamma}$ uniformly. Since $\tilde{\gamma}$ is concave and the Lipschitz continuity ensures that the slope in each point does not exceed K, the slope of the piecewise linear approximation cannot exceed K either. Hence, the construction method in Lemma 7.3.1 guarantees that $\gamma_N \in G_K^{\text{lin}}$ for every $N \in \mathbb{N}$. □

Hence, treating the bid-ask calibration problem in G_K^{lin} is a sensible substitute for directly solving among all distortion functions in G_K.

Furthermore, the continuity of the error function assures that integrals w.r.t. piecewise linear concave distortion functions converge as well, so optimizing in the piecewise concave distortion functions may not provide significant disadvantages in accuracy compared to optimizing in G_K.

As a second result, we obtain that piecewise linear concave distortion functions are fully characterized by an N-tuple of real numbers (with constraints).

Lemma 7.3.3 (Alternative characterization)
Let $(y_0, \dots, y_N)$ be a decomposition of the unit interval, i.e. $0 = y_0 < y_1 < \cdots < y_N = 1$, and γ be a piecewise linear concave distortion function which is linear on the subintervals $[y_{n-1}, y_n]$ for all $n = 1, \dots, N$. Then the distortion function γ is fully characterized by the vector of differences on the decomposition points $\Delta\gamma \in \mathbb{R}^N$, where

$$\Delta\gamma_n := \gamma(y_n) - \gamma(y_{n-1}), \quad n = 1, \dots, N.$$

Furthermore, $\Delta\gamma$ exhibits the constraints

$$\frac{\Delta\gamma_1}{y_1 - y_0} \geq \frac{\Delta\gamma_2}{y_2 - y_1} \geq \cdots \geq \frac{\Delta\gamma_N}{y_N - y_{N-1}} \geq 0 \quad \text{and} \quad \sum_{n=1}^{N} \Delta\gamma_n = 1. \tag{7.3}$$

Proof
The vector $\Delta\gamma$ can be transformed to the different slopes by dividing each component $\Delta\gamma_j$ by $y_j - y_{j-1}$. Since piecewise linear functions are uniquely determined by the slopes, their grid, and their start and end value, the differences $\Delta\gamma$ fully describes the distortion function γ. Monotonicity, concavity, and $\gamma(0) = 0$, $\gamma(1) = 1$ immediately yield the constraints from the assertion. □

In particular, solving the bid-ask calibration problem can be traced back to finding the best matching $(\Delta\gamma_1, \dots, \Delta\gamma_N)$ fulfilling the conditions in (7.3), reducing the infinite-dimensional bid-ask calibration problem from G_K to a finite-dimensional problem. Moreover, piecewise linearity of the distortion function provides convenient expressions for the distorted expectation, hence for bid-ask pricing. The key result for calculating distorted expectations is presented in the following theorem.

Theorem 7.3.4 (Piecewise linear concave distorted expectations)
Let $N \in \mathbb{N}$, γ_N be a piecewise linear concave distortion function with adjacent decomposition $0 = y_0 < y_1 < \cdots < y_N = 1$ and difference vector $(\Delta\gamma_1, \dots, \Delta\gamma_N) \in [0,1]^N$. Let

furthermore $f(X) \in \mathcal{L}^\infty(R)$ and $y_0, \dots, y_N$ are points of continuity of $F_{f(X)}$. Then the distorted expectation of $f(X)$ w.r.t. $\gamma_N \circ R$ can be calculated via

$$\int X \,\mathrm{d}(\gamma_N \circ R) = \sum_{n=1}^{N} \Delta\gamma_n \mathbb{E}\Big[f(X) \Big| f(X) \in \big[\mathrm{VaR}_{y_n}(f(X)), \mathrm{VaR}_{y_{n-1}}(f(X)) \big] \Big].$$

For the proof of Theorem 7.3.4, we need the following lemma yielding a convenient decomposition for piecewise linear distortion functions.

Lemma 7.3.5 (Decomposition of piecewise linear distortion functions)
Let $N \in \mathbb{N}$, γ_N be a piecewise linear concave distortion function with adjacent decomposition $0 = y_0 < y_1 < \cdots < y_N = 1$, and difference vector $(\Delta\gamma_1, \dots, \Delta\gamma_N) \in [0,1]^N$. Define the slopes of γ by $\tilde{\gamma}_n := \Delta\gamma_n/(y_n - y_{n-1})$ and the help functions $f_n(y) := \min\{y_n, \max\{y_{n-1}, y\}\} - y_{n-1}$ for $n \in \{1, \dots, N\}$. Then γ_N can be decomposed into the slopes and help functions via

$$\gamma_N(y) = \sum_{n=1}^{N} \tilde{\gamma}_n f_n(y).$$

Proof
Let $\nu \in \{1, \dots, N\}$ be arbitrary and $y \in [y_{\nu-1}, y_\nu]$. By definition of $\tilde{\gamma}_n$ and f_n, $n \in \{1, \dots, N\}$, we obtain

$$\sum_{n=1}^{N} \tilde{\gamma}_n f_n(y) = \sum_{n=1}^{N} \frac{\gamma_N(y_n) - \gamma_N(y_{n-1})}{y_n - y_{n-1}} (\min\{y_n, \max\{y_{n-1}, y\}\} - y_{n-1}).$$

Since $y \in [y_{\nu-1}, y_\nu]$, the terms for $n > \nu$ cancel, resulting in

$$= \sum_{n=1}^{\nu} \frac{\gamma_N(y_n) - \gamma_N(y_{n-1})}{y_n - y_{n-1}} (\min\{y_n, \max\{y_{n-1}, y\}\} - y_{n-1}).$$

Furthermore, the terms for $n < \nu$ can be simplified, resulting in

$$\begin{aligned}
&= \sum_{n=1}^{\nu-1} (\gamma_N(y_n) - \gamma_N(y_{n-1})) + \frac{\gamma_N(y_\nu) - \gamma_N(y_{\nu-1})}{y_\nu - y_{\nu-1}} (y - y_{\nu-1}) \\
&= \gamma_N(y_{\nu-1}) + \frac{\gamma_N(y_\nu) - \gamma_N(y_{\nu-1})}{y_\nu - y_{\nu-1}} (y - y_{\nu-1}) \\
&= \gamma_N(y)
\end{aligned}$$

according to the construction in Lemma 7.3.1. □

Proof (of Theorem 7.3.4)
Applying Lemma 7.3.5, we decompose $\gamma_N = \sum_{n=1}^N \tilde{\gamma}_n f_n$ *with* $\tilde{\gamma}_n$, f_n, $n \in \{1\dots,N\}$ *as defined in Lemma 7.3.5. In particular, we obtain for* $f(X) \in \mathcal{L}^\infty(R)$ *by applying linearity of Lebesgue–Stieltjes integrals in the integrators*

$$\begin{aligned}\int f(X)\,\mathrm{d}(\gamma_N \circ R) &= -\int_{-\infty}^{\infty} x\,(\gamma_N \circ F_{-f(X)})(\mathrm{d}x)\\ &= -\int_{-\infty}^{\infty} x\,\left(\sum_{n=1}^N \tilde{\gamma}_n (f_n \circ F_{-f(X)})\right)(\mathrm{d}x)\\ &= -\sum_{n=1}^N \tilde{\gamma}_n \int_{-\infty}^{\infty} x\,(f_n \circ F_{-f(X)})(\mathrm{d}x).\end{aligned}$$

Scrutinizing the Stieltjes integrators $f_n \circ F_{-f(X)}$, $n \in \mathbb{N}$, *one observes that the functions* $f_n \circ F_{-f(X)}$ *are constant on* $[-\infty, q_{y_{n-1}}(-f(X))] \cup [q_{y_n}(-f(X)), \infty]$, $n \in \mathbb{N}$, *denoting by* $q_\alpha := q_\alpha(-f(X))$ *the lower* α*-quantile of* $-f(X)$, *which follows directly from the definition of* f_n. *Since the Stieltjes integral w.r.t. a constant integrator is* 0, *one can restrict the integration borders to* $[q_{y_{n-1}}, q_{y_n}]$. *Hence,*

$$\begin{aligned}&-\frac{1}{y_n - y_{n-1}} \int_{-\infty}^{\infty} x\,(f_n \circ F_{-f(X)})(\mathrm{d}x)\\ &= -\frac{1}{y_n - y_{n-1}} \int_{q_{y_{n-1}}}^{q_{y_n}} x\,F_{-f(X)}(\mathrm{d}x)\\ &= -\mathbb{E}\Big[-f(X)\Big|-f(X) \in \big[q_{y_{n-1}}(-f(X)), q_{y_n}(-f(X))\big]\Big]\\ &= \mathbb{E}\Big[f(X)\Big|f(X) \in \big[\mathrm{VaR}_{y_n}(f(X)), \mathrm{VaR}_{y_{n-1}}(f(X))\big]\Big]\end{aligned}$$

delivers the desired result, since $q_\alpha(-f(X)) = -\,\mathrm{VaR}_\alpha(f(X))$. □

In the theorem above, we have calculated an expression for the Choquet integral of $f(X)$ w.r.t. the distorted probability $\gamma_N \circ R$ which can be used to calculate ask prices. Analoguously, we can obtain a similarly convenient result for bid prices by means of the same techniques.

Corollary 7.3.6
Let $N \in \mathbb{N}$, γ_N *be a piecewise linear concave distortion function with adjacent decomposition* $0 = y_0 < y_1 < \cdots < y_N = 1$, *and difference vector* $(\Delta\gamma_1, \dots, \Delta\gamma_N) \in [0,1]^N$. *Let furthermore* $f(X) \in \mathcal{L}^\infty(R)$ *and* $y_0, \dots, y_N$ *are points of continuity of* $F_{f(X)}$. *Then the dual of the distorted expectation of* $f(X)$ *w.r.t.* $\gamma_N \circ R$ *can be calculated via*

$$-\left(\int -f(X)\,\mathrm{d}(\gamma_N \circ R)\right) = \sum_{n=1}^N \Delta\gamma_n \mathbb{E}\Big[f(X)\Big|f(X) \in \big[q_{y_{n-1}}(f(X)), q_{y_n}(f(X))\big]\Big].$$

Proof

Following the steps of the proof of Theorem 7.3.4, one sees that for $f(X) \in \mathcal{L}^\infty(R)$

$$\begin{aligned} -\int -f(X)\,\mathrm{d}(\gamma_N \circ R) &= \int_{-\infty}^{\infty} x\,(\gamma_N \circ F_{f(X)})(\mathrm{d}x) \\ &= \sum_{n=1}^{N} \tilde{\gamma}_n \int_{-\infty}^{\infty} x\,(f_n \circ F_{f(X)})(\mathrm{d}x) \end{aligned}$$

holds. Similar to the above proof, the Stieltjes integrator is constant on $[-\infty, q_{y_{n-1}}(f(X))] \cup [q_{y_n}(f(X)), \infty]$*,* $n \in \mathbb{N}$*, hence, it follows analoguously*

$$\begin{aligned} &\frac{1}{y_n - y_{n-1}} \int_{-\infty}^{\infty} x\,(f_n \circ F_{f(X)})(\mathrm{d}x) \\ &= -\mathbb{E}\Big[f(X)\Big|f(X) \in \big[q_{y_{n-1}}(f(X)), q_{y_n}(f(X))\big]\Big], \end{aligned}$$

which yields the assertion. □

In particular, the above formulae yield a tractable setting for solving the bid-ask calibration problem within the piecewise linear concave distortion functions, given a decomposition $\mathfrak{Y} = \{y_0, \ldots, y_N : 0 = y_0 < \cdots < y_N = 1\}$ of the unit interval. Hence, solving the bid-ask calibration problem corresponds to finding a vector $\Delta\gamma \in \mathbb{R}^N$ satisfying the constraints in (7.3) and minimizing some distance between model bid-ask prices and quoted market bid-ask prices. Thus, we can now construct a rough algorithm to implement bid-ask calibration to piecewise linear concave distortions, given a decomposition of the unit interval.

Algorithm 3 (Bid-ask calibration to piecewise linear concave distortions)
Let $(C_1^{\star\mathrm{bid}}, C_1^{\star\mathrm{ask}}), \ldots, (C_M^{\star\ \mathrm{bid}}, C_M^{\star\ \mathrm{ask}})$ *be given bid-ask market quotes of contingent claims* $C_1, \ldots, C_M$ *and* $\eta : \mathbb{R}_{\geq 0}^{2M} \to \mathbb{R}_{\geq 0}$ *an error function. The bid-ask calibration problem can be solved by the following algorithm:*

1. *Choose* $N \in \mathbb{N}$*.*
2. *Choose a decomposition* $\mathfrak{Y} = \{y_0, \ldots, y_N : 0 = y_0 < \cdots < y_N = 1\}$ *of the unit interval.*
3. *Calculate*

$$\mathbb{E}\Big[f(C_j)\Big|f(C_j) \in \big[q_{y_{n-1}}(f(C_j)), q_{y_n}(f(C_j))\big]\Big]$$

and

$$\mathbb{E}\Big[f(C_j)\Big|f(C_j) \in \big[\mathrm{VaR}_{y_n}(f(C_j)), \mathrm{VaR}_{y_{n-1}}(f(C_j))\big]\Big]$$

for $n = 1, \ldots, N$*,* $j = 1, \ldots, M$*.*

4. Solve the constrained optimization problem

$$\min_{\Delta\gamma\in\mathbb{R}^N_{\geq 0}} \eta\Bigg(\left|\sum_{n=1}^N \Delta\gamma_n \mathbb{E}[f(C_1)|f(C_1)\in[q_{y_{n-1}},q_{y_n}]] - C_1^{\star\,\mathrm{bid}}\right|,\dots,$$
$$\left|\sum_{n=1}^N \Delta\gamma_n \mathbb{E}[f(C_M)|f(C_M)\in[q_{y_{n-1}},q_{y_n}]] - C_M^{\star\,\mathrm{bid}}\right|,$$
$$\left|\sum_{n=1}^N \Delta\gamma_n \mathbb{E}[f(C_1)|f(C_1)\in[\mathrm{VaR}_{y_n},\mathrm{VaR}_{y_n-1}]] - C_1^{\star\mathrm{ask}}\right|,\dots,$$
$$\left|\sum_{n=1}^N \Delta\gamma_n \mathbb{E}[f(C_M)|f(C_M)\in[\mathrm{VaR}_{y_n},\mathrm{VaR}_{y_n-1}]] - C_M^{\star\,\mathrm{ask}}\right|\Bigg)$$

subject to

$$\sum_{n=1}^N \Delta\gamma_n = 1,\ \Delta\gamma\geq 0,$$
$$\left(\frac{\Delta\gamma_2}{y_2-y_1}-\frac{\Delta\gamma_1}{y_1-y_0},\dots,\frac{\Delta\gamma_N}{y_N-y_{N-1}}-\frac{\Delta\gamma_{N-1}}{y_{N-1}-y_{N-2}}\right) =: D(\Delta\gamma)\leq 0.$$

Thus, the core of the bid-ask calibration problem is reduced to a non-linear constrained optimization problem in the compact convex space

$$G := \left\{\Delta\gamma\in\mathbb{R}^N :\ \Delta\gamma\geq 0,\ D(\Delta\gamma)\leq 0,\ \sum_{n=1}^N \Delta\gamma_n = 1\right\}.$$

Algorithm 3 treats the optimization on a fixed decomposition $\mathfrak{Y}=\{y_0,\dots,y_N\}$ of the unit interval. Obviously, the methodology can be enhanced by varying over the decompositions as well, which also delivers a constrained optimization problem. The problem accompanying optimization over decompositions is performance: Varying decompositions considerably slow down the optimization procedure, since all conditional expectations

$$\mathbb{E}\Big[f(C_j)\Big|f(C_j)\in\big[q_{y_{n-1}}(f(C_j)),q_{y_n}(f(C_j))\big]\Big]$$

and

$$\mathbb{E}\Big[f(C_j)\Big|f(C_j)\in\big[\mathrm{VaR}_{y_n}(f(C_j)),\mathrm{VaR}_{y_{n-1}}(f(C_j))\big]\Big]$$

have to be recalculated in every optimization step, while they only have to be calculated once when fixing the decomposition $\mathfrak{Y}=\{y_0,\dots,y_N : 0=y_0<\dots<y_N\}$.

Remark 7.3.7 (Acceleration of Algorithm 3)
One possibility to accelerate and simplify the bid-ask calibration is equidistant spacing of $y_0 < \cdots < y_N$*:*[4] *Step 3 of Algorithm 3 then reduces to the calculation of*

$$\mathbb{E}\Big[f(C_j)\Big|f(C_j) \in \big[q_{y_{n-1}}(f(C_j)), q_{y_n}(f(C_j))\big]\Big]$$

for $n = 1, \ldots, N$*,* $j = 1, \ldots, M$ *and the concavity constraint* $D(\Delta\gamma) \leq 0$ *simplifies to*

$$\Delta^2\gamma := (\Delta\gamma_2 - \Delta\gamma_1, \ldots, \Delta\gamma_N - \Delta\gamma_{N-1}) \leq 0.$$

Furthermore, a closer look on the optimization problem in Algorithm 3 suggests the following strategies for the concrete numerical implementation. There are several possibilities to use additional knowledge for faster implementation. Very helpful tools arise from the theory of nonlinear optimization, where additional conditions on the objective function help for rapid implementation. A well-known tool in nonlinear constrained optimization is exploiting the knowledge of Lagrangian multipliers, which has been popularized as the *Karush–Kuhn–Tucker conditions* (i.e. (Boyd and Vandenberghe, 2004, p. 243ff)). We can shown that the optimization problem above exhibits the Karush–Kuhn–Tucker conditions under mild technical restrictions (i.e. exclusion of zero-slope parts, continuous differentiability of the goal function).

Proposition 7.3.8 (Karush–Kuhn–Tucker conditions for Algorithm 3)
Let $\eta : \mathbb{R}_{\geq 0}^{2M} \to \mathbb{R}_{\geq 0}$ *be an error function and denote the goal function for the optimization problem described in Algorithm 3 by* $\tilde{\eta} : \mathbb{R}_{>0}^{N} \to \mathbb{R}_{\geq 0}$*, i.e.*

$$\begin{aligned}\tilde{\eta}(\Delta\gamma) = \eta\Bigg(&\left|\sum_{n=1}^{N} \Delta\gamma_n \mathbb{E}[f(C_1)|f(C_1) \in [q_{y_{n-1}}, q_{y_n}]] - C_1^{\star\,\mathrm{bid}}\right|, \ldots, \\ &\left|\sum_{n=1}^{N} \Delta\gamma_n \mathbb{E}[f(C_M)|f(C_M) \in [q_{y_{n-1}}, q_{y_n}]] - C_M^{\star\,\mathrm{bid}}\right|, \\ &\left|\sum_{n=1}^{N} \Delta\gamma_n \mathbb{E}[f(C_1)|f(C_1) \in [\mathrm{VaR}_{y_n}, \mathrm{VaR}_{y_n - 1}]] - C_1^{\star\mathrm{ask}}\right|, \ldots, \\ &\left|\sum_{n=1}^{N} \Delta\gamma_n \mathbb{E}[f(C_M)|f(C_M) \in [\mathrm{VaR}_{y_n}, \mathrm{VaR}_{y_n - 1}]] - C_M^{\star\,\mathrm{ask}}\right|\Bigg)\end{aligned}$$

for $\Delta\gamma \in \mathbb{R}_{>0}^{N}$[5]*. If the goal function* $\tilde{\eta}$ *is continuously differentiable, then the Karush–Kuhn–Tucker conditions for nonlinear optimization are fulfilled, i.e. for a minimizing*

[4] Actually, this can be relaxed further: If $y_0 < \cdots < y_N$ are symmetrically spaced around 0.5, the same acceleration argument applies.

[5] To guarantee for an open domain, we restrict ourselves only to the distortion functions that have non-zero slopes in every point.

piecewise linear distortion function $\gamma^\star$, there exist constants $\lambda_1, \ldots, \lambda_{N-1} \in \mathbb{R}_{\geq 0}$, $\nu \in \mathbb{R}$, such that

$$\nabla\tilde{\eta}(\Delta\gamma^\star) + \sum_{j=1}^{N-1} \lambda_j \Xi_j + (\nu, \ldots, \nu) = 0$$

holds, denoting by $\nabla\tilde{\eta}$ the gradient of the goal function $\tilde{\eta}$ w.r.t. $\Delta\gamma$ and

$$\Xi_j := -\frac{e_j}{y_j - y_{j-1}} + \frac{e_{j+1}}{y_{j+1} - y_j},$$

where $(e_j)_{j=1}^N$ denote the the standard unit vectors in $\mathbb{R}^N$ with $e_j = (\delta_{1j}, \ldots, \delta_{Nj})$ and δ_{kl} is the Kronecker symbol.

Proof

We use the Karush–Kuhn–Tucker conditions as described in (Boyd and Vandenberghe, 2004, p. 243) and obtain that

$$\nabla\tilde{\eta}(\Delta\gamma^\star) + \sum_{j=1}^{L} \lambda_j \nabla f_j(\Delta\gamma^\star) + \nu\nabla h(\Delta\gamma^\star) = 0$$

holds for some $\lambda_1, \ldots, \lambda_L \geq 0$, $\nu \in \mathbb{R}$, and optimal $\Delta\gamma^\star \in \mathbb{R}^N_{>0}$, denoting by f_j all inequality constraints with $f_j(\Delta\gamma) \leq 0$ and by h the equality constraint

$$h(\Delta\gamma) = -1 + \sum_{j=1}^{N} \Delta\gamma_j = 0.$$

If we scrutinize the concrete specification of f_j from Algorithm 3, we obtain $L = N - 1$ and

$$f_j(\Delta\gamma) = \frac{\Delta\gamma_{j+1}}{y_{j+1} - y_j} - \frac{\Delta\gamma_j}{y_j - y_{j-1}}.$$

Calculating the partial derivatives of f_j and h w.r.t. $\Delta\gamma_k$ exactly yields the gradients Ξ_j as described in the assertion. □

Often, some Euclidean distance function is used as an error function (as we do in the numerical case study described in Section 6.3). In this case, we can even use more information than in the Karush–Kuhn–Tucker case described in Proposition 7.3.8 and apply further specific algorithms as a strategy for finding a solution.

Remark 7.3.9 (Linear constrained least-squares problem)

If we use the Euclidean distance (or some strictly monotone transformation of it) as an error function (e.g. the popular RMSE error function), the bid-ask calibration to piecewise linear concave distortions reduces to a linearly constrained least-squares optimization problem which is well treated in literature (see, e.g., (Hanson and Haskell, 1982; Hanson, 1986)).

7.3.2 Application to parameter risk-captured prices

We now want to apply the nonlinear calibration procedure described in Section 7.3.1 to the parameter risk-captured pricing framework presented in Section 3.2. The central assumption in this framework is the presence of a distribution R on the parameter space Θ. In some cases of historical estimation of parameters (e.g. correlation estimation as in Example 5), the distribution R on Θ is given by the distribution of the parameter's estimator. In other cases, e.g. the calibration to market prices, the distribution R may be recovered from the calibration to mid prices by an algorithm as described in Chapter 6. While in the examples in the Chapters 5 and 6, the choice of risk measure was regarded to be subjective, Algorithm 3 allows us to calibrate parameter risk-captured prices to bid-ask prices using a broad and flexible class of risk measures, represented by piecewise linear concave distortion functions.

Using the suggested procedure for obtaining a distribution R on the parameters, when calibrating to market prices of vanillas, the result is a three-step calibration scheme:

First, we calibrate to mid market prices and obtain a parameter $\theta_0 \in \Theta$. Second, using the parameter θ_0, we construct the distribution R on the parameter space Θ as suggested in Chapter 6. Finally, we calibrate to bid-ask prices using Algorithm 3 to obtain the best matching concave distortion function γ_0.

The choice of the error function is also somewhat delicate: (Detlefsen and Härdle, 2007) observe that different choices of error functions deliver different calibration results. Furthermore, (Guillaume and Schoutens, 2011) scrutinize different calibration methods (historical calibration, calibration to vanillas, etc.), restricted to the Heston model. We omit further interdependencies to those issues in the current investigation and concentrate on one error function and one calibration method.

7.4 Application to data

In the previous section, we have presented and discussed a non-parametric calibration scheme based on piecewise linear concave distortion functions to bid-ask prices. In this section, we apply our piecewise linear calibration scheme and compare it to a parametric calibration scheme à la (Cherny and Madan, 2010), using the introduced parametric families of AVaR- and *minmaxvar*-type distortion functions. Therefore, we calibrate a Γ-OU-Barndorff-Nielsen–Shephard model to the quoted mid prices of plain vanilla options.

Afterwards, we induce a distribution on the parameter space employing the normal transformation function as in Chapter 6.3. Then, we calibrate the parameter risk-captured prices to quoted bid-ask prices of vanillas, comparing parametric ansatzes following (Cherny and Madan, 2010) and the non-parametric calibration approach from Algorithm 3. As a result, we obtain a very characteristic shape of the "market-implied" piecewise linear approximation. From our empirical results, we consequently suggest using a different parametric family of concave distortion functions: The ess sup-expectation convex combinations, simply interpolating between the essential supremum and the vanilla expectation w.r.t. the parameter distribution R. This family turns out to match the results from the piecewise linear calibration approach and allows for a fast and efficient calibration.

Our set of data is a DAX option surface, consisting of 501 bid and ask vanilla prices as of December 2, 2011, with different maturities and strikes. For simplicity, we assume no bid-ask spreads in the DAX spot price and EUR interest rates.[6]

We use a Γ-OU-Barndorff-Nielsen–Shephard model as in Section 6.3, for details see (Barndorff-Nielsen and Shephard, 2001; Cont and Tankov, 2004). The risk-neutral dynamics of the log-index price $(X_t)_{t\geq 0}$ in the Barndorff-Nielsen–Shephard model are given by

$$\begin{aligned} \mathrm{d}X_t &= \left(r - \frac{\sigma_t^2}{2} + \frac{\lambda c \rho}{\alpha + \rho}\right) \mathrm{d}t + \sigma_t \,\mathrm{d}W_t - \rho \,\mathrm{d}Z_{\lambda t}, \\ \mathrm{d}\sigma_t^2 &= -\lambda \sigma_t^2 \,\mathrm{d}t + \mathrm{d}Z_{\lambda t}, \end{aligned}$$

with parameters $r, S_0, \sigma_0^2, \lambda, c, \rho, \alpha > 0$, $(W_t)_{t\geq 0}$ is a Brownian motion, and $(Z_t)_{t\geq 0}$ is a compound Poisson process with exponentially distributed jump size, i.e. $Z_t = \sum_{j=1}^{N_t} U_j$ with a Poisson process $(N_t)_{t\geq 0}$ with intensity $c > 0$ and $(U_j)_{j\in\mathbb{N}}$ are exponentially i.i.d. with parameter $\alpha > 0$. $(W_t)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ are independent. Since we assume the DAX spot price S_0 and risk-free rate r to be given, the unspecified parameters with exposure to parameter risk are gathered in the quintuple $(\sigma_0^2, c, \alpha, \lambda, \rho)$.

For calculating vanilla prices, we use the Fourier pricing method of (Carr and Madan, 1999; Raible, 2000) and calculate call prices for the moneyness dimension simultaneously via FFT. As an error function for both the initial calibration to mid prices and the calibration to bid-ask prices, we use the RMSE error function without any weighting (for the impact of using different error functions, we refer to (Detlefsen and Härdle,

[6] Since the DAX is a performance index, dividends are included and do not have to be modeled seperately.

2007)), standardized by the mean option price, similar to the setting in Section 6.3. Thus, our error function is

$$\eta(\theta) = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}(\mathbb{E}_\theta[C_m] - C_M^{\star\,\mathrm{mid}})^2}}{\frac{1}{M}\sum_{m=1}^{M} C_M^{\star\,\mathrm{mid}}}.$$

Parameter risk setup and calibration

We apply the methods presented in Chapter 6 and obtain a parameter risk distribution R on Θ by discretizing Θ and weighting the parameters with their error to market prices, transformed and normed by a decreasing function h such that the sum of all parameters equals one. We hereby incorporate all parameters (on a discrete grid) up to an aggregate market error of 3.5% and weight them by transforming the error function with the normal transformation function

$$h_\lambda^{\mathcal{N}}(t) := c \cdot \exp\left(-\frac{(t-t_0)^2}{2\lambda^2}\right),$$

where $\lambda = 0.005$, $c > 0$ matching, and $t_0 = 1.63\%$ denoting the minimal aggregate market error. Doing so, we obtain a discrete distribution on Θ with a support of 9 430 parameter vectors with an aggregate market error of less than 3.5%.

With a distribution R on Θ at hand, we calibrate to bid and ask prices in various ways: First, we calibrate to bid-ask prices with our non-parametric piecewise linear approximation scheme described in Algorithm 3. As a unit interval decomposition, we use 100 and 1 000 equidistant nodes; for optimization purposes, we use again the mean-standardized RMSE as in the mid prices calibration. Second, we compare our result to a parametric calibration using parametric families of distortion functions. We hereby employ the popular AVaR- and *minmaxvar*-families for calibration and observe differences in calibration performance.

Results

As a first result, we obtain that the calibration performances of the piecewise linear and the parametric approaches do not differ significantly, the standardized RMSEs of all approaches are close to each other, cf. Table 7.1.

Actually, the distortion functions that result from the calibration differ in shape (cf. Figure 7.2), in case of the AVaR-calibration due to its specific shape. The piecewise

Distortion framework	RMSE/mean to bid-ask prices	CPU time
piecewise linear 1 000 nodes	1.64%	301.11 sec
piecewise linear 100 nodes	1.64%	4.26 sec
minmaxvar-family	1.65%	3.17 sec
AVaR-family	1.64%	3.73 sec

Table 7.1 A comparison of the calibration performance of different distortion function families. Actually, the influences of the choice of the distortion function family are minor, i.e. the calibration performance is comparable. Employing much more nodes in the piecewise linear approach does not improve the calibration performance, but results in more computational time.

linear calibration result is clearly the most flexible method and does not exhibit too strong performance drawbacks compared to parametric calibration when choosing a reasonable number of nodes.

As a very remarkable result, one observes that both the 100 and 1 000 nodes approximation follow the same pattern: After a sharp increase close to zero, we can observe linear growth in the argument of the distortion function. This observation is quite robust: When using different transformation functions for creating the distribution on Θ and incorporating more parameters, this pattern remains the same. Even when using a different model (e.g. the Heston model), a similarly shaped distortion function is obtained. Furthermore, the result is stable w.r.t. different choices of starting vectors in the optimization procedure, so it is also unlikely that it results from numerical instabilities.

7.5 Calibration to discontinuous distortion functions

The results from the calibration in the previous section exhibit a characteristic pattern in the shape of the distortion function for the nonparametric piecewise linear approximation. Although the piecewise linear concave distortions are also continuous, the shape of the calibration result suggest that a discontinuity in zero should be considered as well.

Hence, our results motivate to introduce another parametric family of concave distortion functions, see Definition 7.5.1.

Definition 7.5.1 (*ess sup*-expectation convex combinations)
Let $\lambda \in [0,1]$. The distortion function

$$\gamma_\lambda(u) := \begin{cases} 0, & u = 0 \\ \lambda + (1-\lambda)u, & u \in (0,1] \end{cases}$$

is called the ess sup-expectation convex combination distortion *with weight λ on the essential supremum.*

The name of this family stems from the behavior of the Choquet integral w.r.t. a γ_λ-distorted probability: If we have a probability measure P and a bounded random variable X, the Choquet integral w.r.t. $\gamma_\lambda \circ P$ is a convex combination of the essential supremum of X and the ordinary expectation w.r.t. P, weighting the essential supremum with λ and the expectation with $1-\lambda$.

As one can easily see, the suggested class of distortion function from Definition 7.5.1 is not continuous any more, but exhibits a jump at zero. As already pointed out in Section 7.2, zero is the only point where the a concave distortion function may exhibit jumps. Hence, we can easily show that we can decompose *every* distortion function into two parts.

Proposition 7.5.2 (Decomposition of discontinuous distortion functions)
Let $(\Omega, \mathcal{F}, P)$ be a probability space, γ a concave distortion function with a jump at zero, X an integrable random variable that is P-a.s. bounded from above, and let $\lambda = \lim_{v \downarrow 0} \gamma(v)$ be the height of the jump. Then the function

$$\gamma^{\text{cont}}(u) = \frac{\lim_{v \downarrow u} \gamma(v) - \lambda}{1-\lambda}$$

is a continuous concave distortion function and

$$\int_\Omega X \,\mathrm{d}(\gamma \circ P) = \lambda \operatorname{ess\,sup} X + (1-\lambda) \int_\Omega X \,\mathrm{d}(\gamma^{\text{cont}} \circ P).$$

In particular, for every concave distortion function with a jump at zero the Choquet integral w.r.t. $\gamma \circ P$ is representable as a convex combination of the ess sup *and a continuous concave distortion function γ^{cont}.*

We call γ^{cont} the *continuous part* of γ and λ the *jump part* of γ. It may be noted that the continuous part γ^{cont} is not unique, but we denote with it the construction in Proposition 7.5.2. For $\lambda = 1$, we trivially set $\gamma^{\text{cont}} = 0$.

Proof (of Proposition 7.5.2)
It is easy to show that γ^{cont} *is a continuous and concave distortion function (if* $\lambda = 1$*, the proposition is trivially true). Abbreviate* $x_0 := \operatorname{ess\,sup} X$ *(first assuming* $x_0 \geq 0$*), then the definition of the Choquet integral immediately yields*

$$\begin{aligned}
\int_\Omega X \,\mathrm{d}(\gamma \circ P) &= \int_{-\infty}^{0} \gamma(P(X > x)) - 1 \,\mathrm{d}x + \int_0^\infty \gamma(P(X > x)) \,\mathrm{d}x \\
&= \int_{-\infty}^{0} \gamma(P(X > x)) - 1 \,\mathrm{d}x + \int_0^{x_0} \gamma(P(X > x)) \,\mathrm{d}x \\
&= \int_{-\infty}^{0} (1-\lambda)\gamma^{\text{cont}}(P(X > x)) + \lambda - 1 \,\mathrm{d}x + \\
&\quad \int_0^{x_0} (1-\lambda)\gamma^{\text{cont}}(P(X > x)) + \lambda \,\mathrm{d}x \\
&= \lambda x_0 + (1-\lambda) \int_\Omega X \,\mathrm{d}(\gamma^{\text{cont}} \circ P).
\end{aligned}$$

Similar for $x_0 < 0$. □

In particular, if we use γ_λ from Definition 7.5.1 as distortion function, we are interpolating between the essential supremum and the expectation (since the continuous part of γ_λ is just the identity function on $[0,1]$) when calculating the Choquet integral. With this, we can easily generalize the counterexample which was presented in Example 4.2.1 that distortion-induced risk-captured prices do not necessarily exhibit the convergence property (CP) presented in Definition 4.1.1.

Example 7.5.3 (Generalization of Example 4.2.1)
As, in Example 4.2.1, we consider the parameter space $\Theta = \mathbb{R}$ *and the sequence of distributions* $R_N \sim \mathcal{N}(0, 1/N)$*, which converges to a Dirac distribution with probability mass concentrated at* 0*. Let* $f \in \mathcal{C}^b(\mathbb{R})$*. Since we have seen that* $\max_{x\in\mathbb{R}} f(x) = \lim_{N\to\infty} \operatorname{ess\,sup}_{R_N} f \neq \operatorname{ess\,sup}_R = f(0)$ *in general, we can immediately conclude that for every discontinuous distortion function, applying the decomposition from Proposition 7.5.2, the respective distortion risk measures (i.e. the Choquet integrals) do not converge, since the* ess sup*-part does not converge.*

Since our piecewise linear calibration results in Figure 7.1 look similar to an ess sup-expectation convex combination, they deliver an appealing interpretation for trading: When setting the bid-ask prices by means of the calibration risk framework of Chapter 6, we get bid-ask spreads that are fairly in line with the market when calculating the worst case of all parameters, the expectation w.r.t. the delivered distribution, and simply interpolating between them, using the weight of the worst case as a "risk-aversion

Distortion framework	RMSE/mean to bid-ask prices	CPU time
piecewise linear 100 nodes	1.64%	4.26 sec
minmaxvar-family	1.65%	3.17 sec
AVaR-family	1.64%	3.73 sec
ess sup-exp-family	1.64%	0.21 sec

Table 7.2 A comparison of the calibration performance of different distortion function families, now including the ess sup-expectation convex combinations. One can see that the simple structure of the ess sup-expectation convex combinations results in higher computational efficiency compared to other one-parametric distortion function families.

parameter". Strikingly, this simple approach ties with the more sophisticated AVaR- and *minmaxvar*-methodologies in calibration performance and is much easier and faster to implement: In our setting, we just have to calculate the supremum, the infimum, and the expectation, which can efficiently be done using vectorized programming systems (e.g. MATLAB). Afterwards, the optimization procedure only incorporates three values per vanilla option and is much faster compared to other parametric distortion function types or the 100 nodes piecewise linear function:

Furthermore, we obtain a firm economic interpretation of the expectation w.r.t. the distribution: The expectation can be interpreted as the "true mid price" (which is not observable in the market, since we only get bid and ask quotes), with a (non-symmetric) risk premium relative to the parameter risk which is expressed by the essential supremum (for ask prices) and infimum (for bid prices). This is in line with (Cont, 2006) and (Lindström, 2010), who argue seperately for the supremum and the expectation incorporating model/parameter risk.

Since our empirical observations motivate the introduction of discontinuous distortion functions, we finally generalize the existence theorem for a solution to the bid-ask calibration problem which was formulated for continuous distortion functions in Theorem 7.2.2 to discontinuous distortion functions and allow for an additional jump of arbitrary size at 0.

Theorem 7.5.4

Let $f(X)$ be bounded, η a continuous error function, $K > 0$, and define

$$G_K^{\text{jump}} := \{\gamma : [0,1] \to [0,1] : \gamma \text{ is a distortion function with } \gamma^{\text{cont}} \text{ is concave, monotone, Lipschitz with Lipschitz constant } K\}.$$

Then there is a solution of the bid-ask calibration problem in G_K^{jump}, i.e. there exists a minimizing distortion function $\tilde{\gamma} \in G_K^{\mathrm{jump}}$ such that the function

$$\tilde{\eta} : \gamma \mapsto \eta\big(|-\Gamma_\gamma(-C_1) - C_1^{\star\,\mathrm{bid}}|, \dots, |-\Gamma_\gamma(-C_M) - C_M^{\star\ \mathrm{bid}}|, \\ |\Gamma_\gamma(C_1) - C_1^{\star\mathrm{ask}}|, \dots, |\Gamma_\gamma(C_M) - C_M^{\star\ \mathrm{ask}}|\big)$$

is minimized in $\tilde{\gamma}$ in the following sense:

$$\tilde{\eta}(\tilde{\gamma}) \leq \tilde{\eta}(\gamma) \qquad \text{for all } \gamma \in G_K^{\mathrm{jump}}.$$

Proof
Let $\gamma \in G_K^{\mathrm{jump}}$ and decompose $\gamma = (\lambda, \tilde{\gamma})$ via $\lambda := \gamma(0+) - \gamma(0)$ and $\tilde{\gamma}(y) := \gamma(y) - \gamma(0+)$, which is very close to the decomposition which was introduced in Proposition 7.5.2. As a first step, we equip G_K^{jump} with a special topology: Since the decomoposition into a jump part λ and the remaining continuous part $\tilde{\gamma}$ is unique and can be recovered by setting

$$\gamma(y) := \begin{cases} 0, & y = 0 \\ \lambda + \tilde{\gamma}(y), & y > 0, \end{cases}$$

we identify G_K^{jump} with the set

$$S := \{(\lambda, \gamma) \in [0,1] \times \mathcal{C}[0,1] : \gamma(0) = 0, \gamma(1) = 1 - \lambda, \gamma \text{ is monotone,} \\ \text{concave with Lipschitz constant } K\}.$$

Thus, the product topology induced by the usual Euclidean topology on $[0,1]$ and the uniform topology on $\mathcal{C}[0,1]$ is the natural topology to equip S with. From now on, we follow the steps from the proof of Theorem 7.2.2.

Step 1: G_K^{jump} (resp. S) is compact w.r.t. the product topology described above.

First, we define the set of all Lipschitz continuous "subdistortion functions"

$$\tilde{G}_K := \{\gamma \in \mathcal{C}[0,1] : \gamma(0) = 0, \gamma(1) \leq 1, \gamma \text{ is monotone,} \\ \text{concave with Lipschitz constant } K\}.$$

Trivially, we obtain that $S \subset [0,1] \times \tilde{G}_K$ holds. By construction, $\tilde{G}_K$ is uniformly bounded (by 1) and equicontinuous (since all functions are Lipschitz with Lipschitz constant K). As a second step, we note that $\tilde{G}_K$ is closed: For a $\|\cdot\|_\infty$-convergent sequence $(\gamma_n)_{n\in\mathbb{N}}$ from $\tilde{G}_K$ with limit $\gamma(y) := \lim_{n\to\infty} \gamma_n(y)$ for $y \in [0,1]$, we immediately obtain that γ is monotone, concave $\gamma(0) = 0$ and $\gamma(1) \leq 1$. Furthermore, γ is Lipschitz with Lipschitz

constant $K > 0$. Let $\varepsilon > 0$ be arbitrary and choose $N \in \mathbb{N}$ such that $\|\gamma - \gamma_N\|_\infty < \varepsilon/2$. Hence, for arbitrary $x, y \in [0,1]$, we obtain

$$\begin{aligned}|\gamma(x) - \gamma(y)| &\leq |\gamma(x) - \gamma_N(x)| + |\gamma_N(x) - \gamma_N(y)| + |\gamma_N(y) - \gamma(y)| \\ &\leq 2\|\gamma - \gamma_N\|_\infty + |\gamma_N(x) - \gamma_N(y)| \\ &< K|x - y| + \varepsilon.\end{aligned}$$

Since ε was arbitrary, γ is Lipschitz continuous with Lipschitz constant $K > 0$. Thus, $\gamma \in \tilde{G}_K$, which proves that $\tilde{G}_K$ is closed. Hence, the Arzelà–Ascoli theorem yields that $\tilde{G}_K$ is compact in the uniform topology. Thus, Tychonoff's theorem immediately delivers that $[0,1] \times \tilde{G}_K$ is compact in the product topology. Now, it suffices to show that the set S is closed: Therefore we again take a convergent sequence $(\lambda_n, \tilde{\gamma}_n)_{n\in\mathbb{N}}$ with $(\lambda_n, \tilde{\gamma}_n) \in S$ for every $n \in \mathbb{N}$. The limit pair is then defined by $\lambda := \lim_{n\to\infty} \lambda_n$ and $\tilde{\gamma}(y) := \lim_{n\to\infty} \tilde{\gamma}_n(y)$ for $y \in [0,1]$. If we now look at $\tilde{\gamma}_n(1) = 1 - \lambda_n$, $\tilde{\gamma}(y) = \lim_{n\to\infty} \tilde{\gamma}_n(y) = 1 - \lambda$ trivially. But this is exactly the condition that has to hold for $(\lambda, \tilde{\gamma}) \in S$. Hence, S is closed and therefore compact.

Step 2: The Choquet integral as a function of the distortion function, i.e. the map $I_X : G_K^{\text{jump}} \to \mathbb{R}$ defined by

$$I_X(\gamma) := \int f(X)\,\mathrm{d}(\gamma \circ R),$$

is continuous w.r.t. the product topology.

Let $(\lambda_n, \tilde{\gamma}_n)_{n\in\mathbb{N}}$ be a convergent sequence in S with $(\lambda, \tilde{\gamma}(y)) := \lim_{n\to\infty}(\lambda_n, \tilde{\gamma}_n(y))$ for $y \in [0,1]$. According to Proposition 7.5.2, we know that the Choquet integral of $f(X)$ w.r.t. the induced distortion functions γ_n can be represented as

$$\begin{aligned}\int f(X)\,\mathrm{d}(\gamma_n \circ R) &= \lambda_n \operatorname{ess\,sup} f(X) + \int f(X)\,\mathrm{d}(\tilde{\gamma}_n \circ R) \\ &= \lambda_n \operatorname{ess\,sup} f(X) + \int_{-\infty}^{0} \tilde{\gamma}_n(R(f(X) > x)) - \tilde{\gamma}_n(1)\,\mathrm{d}x \\ &\quad + \int_0^\infty \tilde{\gamma}_n(R(f(X) > x))\,\mathrm{d}x\end{aligned}$$

Dominated convergence, which is applicable due to the assumed boundedness of $f(X)$, yields

$$\begin{aligned}
&\to \lambda \operatorname{ess\,sup} f(X) + \int_{-\infty}^{0} \tilde{\gamma}(R(f(X) > x)) - \tilde{\gamma}(1)\,\mathrm{d}x \\
&\qquad + \int_{0}^{\infty} \tilde{\gamma}(R(f(X) > x))\,\mathrm{d}x \\
&= \int f(X)\,\mathrm{d}(\gamma \circ R)
\end{aligned}$$

for $n \to \infty$, which shows that the Choquet integral (as a function of the distortion function) is again continuous w.r.t. the product topology.

Since η is a continuous error function (thus $\eta \circ I_X$ is continuous w.r.t. the product topology) and G_K^{jump} is compact, there exists an element $\tilde{\gamma} \in G_K^{\mathrm{jump}}$ which minimizes the function $\eta \circ (I_{C_1}, \ldots, I_{C_M})$ for given contigent claims $C_1, \ldots, C_M$. Hence, $\tilde{\gamma}$ minimizes the error measured by η to given market bid-ask prices $(C_1^{\star\mathrm{bid}}, C_1^{\star\mathrm{ask}}), \ldots, (C_M^{\star\ \mathrm{bid}}, C_M^{\star\ \mathrm{ask}})$, i.e. $\eta(I(\tilde{\gamma})) \leq \eta(I(\gamma))$ holds for all $\gamma \in G_K^{\mathrm{jump}}$. □

Theorem 7.5.4 shows that even for a broad class of discontinuous functions (i.e. those with a jointly Lipschitz continuous part), the bid-ask calibration problem has always a solution. Similar to the proof of Theorem 7.2.2, the proof heavily relies on the compactness of the subset. Hence, we can again easily generalize the result to compact subclasses of G_K^{jump}.

Corollary 7.5.5 (Existence of bid-ask calibration in subclasses)
Let $K > 0$, $G \subset G_K^{\mathrm{jump}}$ be a subset which is closed in the product topology (G_K^{jump} and the product topology as described in Theorem 7.5.4) and η be a continuous error function. Then, the bid-ask calibration problem has a solution in G.

Proof
Since G is closed and G_K^{jump} is compact in the product topology, G is also compact. Hence, following the arguments of the proof of Theorem 7.5.4, we obtain that the bid-ask calibration problem has a solution in G. □

Corollary 7.5.5 can, e.g., be applied to justify the usage of the parametric distortion function class of the ess sup-expectation convex combinations described in Definition 7.5.1. We can now show that there is always a solution of the bid-ask calibration problem in the set of ess sup-expectation convex combinations.

Example 7.5.6
Considering the ess sup*-expectation convex combinations defined in Definition 7.5.1, we can see that they can be identified with the set*

$$C := \{(\lambda, \tilde{\gamma}) \in [0,1] \times \mathcal{C}[0,1] : \tilde{\gamma}(y) = (1-\lambda)y\}.$$

One can easily calculate that C is closed w.r.t. the product topology and, obviously, $C \subset G_K^{\text{jump}}$. Hence, according to Corollary 7.5.5, the bid-ask calibration problem has a solution in the ess sup*-expectation convex combinations.*

7.6 Outlook

In this chapter, we have treated the calibration of market-quoted bid-ask prices to broad classes of distortion risk measures, i.e. distortion risk measures that are induced by Lipschitz continuous distortion functions, which is later extended to distortion functions where the continuous part (cf. Section 7.5) is Lipschitz continuous. Furthermore, we have proposed a non-parametric calibration scheme based on piecewise linear approximation. Still, there remain some open questions which could be adressed in further research.

- First, in this work, we focus on distortion risk measures, which is a tractable class of convex risk measures. One could argue that due to central limit arguments, one can show that distortion risk measures are somehow the "center of attraction" for all law-invariant coherent risk measures, which was shown in (Belomestny and Krätschmer, 2012). Nevertheless, since the definition of parameter risk-captured prices employs general convex, normalized, law-invariant risk measures, one might want to find solutions in these broader classes. Hence, existence of a solution to the bid-ask calibration problem as well as the numerical treatment to find calibration results in these broader classes is a major obstacle for the future.

- Second, in our work, we always require that the continuous part of the distortion function is Lipschitz continuous (even jointly Lipschitz continuous). We think that this is not a major problem from a practical point of view (i.e. every distortion function is Lipschitz continuous with Lipschitz constant $1/\varepsilon$ on $[\varepsilon, 1]$ for $\varepsilon \in (0,1)$), but from a mathematical point of view, it would be more satisfactory to find solutions for the bid-ask calibration problem in the class of *all* distortion functions.

- Third, one may be interested in empirical studies whether the observed pattern in the calibration exercise is stable or the market-implied distortion function changes

over time. As the calibration result suggested, the market-implied prices are very similar for very different kind of distortion functions (the parametric distortion functions calibrated almost as good as the non-parametric piecewise linear variant, cf. the results from Section 7.4).

Thus, we think that our contribution in this thesis provides good results that can be applied in practice, but may be extended further to incorporate broader classes of convex risk measures and to investigate how market-implied distortion functions may change over time.

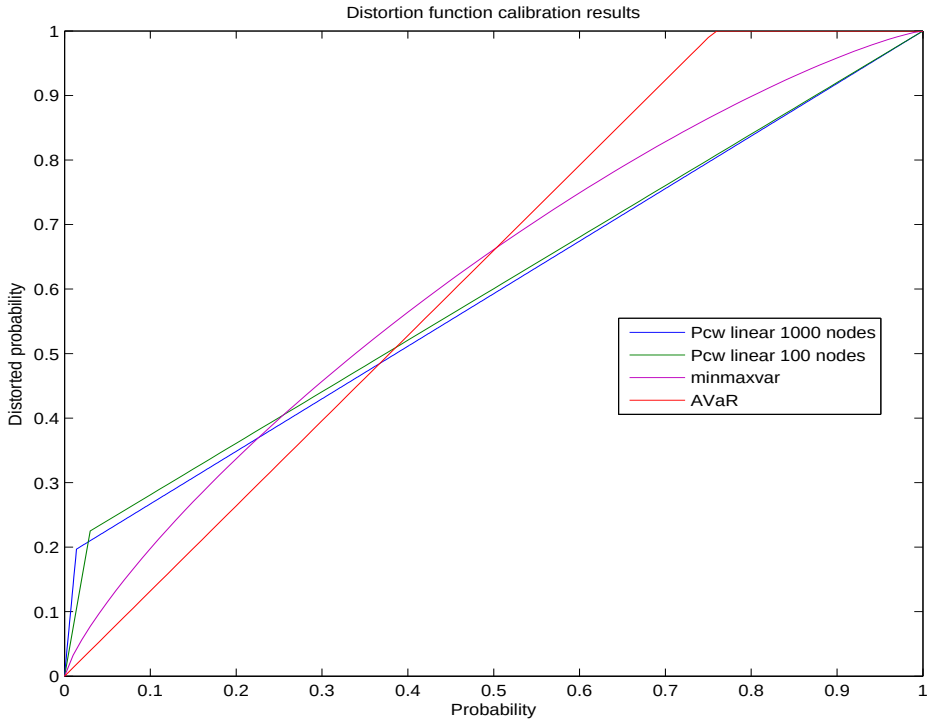


Figure 7.1 Different calibration results for a (non-parametric) piecewise linear calibration with 1 000 and 100 nodes, a parametric calibration to the AVaR-distortion function family, and a parametric calibration to the *minmaxvar*-family. While the piecewise linear calibration results with different nodes are very similar and have a characteristic jump close to zero and afterwards a linear behavior, we get very different results for the *minmaxvar*- and AVaR-results. In the upper probability regions, the *minmaxvar*-result allocates more probability than the piecewise linear result, while the AVaR allocates even more probability. In the lower probability region, we get more similarity for *minmaxvar*- and AVaR-results, but a very different pattern to the jump in the piecewise linear calibration results. Since the calibration results seem to be pretty similar, the higher probability allocation to high-probability events like in the AVaR- and *minmaxvar*-case does not seem to impact the calibration result too much.

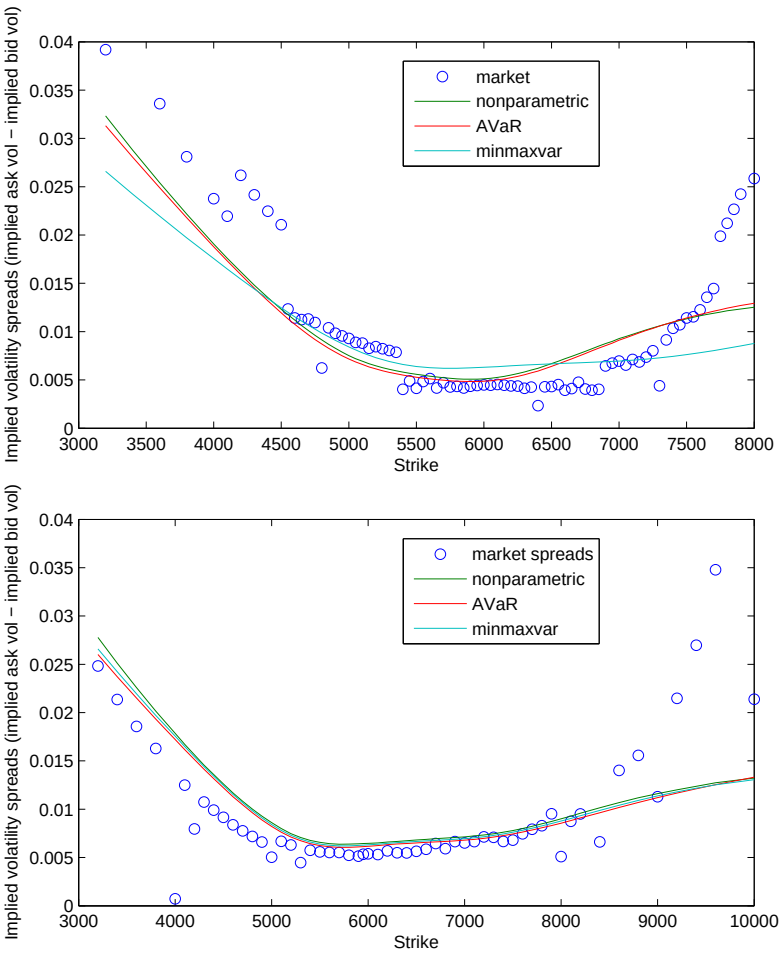


Figure 7.2 Bid-ask spreads for differently calibrated distortions vs. real-world bid-ask spreads, expressed in implicit volatilities for a short-term (above) and a long-term (below) maturity. Since the BNS model with its downward jumps emphasizes the put side of the smile, we exhibit more parameter risk on the left wing and those larger bid-ask spreads can be captured well by all models. For the same reason, the right wing vol spread is matched less accurate, in particular for short-term maturities, but AVaR- and piecewise linear distortions match it more efficiently than the calibrated *minmaxvar*-distortion. One has to encounter that parameter risk may not be the main driver for far-OTM call prices, other effects like illiquidity seem to be more predominant.

8 Conclusion

8.1 Summary

In this dissertation, the phenomenon of model and parameter risk (with an emphasis on parameter risk) in the context of risk-neutral pricing of derivatives is analyzed. Since parameter risk is an omnipresent problem in derivatives pricing, it should be reflected in the prices which derivatives traders are stating. In particular, it should influence the width of bid-ask spreads.

In this thesis, a possible methodology to incorporate parameter risk into risk-neutral derivatives prices is suggested and a framework and theory of model (resp. parameter) risk-capturing functionals is developed based on the extended notion of convex risk measures which was suggested in (Frittelli and Scandolo, 2006; Krätschmer, 2006), where translation invariance is given w.r.t. a linear form on a subspace. Other works on parameter risk (resp. uncertainty) as (Cont, 2006) and (Lindström, 2010) are generalized by the presented framework, which resembles of ideas of (Branger and Schlag, 2004; Gupta, 2009; Gupta et al., 2010). While analyzing continuity properties of risk-capturing functionals, we exhibit continuity properties of convex risk measures as a function of the underlying probability measure. Different from the ansatz of (Weber, 2006) and (Krätschmer et al., 2012), where different topologies are imposed on the domains of the convex risk measures, continuity w.r.t. the weak topology on the set of all probability measures is analyzed w.r.t. different risk-capturing functionals. It can be shown that a broad class of convex risk measures, the spectral risk measures, fulfill weak continuity and an extended version of the Portmanteau theorem is provided as a corollary.

As an application, it is shown that the estimation risk premium that is caused by consistent estimators eventually vanishes, if generated by a good-natured class of convex risk measures. Moreover, in case of asymptotically normal estimators, large-sample approximations can be stated, for selected risk measures in (semi-)closed form, provided that the sensitivities (i.e. the derivatives of the contingent claim's price) w.r.t. the parameters

are known. Parameter risk resulting from historical estimation is treated in numerical examples, discussing parameter risk arising from estimation of the correlation in a two-dimensional Black–Scholes model when pricing an exchange option, and the multivariate estimation of various parameters in a complex multifactor model for gas, electricity, and emissions prices.

For practical use, a method to incorporate calibration risk is discussed in Chapter 6, where the error function measuring the aggregate error to market prices is transformed suitably to result in a distribution on the parameter space. With this methodology, the framework developed in Chapter 3 can be applied and provides calibration risk-captured prices. We stated conditions under which a transformation of the error function is suitable and we provide an algorithm to construct a discrete distribution on the parameter space, approximating the continuous solution. Furthermore, an extensive numerical study was done to compare the parameter risk of different exotic options (Asian options, barrier options, lookback options) as well as the parameter risk of different models (Heston model, Γ-OU-Barndorff-Nielsen–Shephard model, Variance Gamma model). As a result, we found that the more exotic payoff profiles (particularly barrier options and – to a lesser extent – lookback options) exhibit considerably more parameter risk than the relatively good-natured Asian options, regardless of the selected model. Furthermore, in our study, calibration risk was less prevalent in the Heston model than in the BNS model or even the Variance Gamma model.

In case that there are bid-ask prices quoted for plain vanilla derivatives, in Chapter 7 we presented an approach to calibrate to these prices, and to obtain a *market-implied risk-capturing functional*. In case of incomplete markets and parametric distortion risk measures, this has already been done in (Cherny and Madan, 2010). We proofed the existence of a solution to this bid-ask calibration problem in different broad classes of distortion risk measures. Furthermore, a non-parametric calibration scheme was suggested which is based on a piecewise linear approximation. This approach adopts flexibly to virtually all available distortion functions and reduces the bid-ask calibration problem to a constrained optimization problem. In a numerical case study, we calibrated distortion risk measures to bid-ask prices of equity options employing the developed non-parametric approach. As a resulting distortion function, we obtained a characteristic pattern and introduce the parametric family of ess-sup-expectation convex combination, where calibration is considerably faster than in the non-parametric case.

Altogether, this thesis contains a rigorous framework to incorporate parameter risk into bid-ask prices, to apply the framework when parameters have to be estimated from time

series as well as when parameters are the result of a calibration to market prices, and to finally calibrate the suggested ansatz to quoted bid-ask prices of plain vanilla options in a non-parametric manner.

8.2 Critical reflection

This thesis focuses to treat parameter risk, i.e. it is assumed that a parameterized family of models is given and a probability measure on the parameter space describes the likelihood of the different parameters. Although the described framework can be equally used when facing model risk, it is challenging to describe model risk in a tractable scope in practice which is still exhaustive enough to cover the models of interest. First, it is difficult to collect all possible models in a model class $\mathcal{Q}$ which is rich enough to include most possible classes, but tractable enough for calculations. Even fairly general approaches as, e.g., the uncertain volatiliy model of (Avellaneda et al., 1995) rely on very restricting assumptions (in particular for the approach of (Avellaneda et al., 1995): diffusion dynamics, bounded volatility processes) and not nearly describe a model class including the most usual models for asset price dynamics. Second, even after having specified a suitable model class $\mathcal{Q}$ with reasonable candidate models, it may be even harder to determine the likelihood of the models, i.e. to specify a sensible distribution R on the models in doubt $\mathcal{Q}$ which reflects the single "likelihood" of each model. Even the uncertain volatility model of (Avellaneda et al., 1995) does not discuss the likelihood of different paths, but relies on a worst-case approach as described in (Cont, 2006).

Furthermore, the suggested methodology in Chapter 6 treating calibration risk and translating the errors to market prices into a distribution bears some pitfalls. First, it is the decision of the derivatives trader which transformation function to use, i.e. how to weight parameters with different error to market prices. Partly, this may be answered by a calibration to market prices as discussed in Chapter 7, but simultaneously calibrating to the transformation function and the risk measure may be an ill-posed problem. The weighting assigned to parameters by transforming the error function still remains a highly subjective choice. Second, as discussed in (Detlefsen and Härdle, 2007), different measurements of error to market prices may result in tremendous differences in the calibration result. Hence, they may also result in differences in derivatives prices. Thus, also in the choice of the error function, a lot of discretion comes into play.

8.3 Outlook

At least since the financial crisis of 2008, model and parameter risk and uncertainty are a topic which is discussed in a broader scope and even has found its place beyond the professional audience, which is manifested in general media articles as (Salmon, 2009). This thesis provides some ideas and approaches how to treat model and particularly parameter risk in derivatives pricing, but there are lots of other ideas where open questions still arise and may be answered by future research.

Considering the convergence properties of risk-captured prices, one could try to characterize weak (sequentially) convergence of risk-capturing functionals (or their generators), also by employing different domains. In this thesis, we primarily treat the convergence on the space of bounded and continuous functions $\mathcal{C}^b(\Theta)$, but have not obtained a characterization yet even on this good-natured domain. All convex risk measures with "worst-case part" do not exhibit continuity w.r.t. the weak topology, even when restricting the domain to $\mathcal{C}^b(\Theta)$. Meanwhile, further research on similar continuity properties has been done in (Krätschmer et al., 2012), but on different domains (Orlicz spaces) and w.r.t. stronger topologies (the so-called ψ-weak topology w.r.t. a gauge function ψ, which was introduced in (Weber, 2006)). Applying these techniques, one might equally deduct the results from Chapter 4.

A topic which is crucial for practical applications is finding hedging strategies that "replicate" parameter risk-captured prices: In complete models (e.g. the Black–Scholes model), the imposed model directly yields the (theoretical) existence of a dynamic hedging strategy for admissible contingent claims. In incomplete models, dynamic sub- and superhedging strategies can be used for sub-/superreplication. Furthermore, the theory of variance-minimal hedging delivers dynamic strategies minimizing the hedging error due to incompleteness. A hedging strategy $(\xi_t)_{t\geq 0}$ w.r.t. a European-style claim (without early exercise) X can be regarded as a solution of a backward stochastic differential equation (BSDE), satisfying a terminal condition that $\xi_T = X$. Hence, the works of (Drapeau et al., 2012) treating minimal supersolutions of BSDEs – superreplicating the claim payoff X with a hedging strategy $(\xi_t)_{t\geq 0}$ and $\xi_T \geq X$ is a step to establish superhedging strategies. Similar to worst-case pricing, superhedging may not be used in practice and leads to situations which are too conservative. Furthermore, the problem of finding minimal supersolutions is (up to now) only solved for the special case of incomplete markets. A first step could be to define the problem minimial dynamic superhedging problem in general model uncertainty and to look for conditions that and one

would rather "partial superhedging", e.g. by introducing stochastic orders in the style

$$\xi_T \geq_\Gamma X :\Leftrightarrow \Gamma(\xi_T - X) \geq 0.$$

Another possibility how the presented work could be generalized is to incorporate parameter risk as a time-dependent issue. When pricing derivatives, one obtains a snapshot for today's parameters, but is unsure about tomorrow's parameters, which are obtained via recalibration. With such a framework, one might ask whether time-dependent (dynamic) convex risk measures may be applied to capture time-dependent paramater risk. Since dynamic risk measures have close links to BSDEs, a generalization in this sense may lead to a similar question like possible risk-captured super-/subhedging.

Bibliography

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7):1505–1518.

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

Albrecher, H., Mayer, P., Schoutens, W., and Tistaert, J. (2007). The little Heston trap. *Wilmott Magazine*, 1:83–92.

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

Avellaneda, M., Levy, A., and Paras, A. (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance*, 2:73–88.

Bachelier, L. (1900). *Théorie de la spéculation*. Gauthier-Villars.

Balbás, A., Garrido, J., and Mayoral, S. (2009). Properties of distortion risk measures. *Methodology and Computing in Applied Probability*, 11(3):385–399.

Bannör, K., Kiesel, R., Nazarova, A., and Scherer, M. (2013). Model risk and power plant valuation. *Submitted to Energy Economics*.

Bannör, K. and Scherer, M. (2013a). Capturing parameter risk with convex risk measures. *European Actuarial Journal*, 3(1):97–132.

Bannör, K. and Scherer, M. (2013b). Handling model risk and uncertainty; illustrated with examples from mathematical finance. In Klüppelberg, C., Straub, D., and Welpe, L., editors, *Risk - An interdisciplinary introduction*.

Bannör, K. and Scherer, M. (2013c). On the calibration of distortion risk measures to bid-ask prices. *Submitted to Quantitative Finance*.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63:167–241.

Bartoszynski, R. (1961). A characterization of the weak convergence of measures. *The Annals of Mathematical Statistics*, 32(2):561–576.

Belomestny, D. and Krätschmer, V. (2012). Central limit theorems for law-invariant coherent risk measures. *Journal of Applied Probability*, 49(1):1–21.

Benth, F., Kiesel, R., and Nazarova, A. (2011). A critical empirical study of three electricity spot price models. *To appear in Energy Economics*.

Billingsley, P. (2009). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley.

Bion-Nadal, J. (2009). Bid-ask dynamic pricing in financial markets with transaction costs and liquidity risk. *Journal of Mathematical Economics*, 45(11):738–750.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Branger, N. and Schlag, C. (2004). Model risk: A conceptual framework for risk measurement and hedging. *Working Paper*, 98(2):1–37.

Bunnin, F., Guo, Y., and Ren, Y. (2000). Option pricing under model and parameter uncertainty using predictive densities. *Statistics and Computing*, 12.

Burger, M., Graeber, B., and Schindlmayr, G. (2008). *Managing energy risk: An integrated view on power and other energy markets*, volume 425. Wiley.

Carr, P., Geman, H., and Madan, D. (2001). Pricing and hedging in incomplete markets. *Journal of Financial Economics*, 62:131–167.

Carr, P., Geman, H., Madan, D., and Yor, M. (2006). Self-decomposability and option pricing. *Mathematical Finance*, 17(1):31–57.

Carr, P. and Madan, D. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2:61–73.

Cartea, A. and Figueroa, M. (2005). Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335.

Černy, A. (2009). *Mathematical Techniques in Finance: Tools for Incomplete Markets*. Princeton University Press, second edition.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., and Montrucchio, L. (2011). Risk measures: rationality and diversification. *Mathematical Finance*, 21(4):743–774.

Cherny, A. and Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies*, 22:2571–2606.

Cherny, A. and Madan, D. (2010). Markets as a counterparty: An introduction to conic finance. *International Journal of Theoretical and Applied Finance*, 13(8):1149–1177.

Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4):407–432.

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3):519–547.

Cont, R. and Tankov, P. (2004). *Financial Modelling With Jump Processes*. Chapman and Hall/CRC Financial Mathematics Series.

Denneberg, D. (1994). *Non-additive measure and integral*. Kluwer Academic Publishers.

Detlefsen, K. and Härdle, W. (2007). Calibration risk for exotic options. *The Journal of Derivatives*, 14(4):47–63.

Drapeau, S., Heyne, G., and Kupper, M. (2012). Minimal supersolutions of convex BSDEs. *Working Paper*.

Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182.

Figlewski, S. (1998). Derivatives risks, old and new. *Wharton-Brookings Papers on Financial Services*.

Fisher, R. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, 10(4):507–521.

Föllmer, H. and Leukert, P. (1999). Quantile hedging. *Finance and Stochastics*, 3(3):251–273.

Föllmer, H. and Leukert, P. (2000). Efficient hedging. *Finance and Stochastics*, 4(2):117–146.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447.

Föllmer, H. and Schied, A. (2004). *Stochastic Finance*. De Gruyter, second edition.

Föllmer, H. and Schweizer, M. (1990). Hedging of contingent claims under incomplete information. In *Applied Stochastic Analysis*, pages 389–414.

Frittelli, M. and Scandolo, G. (2006). Risk measures and capital requirements for processes. *Wiley: Mathematical Finance*, 16(4):589–612.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer.

Green, T. and Figlewski, S. (1999). Market risk and model risk for a financial institution writing options. *The Journal of Finance*, 54(4):1465–1499.

Guillaume, F. and Schoutens, W. (2011). Calibration risk: Illustrating the impact of calibration risk under the Heston model. *Review of Derivatives Research*.

Gupta, A. (2009). *A Bayesian Approach to Financial Model Calibration, Uncertainty Measures and Optimal Hedging*. PhD thesis, University of Oxford.

Gupta, A. and Reisinger, C. (2012). Robust calibration of financial models using Bayesian estimators. *To appear in Journal of Computational Finance*.

Gupta, A., Reisinger, C., and Whitley, A. (2010). Model uncertainty and its impact on derivative pricing. In Bocker, K., editor, *Rethinking Risk Measurement and Reporting*, pages 137–175.

Gzyl, H. and Mayoral, S. (2008). On a relationship between distorted and spectral risk measures. *Revista de economía financiera*, 15:8–21.

Hambly, B., Howison, S., and Kluge, T. (2009). Modelling spikes and pricing swing options in electricity markets. *Quantitative Finance*, 9(8):937–949.

Hanson, R. (1986). Linear least squares with bounds and linear constraints. *SIAM Journal on Scientific Computing*, 7(3):826–834.

Hanson, R. and Haskell, K. (1982). Algorithm 587: Two algorithms for the linearly constrained least squares problem. *ACM Transactions on Mathematical Software*, 8(3):323–333.

Hardy, G., Littlewood, J., and Pólya, G. (1934). *Inequalities*. Cambridge University Press.

Harrison, J. and Pliska, S. (1983). A stochastic calculus model of continuous trading: Complete markets. *Stochastic Processes and their Applications*, 15(3):313–316.

Heitfield, E. (2009). Parameter uncertainty and the credit risk of collateralized debt obligations. *Working Paper*.

Heston, S. (1993). 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.

Hörmander, L. (1990). *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis*. Springer.

Hürlimann, W. (2004). Distortion risk measures and economic capital. *North American Actuarial Journal*, 8(1):86–90.

Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*. Springer, second edition.

Jessen, C. and Poulsen, R. (2012). Empirical performance of model for barrier option valuation. *To appear in Quantitative Finance*.

Jouini, E., Schachermayer, W., and Touzi, N. (2006). Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, 9(1):49–71.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291.

Kemna, A. and Vorst, A. (1990). A pricing method for options based on average asset values. *Journal of Banking and Finance*, 14(1):113–129.

Klenke, A. (2008). *Probability Theory*. Springer.

Knight, F. (1921). *Risk, uncertainty, and profit*. Hart, Schaffner & Marx.

Knight, K. (2000). *Mathematical Statistics*. Texts in Statistical Science Series. Taylor & Francis.

Knittel, C. and Roberts, M. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, 27(5):791–817.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.

Krätschmer, V. (2006). On σ-additive robust representation of convex risk measures for unbounded financial positions in the presence of uncertainty about the market model. *SFB 649, Deutsche Forschungsgemeinschaft*.

Krätschmer, V., Schied, A., and Zähle, H. (2012). Comparative and qualitative robustness for law-invariant risk measures. *Working Paper*.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95.

Lindström, E. (2010). Implication of parameter uncertainty on option prices. *Advances in Decision Sciences*, page 15 pages.

Lucia, J. and Schwartz, E. (2002). Electricity prices and power derivatives: Evidence from the Nordic Power Exchange. *Review of Derivatives Research*, 5(1):5–50.

Madan, D., Carr, P., and Chang, E. (1998). The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105.

Madan, D. and Senata, E. (1990). The variance gamma model for share market returns. *The Journal of Business*, 63(4):511–524.

Margrabe, W. (1975). The value of an option to exchange one asset for another. *The Journal of Finance*, 23(1):177–186.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144.

Nicolato, E. and Venardos, E. (2003). Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Mathematical Finance*, 13(4):445–466.

Olkin, I. and Pratt, R. (1958). Unbiased estimation of certain correlation coefficients. *The Annals of Mathematical Statistics*, 29(1):201–211.

Raible, S. (2000). *Lévy Processes in Finance: Theory, Numerics, and Empirical Facts*. PhD thesis, Albert-Ludwigs-Universität Freiburg i. Br.

Salmon, F. (2009). Recipe for disaster: The formula that killed Wall Street. *Wired Magazine*.

Samuelson, P. (1965). Rational theory of warrant pricing. *Industrial Management Review*, 6(2):13–39.

Schmeidler, D. (1989). Subjective probability and expected utility wihout additivity. *Econometrica*, 57(3):571–587.

Schoutens, W., Simons, E., and Tistaert, J. (2004). A perfect calibration! Now what? *Wilmott Magazine*, 3:66–78.

Schweizer, M. (1991). Option hedging for semimartingales. *Stochastic Processes and their Applications*, 37(2):339–363.

Shilov, G. (1996). *Elementary functional analysis*. Dover Publications.

Stahl, G., Zheng, J., Kiesel, R., and Rühlicke, R. (2012). Conceptualizing robustness in risk management. *Working Paper*.

Stein, E. and Stein, J. (1991). Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies*, 4(4):727–752.

van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.

Weber, S. (2006). Distribution invariant measures, information, and dynamic consistency. *Mathematical Finance*, 16(2):419–442.

Werner, D. (2011). *Funktionalanalysis*. Springer, seventh edition.

Xu, M. (2006). Risk measure pricing and hedging in incomplete markets. *Annals of Finance*, 2:51–71.

Yang, M., Blyth, W., Bradley, R., Bunn, D., Clarke, C., and Wilson, T. (2008). Evaluating the power investment options with uncertainty in climate policy. *Energy Economics*, 30(4):1933–1950.

List of Tables

5.1 Estimated parameters of the clean spark spread model 82
5.2 Relative width of the parameter risk-captured bid-ask spread 84

6.1 Calibration environment of the assessed models 104
6.2 Statistical properties of Asian price distributions 105
6.3 Statistical properties of barrier price distributions 108
6.4 Statistical properties of lookback price distributions 111

7.1 Calibration performance of different distortion function families 137
7.2 Calibration performance of different distortion function families, including the ess sup-exp-family 140

List of Figures

5.1 Exchange option price as a function of the stocks' correlation 71
5.2 Different risk-captured prices as a function of the sample size 72
5.3 The AVaR-induced bid-ask prices for different significance levels 73
5.4 The entropic-induced bid-ask prices for different risk-aversion parameters . 74
5.5 Evolution of the power, gas, and carbon prices 78
5.6 Evolution of the clean spark spread 79
5.7 Parameter risk-implied bid-ask spread w.r.t. the diffusion components, normal jumps. 85
5.8 Parameter risk-implied bid-ask spread w.r.t. the diffusion components, Laplace jumps. 86

5.9 Parameter risk-implied bid-ask spread w.r.t. the jump size distribution, normal jumps. 86
5.10 Parameter risk-implied bid-ask spread w.r.t. the jump size distribution, Laplace jumps. 87

6.1 Cumulative distribution functions of BNS, Heston and VG price distributions for the Asian call option 106
6.2 Scatterplots of the price distributions and parameter risk-captured prices for the Asian option 107
6.3 Cumulative distribution functions of BNS, Heston and VG price distributions for the ITM barrier call optio 109
6.4 Scatterplots of the price distributions and parameter risk-captured prices for the ITM barrier option 110
6.5 Cumulative distribution functions of BNS, Heston and VG price distributions for the lookback call option 115
6.6 Scatterplots of the price distributions and parameter risk-captured prices for the lookback option 116
6.7 Relative deviations from the plug-in price compared among different option types 117

7.1 Calibration results of different methodologies (parametric and non-parametric) to calibrate to bid-ask prices 146
7.2 Bid-ask spreads for calibrated distortions compared to market bid-ask spreads 147

Index

atomless, 36
Average-Value-at-Risk, 26

Barndorff-Nielsen–Shephard model, 101
bid-ask calibration problem, 121
Black–Scholes model, 39

Choquet integral, 30
convergence property (CP), 55
convex risk measure, 25

distortion function, 34
    ess sup-expectation convex combination, 138
    continuous part, 138
    jump part, 138
    minmaxvar, 35
distortion risk measure, 34

entropic risk measure, 50
error function, 91
    root mean square error, 91
exchange option, 69

Fatou property, 27
Fenchel–Moreau transform, 27

Heston model, 39, 100

incomplete market, 38
indicator function, 30

Karush–Kuhn–Tucker conditions, 132

Margrabe option, 69
model risk, 41
model uncertainty, 38

parameter uncertainty, 38
piecewise linear concave distortion function, 126

risk, 12
risk-captured price, 44
risk-capturing functional, 44

Schwartz function, 95
set function, 29
    additive, 29
    monotone, 29
    submodular, 29
    supermodular, 29
significance level, 26
simple function, 29
spectral risk measure, 28

uncertain volatility model, 40
uncertainty, 12

Value-at-Risk, 26
variance gamma model, 102