Pricing and hedging in incomplete financial markets

Wurth, A.M.

Publication date:
2009

Pricing and Hedging in Incomplete Financial Markets

Andreas Würth

Pricing and Hedging in Incomplete Financial Markets

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Ruth First zaal van de Universiteit op vrijdag 8 mei 2009 om 14.15 uur door

Andreas Martin Würth

geboren op 19 maart 1973 te Zürich, Zwitserland.

PROMOTOR:

prof.dr. J.M. Schumacher

PROMOTIECOMMISSIE:

prof.dr. H. Föllmer
dr. R.J.A. Laeven
prof.dr.ir. C.W. Oosterlee
prof.dr. A.A.J. Pelsser
prof.dr. J.M. Schumacher
prof.dr. B.J.M. Werker

THOMAS STIELTJES INSTITUTE
FOR MATHEMATICS

Preface

This thesis consists of seven chapters, where Chapter 2 and Chapter 5 are joint work with J.M. Schumacher.

Acknowledgments

When I arrived at Tilburg, a lot of administrative things had to be done. Thanks to Tilburg University, and in particular to CentER, for helping me concerning those issues. In particular, CentER organized the housing for me, which was very helpful.

From the beginning of my stay in Tilburg, I had contact with my supervisor, Hans Schumacher. He guided me through the modern financial mathematics, and gave me many useful comments for my thesis. I would like to thank him a lot for all that.

Furthermore, in the last year of my PhD, I had many stimulating discussions with Roger Laeven, who gave me some useful comments for my thesis. Thanks a lot also to him.

I had the opportunity to take some of my doctoral courses at mathematical departments, in particular stochastics at the University of Amsterdam and numerics at the Delft University of Technology. Not only that those courses were very helpful for my research, they also gave me the opportunity to get in contact with Peter Spreij and Kees Oosterlee, which could give me many useful suggestions and with whom I had some stimulating discussions. Thanks to them.

I would like also to thank Tilburg University, and in particular CentER, for having access to many journals, whose articles were important for my research.

Furthermore, CentER offered me the opportunity to attend conferences, at which I could present my papers, contact some people with different backgrounds, have some stimulating discussions, and obtain new research ideas.

Thanks also to all members of my PhD committee, that is, apart from my supervisor, to Hans Föllmer, Kees Oosterlee, Antoon Pelsser, Roger Laeven, and Bas Werker. They gave me useful suggestions for improving my thesis.

Thanks also to the members of the Econometrics and Operations Research department. I had the opportunity to talk with them about the current research. In particular, apart from the committee members, thanks to Feico Drost and Pavel Čížek.

I would like to thank all other people from outside of the university which gave me useful advise for my research. Apart from the ones already mentioned, thanks in particular to F. Delbaen and M. Frittelli.

From the point of view of my nonacademic life in Tilburg, I would like to thank all people that I met here and who made my life enjoyable.

Contents

Chapter 1

Introduction

After the revolutionary works of Black and Scholes [10] as well as Merton [63], a large amount of literature has been published on the topic of Mathematical Finance. From an empirical point of view, the model of Black and Scholes was unsatisfactory because it did not cover some stylized facts of stock prices such as large kurtosis, volatility clusters, and non-constant implied volatility, see e.g. Fama [30], Rubinstein [75], or Bollerslev et al. [12]. Therefore, a lot of attention has been directed towards the study of more sophisticated models, for example Eberlein and Özkan [28], which are more closely related to the real world stock price processes.

From a theoretical point of view, one main question was which properties of the price processes of securities follow from some reasonable economical assumptions. In Merton [63], some basic assumptions have been formulated for obtaining option prices. Those assumptions are in principle similar, but slightly stronger, than the generally accepted assumption of no arbitrage, This assumption means that it is not possible for an investor to make money with zero initial endowment at no risk. It is quite reasonable, because if such an opportunity would exist, everyone would be interested in that, so that this opportunity would not exist for a long time. As a consequence, a model which does not satisfy this property does not seem to be very reasonable, and therefore almost all financial models proposed in the literature satisfy it. The notion of arbitrage, its connection to an economic equilibrium as well as to the existence of a pricing measure, has been established by Kreps [58]. Delbaen and Schachermayer have proved in their famous paper [24] that absence of arbitrage in a financial market is equivalent to the existence of a local martingale measure. This measure gives a price system for the options. The same authors have generalized this result in [26].

From a practical point of view, the question is, if one has a specific model for the security price processes, how to find the corresponding option prices. In the model of Black and Scholes, a replication argument can be used for deriving this price. This means that one derives a trading strategy using only the underlying asset and a risk-free asset, which gives without risk the same terminal payoff as the option. As a consequence, the option price must be the same as the initial endowment for the trading strategy, because otherwise there would be an arbitrage opportunity. This leads to an explicit pricing formula for standard options. Furthermore, this method also shows how to replicate, that is how to hedge, this option.

In more sophisticated models than the one of Black and Scholes, there is in general no explicit pricing formula anymore, and an important research question is how to find a price numerically. Some standard techniques are the Finite Difference method described in Wilmott et al. [84], or Monte Carlo simulation as described in Glasserman [37]. Furthermore an approach using Fourier transformation has been described in Heston [44], and for numerical issues in Carr and Madan [16].

And even if for many models one still has the replication argument as in the model of Black and Scholes, there are some models where not all options are replicable. Such models are called incomplete, whereas the other ones are called complete.

From a mathematical point of view, the market is incomplete if the amount of randomness is too large compared with the amount of possible trading strategies. Typical examples of such incomplete models are models with jumps, such as Lévy processes described in Cont and Tankov [22]. If the jump size of the underlying is continuously distributed, it is not possible to replicate even standard options with a finite number of underlying assets. But even if the price processes are continuous and driven by a multivariate Brownian motion, it is not possible to fully replicate all contingent claims if the dimension of the Brownian motions is larger than the amount of risky assets which can be traded, as it was shown by Karatzas et al. [53].

Another issue are restrictions in the trading strategies. For example, one may study the model where only simple strategies are allowed, or another one where the strategies have to satisfy short-selling constraints. The latter has been considered by Schmock et al. [77]. Finally, there may also be some restrictions for the information that can be used when defining those strategies, i.e. the strategies must be measurable with respect to a smaller filtration than the one under which all financial processes are adapted.

From a practical point of view, incompleteness can arrive in situations where it makes sense or is even necessary to model the financial market in a way as described above. To model it using jump processes makes sense, for example, in situations where credit default risk plays a role. Eberlein and Özkan have also applied jump processes for term structure models [29]. A different example is a hedge fund, where the returns of the underlying hedge fund are only reported on a monthly basis, so that the financial process is actually a discrete-time process.

Practical situations where even a model without jumps leads to incompleteness are, for example, situations where the payoff depends on an insurance risk, such as a catastrophe in reinsurance, or death or longevity in life insurance. There is typically no financial asset which replicates the payoff of such kinds of events. When taking models with stochastic volatility for the stock price as in Heston [44], one needs not only the underlying stock price, but also an option for being able to replicate all contingent claims. For long term claims as well as in markets which are not so well developed, one cannot expect that there are enough options for being able to trade them continuously, and therefore no full hedge is possible.

Finally, if there are transaction costs as for example in Hodges and Neuberger [48], or more recently in Monoyios [66], a full replication would be too expensive, and therefore the market also becomes incomplete.

Incompleteness of the financial market has several consequences for pricing as well as for hedging. Because a full replication of the option payoff is not possible, one cannot take the replication argument as pricing principle. Because a seller of such an option cannot replicate it, he has to accept a certain level of risk. Therefore, as typically done since a long time by actuaries in the insurance industry, the seller charges a certain risk loading for accepting the risk. But neither the magnitude of this risk loading nor the trading strategy for a possibly partial hedge of the payoff is given by the general no arbitrage assumption. The equivalent martingale measure which determines the price system in the market is not unique. The no-arbitrage assumption only gives some bounds for the price, but not a specific price. Such price bounds can typically be very wide. It follows from the optional decomposition theorem (Kramkov [57]) that a payoff is superreplicable at the supremum of all its possible prices which are determined by the set of equivalent martingale measures. This also implies that if this measure is unique, there exists a replicating strategy.

For being able to determine a unique price as well as hedge, additional assumptions on the behavior of the investors have to be made. There exists a considerable number of papers, for example Hodges and Neuberger [48], or Henderson [41], which assume that the investor preferences are given by the expected utility, where a von Neumann-Morgenstern utility function is given. The price of a financial claim is then given by the so-called first-order indifference price, that is the price at which the investor, when investing optimally, is indifferent of first order between buying or selling this claim. Using this assumption about the investor behavior, one can find a unique (first-order indifference) pricing measure, as well as an optimal investment policy in an incomplete market, or an optimal hedge.

Pricing using the indifference pricing measure can be seen as a completion of the market by an equilibrium argument. In the theory of representative agents, see for example Duffie [27], a utility function of a representative agent is given by

$$u_\lambda(x) = \sum_{i=1}^{n} \lambda_i u_i(x),$$

where n is the number of investors and u_i the utility function of investor i. Here, it is assumed that the preferences are expressed in terms of expected utility, an issue which does not need to be assumed in general. In an equilibrium, the market portfolio must be optimal for this representative investor, that is

$$E[u_\lambda(X)] \leq E[u_\lambda(M)]$$

for all random variables X which can be achieved in the market at a price smaller than or equal to the one of M, where M is the market portfolio.

In principle, it is also possible for single investors to trade new assets which have not been traded before, if someone is willing to buy and another one is willing to sell such a non-traded asset. Obviously, such assets have to add up to zero. In this sense, we can think of the market as complete by assuming a pricing measure for every non-traded asset. Obviously, the market portfolio of this completed market is still M.

Let this pricing measure now be fixed and the market be the completed one. If the prices are such that we have an equilibrium, it follows from Duffie [27] that under rather mild conditions, there is a representative investor, i.e. there is a λ such that $u_\lambda(x)$ as defined above is the utility function of the representative investor. It follows that M must be optimal for this investor.

In our research, we assume that this λ is already fixed, i.e. $u := u_\lambda$ is a given utility function characterizing the preferences of the representative agent. Because M is optimal for this investor, it follows for all (new or already existing) assets X that for each $\alpha \in \mathbb{R}$,

$$E[u(M + \alpha(X - \pi(X)))] \leq E[u(M)],$$

where $\pi(X)$ is the price of this asset X. The differential of the left hand side with respect to α at $\alpha = 0$ must therefore be 0, assuming enough regularity in order to take this differential. It follows that π must be the first-order indifference price of X. For simplicity, the first-order indifference price will be denoted simply indifference price for the rest of this thesis.

Actually the two issues, finding the indifference pricing measure and the optimal strategy, are connected to each other via the duality theory. One of the first papers which have applied this theory for the investment problem was Karatzas et al. [53], where the relationships for the specific case where the financial market is given by a multivariate Brownian motion are discussed. In the sequel, a large amount of generalizations has been published. In 1999 and 2001, Kramkov and Schachermayer [56] and Schachermayer [76] have provided a duality theory for general semimartingales if the investor preferences are given by utility functions satisfying the Inada conditions. In Bouchard et al. [14], duality relationships have also been shown for nonsmooth utility functions. The key issue of all the papers about duality theory in the context of expected utility preferences is that the indifference pricing measure is given as the minimum expectation of the function which is conjugate to the utility function, where the minimum is taken over all densities of absolutely continuous martingale measures, modulo a proportionality factor. At least theoretically, this gives a way how to find the indifference price, even if the computation of a specific model still remains a challenge.

From a practical point of view, a question is how to find the pricing measure and the optimal strategy for a specific model. This is a much more delicate issue than in complete markets, as there is typically an optimization problem to solve. In some situations, in particular when using exponential utility functions, efficient numerical algorithms have been provided, such as in Grasselli and Hurd [38]. An alternative approach is followed in Chen et al. [17], where a Taylor approximation is used for obtaining an approximation of the indifference price.

A different idea from indifference pricing is to look for bounds rather than for a specific price. In order to obtain bounds which are narrower than the ones obtained by the no-arbitrage assumption, Cochrane and Saá-Requejo [21] assumed that the Sharpe ratio can only lie within some specified bounds. It has been shown that this is equivalent to the assumption that the Radon-Nikodym density of the pricing measure is only allowed to have a limited variance.

In recent years, a different assumption on the investor behavior has attracted considerable attention. The idea is that investors are risk averse, and that they need some capital for accepting some risk. In 1999, Artzner et al. [3] gave an axiomatic framework of what is a coherent risk measure. One of the first papers on the acceptability of claims in incomplete markets was in Carr et al. [15]. In the sequel, several theoretical papers have been published about the idea of risk measure pricing, a recent one being Klöppel and Schweizer [55]. The idea of this pricing method is that the price of a claim should be the minimal amount of money that is needed for the claim to become acceptable in terms of a coherent risk measure, under optimal hedging. Despite some theoretical work on this issue, there are only few papers which are dedicated to the concrete (numerical or analytical) solution of a specific risk measure pricing problem, and in Ilhan et al. [49], the numerical challenge arising in the solution of such problems is explicitly mentioned.

This thesis focuses on financial markets which are incomplete, and contributes to the existing literature theoretically as well as numerically. In a first part, the question is treated whether the indifference pricing measure is equivalent to the objective measure in a measure-theoretical sense. This question is important because, if this measure would not be equivalent, it would not be reasonable to take it as pricing measure. Indeed, if an investor would price new options in the market using an absolutely continuous but not equivalent pricing measure, he would allow some counterparties to make arbitrage, according to the fundamental theorem of asset pricing (Delbaen and Schachermayer [24]).

The question of equivalence of the indifference pricing measure has already been treated in some papers during the last ten years. For the specific case of the variance-optimal measure, the one which minimizes the expected squared density function, equivalence has been proved in 1996 by Delbaen and Schachermayer [25], provided the market admits at least one square-integrable martingale measure which is equivalent, and provided the price processes are continuous. For price processes given by general semimartingales, equivalence has been proved if the preferences of investors are given by an exponential utility function, in Frittelli [36], provided there exists at least one equivalent martingale measure which is integrable with respect to the conjugate function of the exponential one. In 2002 [52], Kabanov and Stricker generalized this work to essentially all bounded and strictly increasing utility functions. In the same year, Bellini and Frittelli [7] have formulated the notion of the minimax martingale measure, which is essentially the indifference pricing measure, but which does explicitly take into account the current wealth of the investor. This issue is important for most utility functions, because the optimal strategies as well as the indifference pricing measure depend in general on the current wealth. In a recent paper of Kabanov and Sirbu [51], conditions are given for having an indifference price which is at least locally independent of the current wealth. In the same paper as mentioned above, Bellini and Frittelli proved the equivalence of this measure for the case where the utility function is unbounded from above. Finally, in 2004 [14], Bouchard et al. extended this result to the case of nonsmooth utility functions, where the formulation has again been changed slightly, but has essentially the same meaning as the minimax martingale measure.

This thesis contributes to the current research in the way that it proves the equivalence of the indifference pricing measure for the case of utility functions which are nondecreasing but not strictly increasing, that is they have a satiation point. Examples of such functions are the quadratic utility function which gives the variance-optimal measure from Delbaen and Schachermayer [25]. The work in this thesis can therefore be seen as an extension of this example to general satiated utility functions. Further examples are power-utility functions with power larger than 1, but also preferences given by a convex loss function such as in the analysis of the expected shortfall of Föllmer and Leukert [33], see also Gundel and Weber [39].

For utility functions which are satiated, there exist simple counterexamples to the equivalence result when taking general semimartingales as price processes. This thesis focuses on the case where the filtration is continuous, which implies that all price processes as well as density processes are continuous. Furthermore, assumptions on the relative risk aversion of the functions have to be established, which are trivially satisfied by the variance-optimal martingale measure. With those assumptions, a theorem similar to the ones for unsatiated investors is proved in this thesis, namely that the indifference pricing measure is equivalent, provided there exists a martingale measure which is sufficiently integrable. For doing this, the notion of the minimax martingale measure from Bellini and Frittelli [7] is used as indifference pricing measure.

The intention of the proof of equivalence is to do this not only for smooth utility functions, but also for nonsmooth ones as long as they are satiated. Because relative risk aversion is typically defined for smooth utility functions, it is not a priori clear how to extend it to more general situations. Moreover, there does not seem to exist literature on this issue.

As a consequence, before proving the equivalence of the minimax martingale measure for satiated investors, the first part of the thesis, Chapter 2, is dedicated to define risk aversion for nonsmooth utility functions and its consequences. Actually, the well-known results of Pratt [68] from 1964 are extended to utility functions which are not strictly increasing nor necessarily differentiable but only nondecreasing and concave. For this issue, risk aversion is defined as a measure, from which the utility function can be reconstructed up to positive affine transformations, as in the case of classical risk aversion. In the sequel, the notion what it means that one utility function is more risk averse than another one is established, and an extension of [68] is proved. Finally, the fact that a utility function has bounded relative risk aversion is connected to a super-power property of the function, an issue that is used to prove the equivalence of the minimax martingale measure.

For proving the equivalence of the minimax martingale measure for satiated utility functions, which is the subject of Chapter 3, this thesis connects this problem to the one of finding a minimal martingale measure, where minimal means minimal with respect to the expectation of a Young function. This is a generalization of the case of Delbaen and Schachermayer [25] where this Young function is the quadratic one. One can then use Luxemburg norms and Orlicz space techniques. The idea of using this method for a related problem has been used in Biagini and Frittelli [9], where the focus was on proving existence of optimal primal and dual solution of investment problems. For proving equivalence, the same ideas are used as in [25], but for general Young functions. In particular, the Cauchy-Schwartz inequality turns into a generalization of the Hölder inequality, an issue which is established for Young functions in this thesis. Finally, examples are given which show what can go wrong if one of the assumptions on the relative risk aversion is not satisfied.

In the second part of the thesis, the focus is changed from theoretical considerations to some specific models. The aim of this is to calculate numerically a price as well as a hedge of a claim in an incomplete market. First, in Chapter 4, a model with stochastic volatility is considered, where it is assumed that there are only few options in the market, so that only the underlying asset can be used for trading. Incompleteness then arises from the fact there are two Brownian motions as sources of risk (stock price as well as volatility risk), but only one traded asset apart from the risk-free one. Practical examples of such situations may be markets where options are rarely traded such as emerging markets or specific commodity markets, or options with a time horizon which is much larger than the one of typical options in the market. Furthermore, option prices can be used to calibrate utility-based models, as done in this thesis.

Numerical indifference pricing has been done recently using exponential utility functions in Grasselli and Hurd [38], Lim [60], and Monoyios [66]. Using power utility, indifference pricing for stochastic volatility models has been considered theoretically in Monoyios [67] and Hobson [47]. In particular, these papers show that the price of volatility risk in such situations is given by a specific nonlinear partial differential equation, which can in principle be solved by a Finite Difference method as well as by the Feynman-Kac representation using a Monte Carlo simulation.

In this thesis, the results are used in order to compare the Finite Difference algorithm with the Monte Carlo simulation with respect to their numerical efficiency. The partial differential equation is nonlinear, and its solution is a nontrivial problem. The contribution of this thesis to the research is therefore to show how the algorithms can be implemented for solving the indifference pricing problem, as well as to compare them and state their strengths and weaknesses. Furthermore, it is shown how the price of volatility risk is used in order to determine the indifference price of an option by the solution of the pricing PDE. Several methods are compared for this problem, e.g. Hilber et al. [45] or Clarke and Parrot [20]. Finally, a calibration of the price of volatility risk to the S&P option prices is performed.

Another specific situation where incompleteness occurs is is the one of unit-linked insurance products. This is the issue of chapter 5. Typically, the financial claims can be hedged, as long as the financial market is driven by a multivariate Brownian motion. On the other hand, there exists no financial asset which replicates a claim of 1 if someone dies and 0 otherwise. In a unit-linked insurance product where the payoff depends on the financial as well as on the insurance outcome, the claim can therefore not be fully hedged, but partially, because it has a financial component. On the other hand, for the risk of the insurance company, it will ask for some risk loading.

The way in which an optimal hedge is defined, and how large the risk loading is, again depends on the investor preferences. Typically, an insurance company which sells such a claim aims to minimize its risk. In practice, quantile-based risk measures are common, in particular Value at Risk. In this thesis, Conditional Value at Risk (CVaR) is taken, because it is in some way similar to Value at Risk, but coherent, as shown in Artzner et al. [3]. The way of pricing an insurance payoff is therefore the method of risk measure pricing as discussed above. However, instead of making general theoretical considerations, one specific risk measure, the Conditional Value at Risk, is applied to the specific situation of pricing an insurance payoff. The aim is then to have either an explicit formula or a numerical solution for this price.

In general, CVaR pricing is a very challenging issue, when the aim is to obtain specific numerical results. To date, there are only a few papers which are related to this. For example, Sekine [80] treats the problem of minimizing worst conditional expectation. In Melnikov and Skornyakova [62], a type of Value at Risk pricing is treated, and some analytical formulas are obtained. For doing this, results by Föllmer and Leukert [32] are applied which connect the problem of quantile hedging with the Neyman-Pearson lemma. The same authors also wrote a paper on minimization of expected shortfall [33], but this expected shortfall should not be understood in the sense of CVaR.

The contribution of this thesis to the research is that for the CVaR pricing problem of unit-linked insurance payoffs, analytical formulas or at least numerically easily feasible solutions are obtained, which no longer need any optimization algorithm. This goal is achieved first by the frequently used assumption that financial and insurance process are independent, such as has been done in most of the considerations in Møller [65]. Second, a general property of CVaR stated in Rockafellar and Uryasev [73] is applied, in order to connect the CVaR minimization problem to the problem of minimization of expected shortfall in the sense of Föllmer and Leukert [33]. Furthermore, an extension of one of the main theorems in [33] is proved.

Furthermore, one nontrivial and unrealistic assumption is used, namely that the information about the insurance process is only available at the end of the time period in which the insurance policy runs. However, the CVaR price which is obtained by the algorithm proposed in this thesis gives an upper bound for the true CVaR price without this assumption, and numerical examples suggest that this bound is already quite good.

The question arises if one could approximate the case of continuously arriving information by the methodology proposed in Chapter 5 and making the information time steps smaller and smaller. This is indeed possible, and it is treated in Chapter 6. The idea is already given in Föllmer and Leukert [32]. However, this thesis gives a proof which shows that making the information time steps smaller and smaller leads indeed to a CVaR price which converges to the one when information arrives continuously. However, the computational effort for doing this increases a lot, with the result that numerical efficiency again becomes a big issue.

Chapter 2

Risk aversion for nonsmooth utility functions

2.1 Introduction

In the economic literature, the notion of risk aversion plays quite a large role in characterizing investor preferences. Risk aversion is often defined as the Arrow-Pratt coefficient of absolute or relative risk aversion. However, the classical definition of this coefficient assumes differentiability of order two of the utility function, which is not always satisfied in examples. One typical example would be a piecewise linear utility function. Our aim is therefore to give a definition which coincides with the classical one in the case of differentiability of order two, but is also applicable in all other cases.

Nonsmooth utility functions have been applied in some papers, for example in Bouchard et al. [14]. However, to our knowledge, a definition of the absolute or relative risk aversion for such cases has not been formulated before.

Why is the question of risk aversion for nonsmooth utility functions interesting? In Segal and Spivak [79], the authors argue that a nondecreasing concave utility function is differentiable almost everywhere, and non-differentiability is therefore of no importance in expected utility theory. However, in the same paper, it is noted that first-order risk aversion arises in some non-expected utility theories, and they prove that locally approximating utility functions must in this case be nondifferentiable. Furthermore, a smooth utility function can be reconstructed from its risk aversion. This cannot be done for utility functions which are concave and piecewise linear but not linear without considering the null set where the function is not differentiable.

The risk aversion of such a function is almost everywhere zero, and a reconstruction would give a linear utility function, which is not what we want. Moreover, the fact that a piecewise linear utility function has the same risk aversion as a linear one does not make sense even in the case of expected utility theory, and even if the random variables would have a continuous distribution. It follows that the risk aversion at the null set where the function is not differentiable has also to be considered.

For nonsmooth functions, there are several ways of defining derivatives. The general method uses distribution theory. Another generalization can be made specifically for concave functions, for which there exists the theory of the generalized first differential (subdifferential for convex functions, see for example Rockafellar [70]), as well as some literature about a generalized second derivative such as in Rockafellar [71]. The idea of defining the second differential for convex functions as a Lebesgue-Stieltjes measure, which we will partially follow, is also not new. In principle, one could try to use one of those definitions for generalizing the classical formula for risk aversion. However, this would still require an appropriate definition of a quotient of those generalized derivatives, an issue which is not trivial.

In this chapter, we follow a slightly different approach. Instead of defining a quotient of the generalized first and second differential, we define the risk aversion directly as a Lebesgue-Stieltjes measure. This measure is generated from the logarithm of the marginal utility (defined again by subdifferential calculus). The absolute as well as the relative risk aversion is then defined as a Radon-Nikodym derivative, provided it exists. If the utility function is strictly increasing and twice differentiable, this Radon-Nikodym derivative coincides with the classical definition of the (absolute or relative) risk aversion. Therefore, our definition is indeed an extension of the classical one.

Our definition allows us to compare the risk aversion measures of different utility functions. This gives us the opportunity to compare risk aversions of different investors. Those comparisons always hold for absolute as well as for relative risk aversion. Using this comparison, we prove an extension of a classical result of Pratt [68], which connects the fact that u_1 is more risk averse than u_2 to the existence of a concave function T such that $u_1(x) = T(u_2(x))$, for the case of nonsmooth utility functions. We show also that our definition is consistent with another definition about what the notion of being more risk averse means, and which does not assume utility functions.

For the absolute risk aversion, we can typically take the whole real line as domain. However, the relative risk aversion has to been taken relative to a specific wealth. Typically, it is calculated relative to the current wealth, by the formula that relative risk aversion is current wealth times absolute risk aversion. As a consequence, the relative risk aversion is always zero at wealth zero. Therefore, it cannot be expected that there exists a good definition for utility functions which are defined on the positive as well as on the negative real line. Typically, the relative risk aversion is taken for functions which are defined only for positive wealth. In this chapter, we will also focus on this case, but alternatively consider the case where it is negative as well. We will treat both cases in a unifying way whenever this is possible, and separately if not.

We also introduce a second, weaker notion of ordering between risk aversions, which we call essential bounds for the risk aversion. We will give an example of a piecewise linear utility function which has essentially constant relative risk aversion. Furthermore, we will show that strict bounds are always essential bounds for the risk aversion, and therefore the definition of essential bounds is indeed a relaxation of the former definition.

Constant relative risk aversion is connected to power utilities. In the same way, constant absolute risk aversion is connected to exponential utility functions. We will generalize this feature to functions whose associated risk aversion is bounded from above or from below, where we no longer have the power (or exponential) property, but at least an inequality which gives in some sense a super-power, respectively super-exponential property. We will formulate this issue first for strict bounds for the risk aversion, and then for essential bounds as well.

Finally, we will show that the upper bound of the relative risk aversion of such a function translates into a lower bound of the concave conjugate function, and vice versa, an issue which is well-known for power utility functions. We will do this for strict as well as for essential bounds.

The outline of the chapter is as follows. In section 2.2, we treat smooth utility functions, where we review inequality relationships between absolute risk aversion and exponential functions (as well as relative risk aversion and power functions). In section 2.3, we give our generalized definition of the risk aversion measure, and prove the connection to power utility as well as the translation to bounds of the conjugate function. In section 2.4, we do the same for essential bounds for the risk aversion. Section 2.5 concludes.

2.2 Definition and relationships for smooth utility functions

The classical definition of risk aversion from the literature is the following.

Definition 2.2.1. [Absolute and relative risk aversion] Let $u(x)$ be strictly increasing and concave for all $x \in D$, where $D \subset \mathbb{R}$ is the domain, and at least twice differentiable. Then the *absolute risk aversion* of $u(x)$ for $D \subset \mathbb{R}$ is defined by

$$ara(x) := -\frac{u''(x)}{u'(x)}, \tag{2.2.1}$$

and, in the case $D \subset]0, \infty[$, the *relative risk aversion* of $u(x)$ is defined by

$$rra(x) := -\frac{xu''(x)}{u'(x)}. \tag{2.2.2}$$

We will call the functions $ara(x)$ and $rra(x)$ risk aversion densities, for reasons which will become clear later. It is a well-known fact that constant absolute risk aversion density is connected to exponential utility functions, whereas constant relative risk aversion density is connected to power utility functions. The following proposition says that for bounded risk aversion density, one has a super-exponential or a super-power property in the case of the absolute or relative risk aversion density, respectively. Furthermore, for the relative risk aversion density, one has a translation of the risk aversion density to the one of its dual function.

Proposition 2.2.2. *Let the utility function be strictly increasing, concave and twice differentiable. Furthermore, assume that it is defined and strictly larger than $-\infty$ on the whole real axis for the absolute risk aversion, and on $]0, \infty[$ for the relative risk aversion case. Then*

1. *The absolute risk aversion density is bounded from below by a nonnegative constant γ if and only if the function $e^{\gamma x}u'(x)$ is nonincreasing, i.e. if*

$$\frac{u'(y)}{u'(x)} \leq e^{-\gamma(y-x)} \tag{2.2.3}$$

for all $x < y$, $x, y \in \mathbb{R}$. Analogously, it is bounded from above by a nonnegative constant γ if and only if $e^{\gamma x}u'(x)$ is nondecreasing.

2. *The relative risk aversion density is bounded from below by a nonnegative constant γ if and only if the function $x^{\gamma}u'(x)$ is nonincreasing, i.e. if*

$$\frac{u'(y)}{u'(x)} \leq \left(\frac{y}{x}\right)^{-\gamma} \tag{2.2.4}$$

for all $x < y$, $x, y \in]0, \infty[$. Analogously, it is bounded from above by a nonnegative constant γ if and only if $x^{\gamma}u'(x)$ is nondecreasing.

3. *If the utility function satisfies the Inada conditions, that is if $u :]0, \infty[\to \mathbb{R}$ is strictly increasing and concave, with $u'(0^+) = \infty$ and $u'(x) \downarrow 0$ as $x \to \infty$,*

then the dual function $u^*(y) := \inf_x(xy - u(x))$ *is strictly increasing and twice differentiable as well, and for the relative risk aversion density, one has*

$$rra_{u^*}(y) = \frac{1}{rra_u((u^*)'(y))}. \tag{2.2.5}$$

Remark 2.2.3. For the absolute risk aversion density, there is no relation of the type (2.2.5).

Proof.

1. Because $u'(x)e^{\gamma x}$ is differentiable, this function is nonincreasing if and only if

$$u''(x)e^{\gamma x} + \gamma u'(x)e^{\gamma x} = e^{\gamma x}\left(u''(x) + \gamma u'(x)\right) \leq 0,$$

which is equivalent to the condition that the term in the brackets on the right hand side is nonpositive. This is equivalent to

$$\gamma \leq -\frac{u''(x)}{u'(x)} = ara(x)$$

by the fact that $u'(x) > 0$.

2. Again by the differentiability, the condition that $x^\gamma u'(x)$ is nonincreasing is equivalent to

$$u''(x)x^\gamma + u'(x)\gamma x^{\gamma-1} = x^{\gamma-1}\left(xu''(x) + \gamma u'(x)\right) \leq 0$$

which is equivalent to the condition that the term in the brackets on the right hand side is nonpositive, by the fact that $x > 0$. This is equivalent to

$$\gamma \leq \frac{-xu''(x)}{u'(x)} = rra(x).$$

3. By the fact that u is smooth, the infimum of the function $x \mapsto xy - u(x)$ is attained at the point x satisfying $y = u'(x)$, and it follows that

$$u^*(y) = y(u')^{-1}(y) - u((u')^{-1}(y)),$$

where the inverse exists by the fact that u satisfies the Inada conditions. Furthermore, one has $(u')^{-1}(y) = (u^*)'(y)$ by standard differential calculus, from which it follows that u^* is strictly increasing and twice differentiable. Applying formula (2.2.2) to $u^*(y)$ one obtains

$$rra_{u^*}(y) = -\frac{y(u^*)''(y)}{(u^*)'(y)} = -\frac{u'(x)}{xu''(x)} = \frac{1}{rra_u(x)} = \frac{1}{rra_u((u^*)'(y))}$$

by standard differential calculus and the fact that $(u')^{-1}(y) = x$.

□

The following corollary shows how Proposition 2.2.2 can be used to make a connection between bounded risk aversion density and a super-exponential or super-power property of the utility function.

Corollary 2.2.4.

1. *Assume that relation (2.2.3) holds with $\gamma > 0$, and that $u(\infty) = 0$. Then, for all $x < y$, $x, y \in \mathbb{R}$, we have*

$$u(y) \geq u(x)e^{-\gamma(y-x)}.$$

2. *Assume that relation (2.2.4) holds, and assume that $\gamma < 1$ (an analogous statement also holds for $\gamma > 1$, which we will state later in a more general context). Further assume that $u(0) = 0$. Then, for all $x < y$, $x, y \in]0, \infty[$, we have*

$$u(y) \leq u(x)\left(\frac{y}{x}\right)^{1-\gamma}.$$

Proof.

1. Let $x < y < \infty$. Then

$$\begin{aligned}-u(x) &= -u(y) + \int_x^y u'(\xi)d\xi \geq -u(y) + u'(y)\int_x^y e^{-\gamma(\xi-y)}d\xi \\ &= -u(y) + \frac{u'(y)}{-\gamma}\left(1 - e^{-\gamma(x-y)}\right)\end{aligned}$$

by equation (2.2.3) and the fact that $\xi < y$ in the domain of integration. Furthermore,

$$-u(y) = \int_y^\infty u'(\xi)d\xi \leq u'(y)\int_y^\infty e^{-\gamma(\xi-y)}d\xi = \frac{u'(y)}{\gamma}$$

again by (2.2.3) and the fact that $\xi > y$ in the domain of integration. It follows that

$$\begin{aligned}-u(x) &\geq -u(y) + \frac{u'(y)}{\gamma}\left(e^{-\gamma(x-y)} - 1\right) \geq -u(y)\left(1 + e^{-\gamma(x-y)} - 1\right) \\ &= -u(y)e^{-\gamma(x-y)}\end{aligned}$$

by the fact that $e^{-\gamma(x-y)} > 1$ for $y > x$. Moving the exponential function to the left hand side, the result follows.

2. Let $0 < x < y$. Then

$$\begin{aligned} u(y) &= u(x) + \int_x^y u'(\xi)d\xi \leq u(x) + u'(x) \int_x^y \left(\frac{\xi}{x}\right)^{-\gamma} d\xi \\ &= u(x) + \frac{u'(x)x}{1-\gamma} \left(\left(\frac{y}{x}\right)^{1-\gamma} - 1\right), \end{aligned}$$

since $\xi > x$ in the integration domain and (2.2.4). Furthermore, again by (2.2.4),

$$u(x) = \int_0^x u'(\xi)d\xi \geq u'(x) \int_0^x \left(\frac{\xi}{x}\right)^{-\gamma} d\xi = \frac{u'(x)x}{1-\gamma}$$

because here, $\xi < x$. It follows that

$$\begin{aligned} u(y) &\leq u(x) + \frac{u'(x)x}{1-\gamma} \left(\left(\frac{y}{x}\right)^{1-\gamma} - 1\right) \leq u(x) \left(1 + \left(\frac{y}{x}\right)^{1-\gamma} - 1\right) \\ &= u(x) \left(\frac{y}{x}\right)^{1-\gamma}, \end{aligned}$$

since $(\frac{y}{x})^{1-\gamma} > 1$.

□

Remark 2.2.5. A normalization to $u(\infty) = 0$ is possible if the utility function is bounded from above, which is for example the case if the absolute risk aversion density is bounded from below by a positive constant. Indeed, if $u'(x)e^{\gamma x}$ is nonincreasing, then for $x > 0$, $u'(x)e^{\gamma x} \leq c$ with the constant $c = u'(0) > 0$, which implies that $u(x) \leq u(0) + \frac{c}{\gamma}(1 - e^{-\gamma x})$.

Similarly, a normalization to $u(0) = 0$ is possible in the case where the domain is $]0, \infty[$, if the utility function is bounded from below. This is for example the case if the relative risk aversion density is bounded from above and $\gamma < 1$. Because $u'(x)x^\gamma$ is nondecreasing, we have $u'(x)x^\gamma \leq u'(1) =: c$, and therefore $u(1) - u(\epsilon) \leq \frac{c}{1-\gamma}(1 - \epsilon^{1-\gamma})$. As $\epsilon \downarrow 0$, the left hand side of this equation must remain bounded.

2.3 A generalized definition of risk aversion

2.3.1 Assumptions

First, we have to specify the domain. For absolute risk aversion, the domain can be chosen to be the whole real line. On the other hand, the typical case where relative risk aversion makes sense is when the utility function is concave and defined on the positive real line. This is the typical case which we will treat with the most emphasis. Alternatively, we will also treat the case where the utility function is only defined on the negative real line. We will see that mathematically, this is the same as if we talk about risk loving instead of risk averse investors, and a utility function which is convex and defined on the positive real line.

Assumption 2.3.1. In the case of absolute risk aversion, the wealth can be any value in $\mathbb{R}$, i.e. $D = \mathbb{R}$. On the other hand, for relative risk aversion, the wealth is either positive or negative, that is the domain is either $D =]0, \infty[$ or $D =]-\infty, 0[$.

Assumption 2.3.2. The utility function $u : D \to \mathbb{R} \cup \{-\infty\}$ is nondecreasing, concave and upper semicontinuous on D.

Assumption 2.3.3. The utility function u is proper, i.e. there exists a point $x \in D$ with $u(x) > -\infty$.

Remark 2.3.4. If $D =]-\infty, 0[$ and $u(x)$ satisfies Assumption 2.3.2, then $\tilde{u}(x) := -u(-x)$ is defined on the positive domain, and is nondecreasing, convex and lower semicontinuous.

2.3.2 Risk aversion measure

The aim of this section is to provide a definition of the risk aversion which can be applied to all utility functions satisfying Assumptions 2.3.1 and 2.3.2. For this, we will introduce a measure which we denote risk aversion measure, of which the absolute as well as relative risk aversion will turn out to be a Radon-Nikodym derivative, if these derivatives exist.

Definition 2.3.5. [Superdifferential] Let $u : D \subset \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ be a concave function. Then the *superdifferential* δu of u at a point $x \in D$ is the set

$$\delta u(x) := \{s \in \mathbb{R} \cup \{\infty\} \mid u(y) - u(x) \leq s(y - x) \ \forall \ y \in D\}.$$

By the fact that u is concave, the superdifferential $\delta u(x)$ is always nonempty in the region where $u(x)$ is finite, and by Assumption 2.3.2, it is nonincreasing in the sense that for $x < y$ we have $z_x \geq z_y$ for every $z_x \in \delta u(x)$, $z_y \in \delta u(y)$. We may therefore uniquely define

$$u'_r(x) := \inf_y \{y \in \delta u(x)\}. \tag{2.3.6}$$

It is easy to show that $u'_r(x) \in \delta u(x)$ and the function $u'_r(x)$ is right-continuous and monotonically decreasing and therefore of finite variation, and we may define the (positive) Lebesgue-Stieltjes measure $-du'_r(x)$ without ambiguity if $-u(x) < \infty$. It is clear that if $x = \sup\{y \mid -u(y) = \infty\}$, we must have $\sup_y\{y \in \delta u(x)\} = \infty$. For consistency and for preserving the monotonicity of u'_r, we define therefore $u'_r = \infty$ for all x with $-u(x) = \infty$. If u would be twice continuously differentiable and strictly increasing, we would have, by rearranging the terms in equation (2.2.2), that $-d\ln u'$ is absolutely continuous with respect to $\operatorname{sgn}(x)d\ln|x|$, and hence there exists a unique Radon-Nikodym derivative $\gamma(x)$. The function $\operatorname{sgn}(x)$ is 1 if $x > 0$ (positive wealth) and -1 if $x < 0$ (negative wealth). This can also be expressed in the following way. The absolute risk aversion density on an interval I is $\gamma(x)$ if

$$\int_B -d\ln u_r' = \int_B \gamma(x)dx \tag{2.3.7}$$

for every Borel set $B \subset I$, or in differential notation

$$-d\ln u_r' = \gamma(x)dx,$$

and the relative risk aversion density is $\gamma(x)$ on an interval $I \subset D$ if

$$\int_B -d\ln u_r' = \int_B \gamma(x)\operatorname{sgn}(x)d\ln|x| \tag{2.3.8}$$

for every Borel set $B \subset I$, or in differential notation

$$-d\ln u_r' = \operatorname{sgn}(x)\gamma(x)d\ln|x|.$$

This definition is equivalent to 2.2.1 in the sense that the risk aversion densities according to both definitions coincide up to a Lebesgue null set. It now becomes clear why we call $\gamma(x)$ the risk aversion density.

The measure $-d\ln u_r'$ is defined principally for all utility functions satisfying Assumptions 2.3.1 and 2.3.2, but not necessarily on the whole domain D. Indeed, if $u(x)$ is constant after some point, or if $-u(x) = \infty$ for small values of x, the measure $-d\ln u_r'$ is not defined. Therefore, we define for the utility function u the domains

$$\begin{aligned} D^u_{\text{eff}} &:= \operatorname{int}(D \setminus (D^u_{\text{inf}} \cup D^u_{\text{sup}})), \\ D^u_{\text{ra}} &:= D \setminus \operatorname{int}(D^u_{\text{sup}}), \end{aligned} \tag{2.3.9}$$

where the function $\operatorname{int}(\cdot)$ means the interior of the set, and

$$\begin{aligned} D^u_{\text{inf}} &:= \{x \in D \mid u(x) = -\infty\}, \\ D^u_{\text{sup}} &:= \{x \in D \mid u(x) = \sup_{z\in D} u(z)\} \setminus D^u_{\text{inf}}. \end{aligned} \tag{2.3.10}$$

Remark 2.3.6. The effective domain as defined above for a utility function u is

$$D^u_{\text{eff}} = \{x \in D \mid \exists\epsilon > 0 : 0 < u_r'(y) < \infty \ \forall y \in B_\epsilon(x)\}.$$

It follows that the measure $-d\ln u_r'$ is sigma-finite on D^u_{eff}. Actually the measure is finite on all compact intervals of D^u_{eff}.

We turn now to the general definition of the risk aversion measure.

Definition 2.3.7. [Risk aversion measure] Let $u(x)$ be a utility function satisfying Assumptions 2.3.1 and 2.3.2. Then the *risk aversion measure* of u is defined as the following measure ρ on the Borel sets of D^u_{ra}: $\rho = -d\ln(u'_r)$ on D^u_{eff}, and for all $x \in D^u_{\mathrm{ra}} \setminus D^u_{\mathrm{eff}}$, $\rho(\{x\}) = \infty$.

Remark 2.3.8. It follows that ρ is sigma-finite on D_{eff}, whereas on $D_{\mathrm{ra}} \setminus D_{\mathrm{eff}}$ it is obviously not sigma-finite.

If this measure is absolutely continuous with respect to $d\ln|x|$, then u has a relative risk aversion density $\gamma(x)$ on an interval $I \subset D$ if the measure $-d\ln u'_r$ satisfies equation (2.3.8) on every Borel subset $B \subset I$. The same holds for the absolute risk aversion density if the measure is absolutely continuous with respect to dx.

Remark 2.3.9. If $u'(x)$ exists, the measure $-d\ln u'(x)$ is identical to $-d\ln u'_r(x)$. But replacing $-d\ln u'(x)$ by $-d\ln u'_r(x)$, we may apply the notion of risk aversion measures for all utility functions satisfying Assumptions 2.3.1 and 2.3.2.

In the classical utility theory, there is a one-to-one relationship between risk aversions and equivalence classes of utility functions, where two utility functions are equivalent if they can be mapped to each other by a positive affine transformation. This has been pointed out for instance in Pratt [68]. Similarly, using our definition of the risk aversion measure, one has a one-to-one relationship between a suitable class of measures and the set of equivalence classes of utility functions. This issue will be treated in the sequel.

Definition 2.3.10. A *risk aversion set* D_{ra} associated to a given domain D is a subset of the domain D as stated in Assumption 2.3.1, which is of the form $D_{\mathrm{ra}} = D\cap]-\infty, b]$ for some b $\in \mathbb{R}$.

Definition 2.3.11. A measure μ on the Borel sets of a risk aversion set $D_{\mathrm{ra}} = D\cap]-\infty, b]$ is said to be a *ρ-finite* measure if there exists a constant $a \leq b$ such that $\mu(\{x\}) = \infty$ if $x \in D_{\mathrm{ra}} \setminus]a, b[$ and if for each compact subset $K \subset]a, b[$ we have $\mu(K) < \infty$.

Let us recall the classical definition of equivalence classes of utility functions.

Definition 2.3.12. Two utility functions u and v satisfying Assumptions 2.3.1 and 2.3.2 are *equivalent* if there exist constants $c \in \mathbb{R}$ and $d > 0$ such that $u = c + dv$.

Proposition 2.3.13. *For each equivalence class of utility functions u, there exist a unique risk aversion set D_{ra} and a unique ρ-finite measure μ on D_{ra} such that $D_{\mathrm{ra}} = D^u_{\mathrm{ra}}$ according to equation (2.3.9) and μ is the risk aversion measure of any* *utility function u of this class. On the other hand, for each risk aversion set D_{ra} and each ρ-finite measure μ on it, there exists a utility function u, unique up to equivalence, such that $D_{\mathrm{ra}} = D^u_{\mathrm{ra}}$ and μ is the risk aversion measure of u.*

Proof. For a specific utility function u, we have that D^u_{ra} from equation (2.3.9) is a risk aversion set, since by the concavity $D^u_{\sup}$ must be of the form $[b,\infty[$ with $b \leq \infty$. It follows from what has been done before that there exists a measure μ which is ∞ at each point of $D^u_{\mathrm{ra}} \setminus D_{\mathrm{eff}}$, where $D_{\mathrm{eff}} =]a,b[$ is an open interval, and with $\mu = -d\ln u'_r$ on D_{eff}. Let $K \subset]a,b[$ be compact. Then there exist constants $a < \gamma < \delta < b$ with $K \subset]\gamma,\delta]$, and $\mu(]\gamma,\delta]) = \ln u'_r(\gamma) - \ln u'_r(\delta) < \infty$, i.e. μ is a ρ-finite measure. If v is of the same equivalence class as u, then $D^u_{\mathrm{ra}} = D^v_{\mathrm{ra}}$, $D^u_{\mathrm{eff}} = D^v_{\mathrm{eff}}$ and $v = c + du$ with $d > 0$, $v'_r = du'_r$, and $\ln v'_r = \ln d + \ln u'_r$. It follows that $d\ln v'_r = d\ln u'_r$ on $D^u_{\mathrm{eff}} = D^v_{\mathrm{eff}}$, and therefore the risk aversion measures ρ_u and ρ_v corresponding to u and v are the same.

Let now a risk aversion set D_{ra} and a ρ-finite measure μ be given. Let us first assume that $D_{\mathrm{eff}} \neq \emptyset$, where $D_{\mathrm{eff}} := \{x \in D \mid \mu(\{x\}) < \infty\}$. Then, for an $x_0 \in]a,b[= D_{\mathrm{eff}}$, we define a function

$$F(x) := \begin{cases} \int_{]x_0,x]} d\mu(\xi) & \text{if} \quad b > x > x_0, \\ -\int_{]x,x_0]} d\mu(\xi) & \text{if} \quad a < x < x_0, \\ 0 & \text{if} \quad x = x_0. \end{cases}$$

It follows that $F(x)$ is right-continuous, finite for all $x \in]a,b[$, and nondecreasing. Now we define $g(x) := e^{-F(x)}$, then it follows that $g(x)$ is right-continuous, strictly positive, nonincreasing, and finite on $]a,b[$, and therefore also integrable on compact sets in $]a,b[$. Define

$$u(x) := \begin{cases} \lim_{\xi\uparrow b} u(\xi) & \text{if} \quad x \geq b, \\ \int_{x_0}^x g(\xi)d\xi & \text{if} \quad a < x < b, \\ \lim_{\xi\downarrow a} u(\xi) & \text{if} \quad x = a, \\ -\infty & \text{if} \quad x < a. \end{cases}$$

Then u is nondecreasing, concave, and upper semicontinuous, and because $u(x) = -\infty$ for $x < a$ and $u(x)$ is constant for $x \geq b$, $D^u_{\mathrm{eff}} \subset]a,b[$. It is clear that g is in the superdifferential of u, and because it is nonincreasing and right-continuous, $u'_r = g$ on $]a,b[$. It follows that $0 < u'_r < \infty$ on $]a,b[$, and $D^u_{\mathrm{eff}} =]a,b[= D_{\mathrm{eff}}$, and $D_{\sup} = [b,\infty[$, from which it follows that $D_{\mathrm{ra}} = D^u_{\mathrm{ra}}$ according to equation (2.3.9). Furthermore, on $]a,b[$, $\ln u'_r = -F$, and therefore, for a half-open interval $]x,y] \subset]a,b[$, $x > a$, one has $-d\ln u'_r(]x,y]) = F(y) - F(x) = \mu(]x,y])$. By the right-continuity of F and the properties of measures, this must also hold as $x \downarrow a$.

It remains to show that for any other v satisfying Assumptions 2.3.1 and 2.3.2 and for which $\rho_v = \mu$, it follows that $v = c + du$ with $c \in \mathbb{R}$ and $d > 0$. We have that $D_{\text{ra}} = D^u_{\text{ra}} = D^v_{\text{ra}}$, and from the fact that the risk aversion measures of u and v are the same, it follows that the domains $D_{\text{eff}} =]a, b[$ must coincide. Let now $a < x < y < b$. Then

$$\ln v'_r(x) - \ln v'_r(y) = -d \ln v'_r(]x, y]) = -d \ln u'_r(]x, y]) = \ln u'_r(x) - \ln u'_r(y)$$

and therefore (because $u'_r > 0$, $v'_r > 0$)

$$v'_r(y) = \frac{v'_r(x)}{u'_r(x)} u'_r(y).$$

Fixing an $x_0 \in]a, b[$, and applying the fundamental theorem of differential calculus [8] we have

$$v(x) = v(x_0) + \int_{x_0}^{x} v'_r(\xi) d\xi = v(x_0) + \frac{v'_r(x_0)}{u'_r(x_0)} \int_{x_0}^{x} u'_r(\xi) d\xi = v(x_0) + \frac{v'_r(x_0)}{u'_r(x_0)} (u(x) - u(x_0)),$$

hence v is equivalent to u on D_{eff}. For $x \geq b$, it follows by the concavity and monotonicity and the fact that $D \setminus D_{\text{ra}} = \text{int}(D_{\text{sup}})$ that $v(x)$ must be constant for $x \geq b$ as well, and $v(x) = \lim_{\xi \uparrow b} v(\xi) = c + d \lim_{\xi \uparrow b} u(\xi) = c + du(x)$. For $x = a$, one has the limiting argument ($\xi \downarrow a$) by the fact that the functions are nondecreasing and upper semicontinuous. For $x < a$, both utility functions are $-\infty$ because $D^u_{\text{eff}} = D^v_{\text{eff}}$, and the result still holds.

If $D_{\text{eff}} = \emptyset$, the following cases are possible:

1. $D_{\text{ra}} = \emptyset$, then any utility function u must be constant.

2. $D_{\text{ra}} = D$, then $u(x) = -\infty$.

3. $D_{\text{ra}} =] -\infty, b]$, then $u(x) = -\infty$ for $x < b$, and equal to a constant for $x \geq b$.

In all cases, those properties require uniqueness up to positive affine transformations.

□

2.3.3 Comparison of risk aversions

If one has the classical absolute or relative risk aversion density, one has a partial ordering on the set of utility functions according to which u_1 is more risk averse than u_2 if their absolute or relative risk aversion densities γ_1 and γ_2 satisfy $\gamma_1 \geq \gamma_2$. With the notion of the risk aversion measure, one can extend this definition to a partial ordering of the risk aversions for all utility functions satisfying Assumptions 2.3.1 and 2.3.2. This partial ordering coincides with the classical ordering in the case of absolute continuity of the risk aversion measure.

Definition 2.3.14. [Comparison of risk aversions] Let $I \subset D$ be an interval, and $u_1(x), u_2(x)$ two utility functions with ρ_1 and ρ_2 their associated risk aversion measures. Then we say that $\rho_1 \leq \rho_2$ on I if $D_{\text{ra}}^{u_1} \cap I = D_{\text{ra}}^{u_2} \cap I =: D_{\text{ra}} \cap I$ and for all Borel sets $B \subset D_{\text{ra}} \cap I$ we have that

$$\rho_1(B) \leq \rho_2(B).$$

Remark 2.3.15. If γ_1 and γ_2 are the (relative or absolute) risk aversion densities corresponding to ρ_1 and ρ_2, it follows that $\gamma_1 \leq \gamma_2$ on I if $\rho_1 \leq \rho_2$ on I, provided the densities exist.

Remark 2.3.16. It follows that if $\gamma_1 \leq \gamma_2$, then ρ_1 is absolutely continuous with respect to ρ_2. Furthermore, if ρ_{u_2} is absolutely continuous with respect to $\operatorname{sgn}(x) d \ln |x|$, it follows that $\gamma_1(x) \leq \gamma_2(x)$, almost surely with respect to the measure $\operatorname{sgn}(x) d \ln |x|$ on I.

One may think about a general definition of the notion "more risk averse" without use of utility functions. Such a definition could be given in the following way:

Definition 2.3.17. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\succeq$ a partial ordering on its random variables. Then a random variable X on $(\Omega, \mathcal{F}, P)$ is *unacceptable* if for all random variables Y on $(\Omega, \mathcal{F}, P)$ we have $Y \succeq X$ and there exists a random variable $\hat{Y}$ with $\hat{Y} \succ X$.

Definition 2.3.18. [Comparison of risk aversions without utility functions] Let I_1 and I_2 be two investors and $(\Omega, \mathcal{F}, P)$ a probability space. Let there be preference relations $\succeq_1$ and $\succeq_2$ on the set of random variables on $(\Omega, \mathcal{F}, P)$, corresponding to the preferences of I_1 and I_2. Then I_1 is *more risk averse* than I_2 if for all constants $w \in D$ and for all random variables X on $(\Omega, \mathcal{F}, P)$ which map to D one has

$$w \succeq_2 X \Rightarrow w \succeq_1 X \tag{2.3.11}$$

as well as, for all X on $(\Omega, \mathcal{F}, P)$ which are not unacceptable for investor I_1 one has

$$w \preceq_1 X \Rightarrow w \preceq_2 X. \tag{2.3.12}$$

In the case in which preferences are given by expected utilities, this implication may also be written in the following way:

$$\{x \mid v(x) \geq Ev(X)\} \subset \{x \mid u(x) \geq Eu(X)\} \tag{2.3.13}$$

for all random variables X on $(\Omega, \mathcal{F}, P)$, and

$$\{x \mid u(x) \leq Eu(X)\} \subset \{x \mid v(x) \leq Ev(X)\} \tag{2.3.14}$$

for all random variables X with $Eu(X) > -\infty$, where investor 1 and investor 2 have the utility functions $u(\cdot)$ and $v(\cdot)$ respectively. In this case we shall also simply say that u is more risk averse than v.

Roughly speaking, this definition says that a more risk averse investor always prefers a certain outcome to the risk if the less risk averse investor does.

We will now show that Definition 2.3.14 is equivalent to Definition 2.3.18 in the case where preferences are expressed in terms of expected utilities, no matter whether or not the utility functions are smooth. Actually, we prove a theorem first given by Pratt [68], but now without any assumption on differentiability or strict monotonicity. This gives another argument that our definition of the risk aversion measure is reasonable.

Theorem 2.3.19. *Let u and v be two utility functions that satisfy Assumptions 2.3.1 and 2.3.2. Then the following statements are equivalent:*

1. *$D^u_{\text{ra}} = D^v_{\text{ra}}$ and on this domain, $\rho_u \geq \rho_v$, where ρ_u and ρ_v are the risk aversion measures corresponding to u and v.*
2. *There exists a nondecreasing concave function $T : \mathbb{R} \cup \{-\infty\} \mapsto \mathbb{R} \cup \{-\infty\}$ with $u(x) = T(v(x))$ for all $x \in D$, which is strictly increasing on $v(D)$ unless $T(y) = -\infty$, i.e. for all $y_1 < y_2 \in v(D)$ we have either $T(y_1) < T(y_2)$ or $T(y_2) = -\infty$.*

Assume furthermore that we have a probability space $(\Omega, \mathcal{F}, P)$ which admits a random variable with a continuous distribution function. Then statement 2 is equivalent to

3. *u is more risk averse than v in the sense of Definition 2.3.18.*

Remark 2.3.20. Statement 3 is similar to the ones about risk premium and certainty equivalent of Pratt. However, our statement connects to a situation which does not use utility functions and can therefore be applied to more general situations.

The proof of Theorem 2.3.19 uses some technical lemmas. The first one contains a chain rule of superdifferentials. A lemma for subdifferentials for nondecreasing convex functions has been stated in Hiriart-Urruty and Lemarechal [46]. We refer to this lemma.

Lemma 2.3.21. *Let $v : U \subset \mathbb{R} \to V \subset \mathbb{R} \cup \{-\infty\}$ and $T : V \to W \subset \mathbb{R} \cup \{-\infty\}$ be two upper semicontinuous, concave and nondecreasing functions, and $u(x) := T(v(x))$. Then, for any $x \in \mathbb{R}$,*

$$\delta u(x) = \{\Delta_1 \Delta_2 : \Delta_1 \in \delta T(v(x)), \Delta_2 \in \delta v(x)\}.$$

Proof. Set $f(x) := -T(-x)$ and $g(x) := -v(x)$. Then f is a nondecreasing convex and g a convex function, and Theorem 4.3.1 in [46] about the chain rule for subdifferentials can be applied. The statement of Lemma 2.3.21 now follows. □

Lemma 2.3.22. *If u is more risk averse than v, then $D^v_{\text{inf}} \subset D^u_{\text{inf}}$ and $D^v_{\text{sup}} = D^u_{\text{sup}}$.*

Proof. If $D^v_{\text{sup}} \not\subset D^u_{\text{sup}}$, then there exist x_1 and x_2 in D such that $u(x_2) > u(x_1)$ and $v(x_2) = v(x_1)$. We have $x_1 \in \{x \mid v(x) \geq v(x_2)\}$, whereas $x_1 \notin \{x \mid u(x) \geq u(x_2)\}$, hence (2.3.13) is violated for the degenerate random variable X that always takes the value x_2.

If $D^u_{\text{sup}} \not\subset D^v_{\text{sup}}$, then there exist x_1 and x_2 in D such that $u(x_2) = u(x_1) > -\infty$ and $v(x_2) > v(x_1)$. It follows that $x_2 \in \{x \mid u(x) \leq u(x_1)\}$, but $x_2 \notin \{x \mid v(x) \leq v(x_1)\}$, so that (2.3.14) is violated for $X = x_1$.

If $D^v_{\text{inf}} \not\subset D^u_{\text{inf}}$, then there exists $x \in D$ such that $u(x) > -\infty$, whereas $v(x) = -\infty$. It follows that $\{y \mid v(y) \geq v(x)\} = D$, but, for $x_1 < x$, we have $u(x_1) < u(x)$. This is true because otherwise, $x \in \text{int}(D^u_{\text{sup}})$ but also $x \in D^v_{\text{inf}}$, which is impossible by the former considerations. The inclusion (2.3.13) is therefore violated for the degenerate random variable $X = x$, because $x_1 \notin \{y \mid u(y) \geq Eu(X)\}$. □

Lemma 2.3.23. *If the risk aversion measure of u is larger than or equal to the one of v, then $D^u_{\text{eff}} \subset D^v_{\text{eff}}$.*

Proof. It is clear from the definition that $D^u_{\text{ra}} = D^v_{\text{ra}}$. Let x be an element of D^u_{eff}. It follows that $x \in D^v_{\text{ra}}$, and that $\{x\}$ is a compact subset of D^u_{eff}. Therefore the risk aversion measure of u, ρ_u, satisfies $\rho_u\{x\} < \infty$. If $x \in (D^v_{\text{eff}})^c$, then $\rho_v(\{x\}) = \infty$, which is a contradiction. □

Proof of Theorem 2.3.19. First we will prove the equivalence of statements 1 and 2, because they do not use any probability space. Afterwards, we will prove the equivalence of statements 2 and 3.

Let statement 2 be correct, i.e. assume that there is a concave function T with $u(x) = T(v(x))$. Then, by Lemma 2.3.21, we must have that $u'_r(x) = T'_r(v(x))v'_r(x)$, from which it follows for any interval $]x, y] \subset D^u_{\text{eff}}$ that

$$-d \ln u'_r(]x, y]) = -d \ln v'_r(]x, y]) + \ln \left(\frac{T'_r(v(x))}{T'_r(v(y))} \right).$$

If $y \in D^u_{\text{sup}}$, then $y \in D^v_{\text{sup}}$ since T is strictly increasing on $v(D)$. It follows that $D^u_{\text{ra}} = D^v_{\text{ra}}$. Otherwise, if $y \in D^u_{\text{eff}}$, we have $0 < u'_r(y) = T'_r(v(y))v'_r(y) < \infty$, and therefore the argument of the logarithm is strictly larger than 0 and hence a valid expression. If $y \in (D^u_{\text{eff}})^c$, then statement 1 follows because $\rho_u(]x, y]) = \infty$. We can therefore restrict the discussion to values of $y \in D^u_{\text{eff}}$.

Because T_r' is nonincreasing and $v(x)$ is nondecreasing, it follows from $y > x$ that $-d\ln u_r'(]x,y]) \geq -d\ln v_r'(]x,y])$ on $D_{\text{eff}}^u \cap D_{\text{eff}}^v$. This holds for all intervals in this set. If B is a general Borel set in $D_{\text{eff}}^u \cap D_{\text{eff}}^v$, the result follows from Lemma 2.3.27, which we will prove in the next section. But $D_{\text{eff}}^v \supset D_{\text{eff}}^u$ because T is monotonic and concave, and it follows that $T(-\infty) = -\infty$ and therefore $D_{\text{inf}}^v \subset D_{\text{inf}}^u$. If B is not contained in D_{eff}^u, then $\rho_u(B) = \infty$, and therefore the result still holds. Therefore the inequality holds in all cases, and the implication is proved.

Conversely, let statement 1 be correct. Define the measure ϕ by

$$\phi(B) := \rho_u(B) - \rho_v(B)$$

for all Borel sets B on D_{ra}. Obviously, ϕ is a positive measure. Let x_0 be contained in D_{eff} with respect to u and v. If such a point would not exist, then, because $D_{\text{eff}}^u \subset D_{\text{eff}}^v$ by Lemma 2.3.23, $D_{\text{eff}}^u = \emptyset$. In this case, we have either $D_{\text{ra}} = D$, or $D_{\text{ra}} =]-\infty, b] \cap D$. In the first case, $u(x) = -\infty$, $T(y) = -\infty$ satisfy the requirements. In the second case, $u(x) = -\infty$ for $x < b$, $u(x) = const$ for $x \geq b$ is the utility function corresponding to statement 1, and $T(y) = -\infty$ for $y < v(b)$, $T(y) = const$ for $y \geq v(b)$ the corresponding concave function. We can therefore for the rest of the proof assume that $D_{\text{eff}}^u \cap D_{\text{eff}}^v$ is nonempty, and define the function $F : \mathbb{R} \to \mathbb{R}$ by $F(x_0) := \ln u_r'(x_0) - \ln v_r'(x_0)$, and

$$F(x) := \begin{array}{lll} F(x_0) + \phi(]x, x_0]) & \text{if} & x < x_0, \\ F(x_0) - \phi(]x_0, x]) & \text{if} & x > x_0, \end{array}$$

from which it follows that F is nonincreasing and that $F(x) - F(y) = \phi(]x,y])$ for all $x < y$, and

$$u_r'(x) = e^{F(x)} v_r'(x).$$

On D_{eff}^v, $v : D_{\text{eff}} \to v(D_{\text{eff}})$ is invertible, with $D_{\text{eff}} =]a,b[$, and $v(D_{\text{eff}}) =]v(a), v(b)[$ is an open interval. Moreover, $u_r' = 0$ as well as $v_r' = 0$ for $x \geq b$, and $u_r'(x) = \infty$ as well as $v_r'(x) = \infty$ for $x < a$, and for $x \downarrow a$ both functions converge to their value at a. We can therefore define the function g by

$$g(y) = \begin{array}{lll} \infty & \text{if} & y < v(a), \\ \lim_{x \downarrow v(a)} g(x) & \text{if} & y = v(a), \\ e^{F(v^{-1}(y))} & \text{if} & v(a) < y < v(b), \\ 0 & \text{if} & y \geq b, \end{array}$$

from which it follows that g is nonincreasing, nonnegative, and strictly positive on $]v(a), v(b)[$, and

$$u_r'(x) = g(v(x)) v_r'(x).$$

Let now x_0 again be a point of D_{eff}, with respect to both functions u and v. If we define $G(x) := u(x_0) + \int_{v(x_0)}^{x} g(\xi)d\xi$, then G is a nondecreasing concave function, strictly increasing on $[v(a), v(b)] \supset v(D) \setminus \{G = -\infty\}$, which can be chosen to be upper semicontinuous, and we obtain

$$u(x) = u(x_0) + \int_{x_0}^{x} u_r'(\xi)d\xi = u(x_0) + \int_{x_0}^{x} g(v(\xi))v_r'(\xi)d\xi = G(v(x))$$

by the differential calculus [8].

Let now statement 2 be satisfied. Then there exists a nondecreasing, concave, and upper semicontinuous function T, strictly increasing on $v(D) \setminus \{T = -\infty\}$, with $u(x) = T(v(x))$ for all $x \in D$. Let X be an arbitrary random variable on $(\Omega, \mathcal{F}, P)$, and let x be an element of the set $\{y \mid v(y) \geq Ev(X)\}$. Because T is nondecreasing and concave, we have

$$u(x) = T(v(x)) \geq T(Ev(X)) \geq E[T(v(X))] = Eu(X)$$

and therefore $x \in \{y \mid u(y) \geq Eu(X)\}$, so that inequality (2.3.13) is satisfied. Let X be a random variable with $Eu(X) > -\infty$, and let x be in the set $\{y \mid u(y) \leq Eu(X)\}$. Then

$$T(v(x)) = u(x) \leq Eu(X) = ET(v(X)) \leq T(Ev(X)).$$

It follows that $T(Ev(X)) > -\infty$, and by the fact that T is strictly increasing on $v(D) \setminus \{T = -\infty\}$, we have $v(x) \leq Ev(X)$, and inequality (2.3.14) is satisfied.

Let now statement 3 be satisfied. Define

$$\begin{aligned} T(y) &:= \tilde{T}(y_+), \\ \tilde{T}(y) &:= \inf\{u(z_-) \mid v(z) \geq y\}. \end{aligned} \tag{2.3.15}$$

The definition is understood in the sense that $\tilde{T}(y) = \sup_{z \in D} u(z)$ when $y > \sup_{z \in D} v(z)$. Furthermore, if $D =]0, \infty[$, u and v have to be interpreted as their extension to the whole real line, with values $-\infty$ for negative x. To show that $u(x) = T(v(x))$ for all $x \in D$, we need to show that

$$u(x) = \lim_{\epsilon \downarrow 0} \inf\{u(z_-) \mid v(z) \geq v(x) + \epsilon\} \tag{2.3.16}$$

for all $x \in D$. First of all, consider the situation in which v is constant, say $v(x) = v_0$. In that case it follows from Lemma 2.3.22 that $D^u_{\text{sup}} = D$, so that u is constant as well, say $u(x) = u_0$ for all $x \in D$. Since $\inf_{z \in D} u(z) = \sup_{z \in D} u(z) = u_0$, it follows from Definition (2.3.15) that $T(y) = u_0$ for all y. Hence the relation $u(x) = T(v(x))$ is satisfied for all x.

In the rest of the proof, we assume that v is not constant. First consider the case in which $x \in D^v_{\text{sup}}$. The set D^v_{sup} is nonempty if and only if the function v has a saturation point. In this case, write $D^v_{\text{sup}} = [b, \infty[$ if $D \neq]-\infty, 0[$ and $[b, 0[$ if $D =]-\infty, 0[$, respectively. For all $x \in D^v_{\text{sup}}$, we have $\{z \mid v(z) \geq v(x) + \epsilon\} = \emptyset$ for all $\epsilon > 0$. Because $D^v_{\text{sup}} = D^u_{\text{sup}}$ by Lemma 2.3.22, we must have $u(x) = \sup_{z \in D} u(z) = u(b)$ for $x \in D^v_{\text{sup}}$, and by definition $T(v(x)) = \sup_x u(x) = u(b)$ for $x \in D^v_{\text{sup}}$.

Next assume that $x \in D^v_{\text{ra}} \setminus \text{int}(D^v_{\text{inf}})$. Because v is not constant, this set is nonempty and is an interval which is closed in D. If we add the limit points of D which may be $\pm\infty$, we can without loss of generality assume that this interval is $[a, b]$. On this interval, the function $v : [a, b] \to [v(a), v(b)]$ is continuous and strictly increasing, and has therefore an inverse function $v^{-1} : [v(a), v(b)] \to [a, b]$ which is again continuous and strictly increasing. If $x \in (D^v_{\text{ra}} \setminus \text{int}(D^v_{\text{inf}})) \setminus \{b\}$, we still have for ϵ small enough that $v(x) + \epsilon < v(b)$ and therefore $a < v^{-1}(v(x) + \epsilon) < b$. It follows that $\tilde{T}(v(x) + \epsilon) = u(v^{-1}(v(x) + \epsilon)_-) \geq u(v^{-1}(v(x))) = u(x)$ by the monotonicity of u and the strict monotonicity of v^{-1}. By the right-continuity of u which is the same as upper semicontinuity for nondecreasing functions, the left hand side converges to $u(x)$ as $\epsilon \downarrow 0$, so that the relation (2.3.16) holds.

Finally, consider the situation where $x \in D^v_{\text{inf}}$. This means that $v(x) = -\infty$ for $x \in D^v_{\text{inf}}$. Since $D^u_{\text{inf}} \supset D^v_{\text{inf}}$, we also have $u(x) = -\infty$ for $x \in D^v_{\text{inf}}$. Definition (2.3.15) implies that $T(-\infty) = -\infty$, so that in this case the relation $u(x) = T(v(x))$ is satisfied as well.

It is clear that T is nondecreasing, because if $y_1 < y_2$, the set from Definition (2.3.15) for y_1 is a superset of the one for y_2, and in general for two sets $A \subset B \subset \mathbb{R}$ one has $\inf A \geq \inf B$. If a function is nondecreasing, upper semicontinuity is equivalent to right-continuity. But the right-continuity of T directly follows from Definition (2.3.15).

Next we show that T is strictly increasing on $v(D) \cap \{T \neq -\infty\}$. Let $y_1 < y_2$ in $v(D)$. If such two values do not exist, then there is nothing to prove. If $y_1 = -\infty$, we have $T(y_1) = -\infty$, and the statement is trivially satisfied. If $y_1 > -\infty$, then $y_1 \in [v(a), v(b)]$, where v is strictly increasing and continuous on $[a, b]$. In this case, $\tilde{T}(y_i) = u(v^{-1}(y_i)_-)$. But in this range, v^{-1} is strictly increasing and continuous as well, and $v^{-1}(y_1) < v^{-1}(y_2)$. Let $\tilde{T}(y_1) = \tilde{T}(y_2) > -\infty$. Then $u(v^{-1}(y_1)) = u(v^{-1}(y_2))$, and $v^{-1}(y_1)$ is an element of D^u_{sup} but not of D^v_{sup}, a contradiction to Lemma 2.3.22. It follows that $\tilde{T}(y_1) < \tilde{T}(y_2)$ for $y_1 < y_2$ unless $\tilde{T}(y_i) = -\infty$. If we choose $\epsilon > 0$ so small that $y_1 + \epsilon < y_2$, then

$$T(y_1) \leq \tilde{T}(y_1 + \epsilon) < \tilde{T}(y_2) \leq T(y_2).$$

It remains to show that T is concave. If not, then there exist $y_1, y_2 \in \mathbb{R}$ and $t \in]0, 1[$ with

$$tT(y_1) + (1 - t)T(y_2) > T(ty_1 + (1 - t)y_2). \tag{2.3.17}$$

Without loss of generality, we may assume that $y_1 < y_2$. If u is constant, it follows that T is constant, and we have a direct contradiction. If not, we conclude from the arguments above and a little extension that there exists an interval $[a, b]$ such that $T(y) = u(v^{-1}(y))$ for $y \in [v(a), v(b)]$, $T(y) = u(b)$ for $y > v(b)$, and $T(y) = -\infty$ for $y < v(a)$.

Now assume that $y_1 < v(a)$. It follows that $T(y_1) = -\infty$, which is a contradiction to (2.3.17). Assume now that both y_1 and y_2 are contained in $[v(a), v(b)]$. Then inequality (2.3.17) implies that

$$tu(v^{-1}(y_1)) + (1-t)u(v^{-1}(y_2)) > u(v^{-1}(ty_1 + (1-t)y_2)).$$

If $X := v^{-1}(y_1)$ with probability t, $v^{-1}(y_2)$ with probability $1-t$ (such a random variable exists because the probability space admits a continuous distribution), this equation means

$$Eu(X) > u(v^{-1}(Ev(X))),$$

where by the assumption that $y_1, y_2 \in [v(a), v(b)]$, $Ev(X) \in [v(a), v(b)]$ as well. Define $x^* := v^{-1}(Ev(X))$. It is obvious that $v(x^*) \geq Ev(X)$. But on the other hand, $Eu(X) > u(x^*)$. This is a contradiction to statement 3.

If $v(b) < y_1 < y_2$, we again have that $T(y_1) = T(y_2)$, and equation (2.3.17) gives a direct contradiction. It remains to consider the case where $v(a) \leq y_1 \leq v(b) < y_2$. But then we have

$$tT(y_1)+(1-t)T(v(b)) = tT(y_1)+(1-t)T(y_2) > T(ty_1+(1-t)y_2) \geq T(ty_1+(1-t)v(b))$$

that means inequality (2.3.17) is satisfied as well if we replace y_2 by $v(b)$. But this case we have already treated before. □

2.3.4 Connection to power utility functions

In this section, we will show that the connection to power utility functions, as established in Proposition 2.2.2 for smooth utility functions, continues to hold in the general case with our generalized definition of risk aversion. A similar connection also holds for the absolute risk aversion. Here we omit the proof and restrict the discussion to the case of relative risk aversion.

First we recall from Definition 2.3.14 what bounded risk aversion means.

Definition 2.3.24. Let u be a utility function with risk aversion measure ρ, and $\gamma \geq 0$ a constant. Then the absolute risk aversion is *bounded from above by* γ on an interval I if ρ satisfies

$$\rho \leq \gamma dx \quad \text{on} \quad I$$

where dx is the Lebesgue measure. It has a relative risk aversion *bounded from above by* γ on I if ρ satisfies

$$\rho \leq \operatorname{sgn}(x)\gamma d\ln(|x|) \quad \text{on} \quad I.$$

Boundedness from below is defined analogously with the reverse inequalities.

Remark 2.3.25. The measures γdx and $\operatorname{sgn}(x)\gamma d\ln(|x|)$ are risk aversion measures on the Borel sets of $\mathbb{R}$ and $]0,\infty[$ or $]-\infty,0[$, respectively, and γ is the corresponding Radon-Nikodym derivative. Definition 2.3.24 is therefore consistent with Definition 2.3.14. Furthermore, from Definition 2.3.24 it follows that I cannot contain elements of D^u_{sup}, because the utility function v corresponding to the risk aversion measure on the right hand side of the inequality is an exponential or power utility function, and therefore $D^v_{\text{sup}} = \emptyset$. But according to Definition 2.3.14, the risk aversion domains must coincide on I, that is $D^u_{\text{ra}} \cap I = D^v_{\text{ra}} \cap I$.

Proposition 2.3.26. *The relative risk aversion of a utility function u satisfying Assumptions 2.3.1 and 2.3.2 is uniformly bounded from below (above) by a constant $\gamma > 0$ in a region $R :=]a,b[\subset D$ if and only if $R \subset D^u_{\text{ra}}$ and for every $a < x \leq y < b$ we have the inequality*

$$u'_r(y) \leq \begin{array}{ll} u'_r(x)\left(\frac{y}{x}\right)^{-\gamma} & \text{if} \quad D =]0,\infty[\\ u'_r(x)\left(\frac{y}{x}\right)^{\gamma} & \text{if} \quad D =]-\infty,0[\end{array} \tag{2.3.18}$$

or the inequality in the other direction if bounded from above.

Proof. By Definition 2.3.24, it is clear that $R \subset D^u_{\text{ra}}$. First let x be in D^u_{eff}, and $x < y < b$, then $y \in D^u_{\text{eff}}$. By assumption (Definition 2.3.24), and because $]x,y] \subset D_{\text{eff}}$, we have

$$\rho(]x,y]) = -d\ln u'_r(]x,y]) \geq \operatorname{sgn}(\xi)\gamma d\ln|\xi|(]x,y])$$

on $]x,y]$. If $D =]0,\infty[$, then by integration

$$\ln u'_r(y) - \ln u'_r(x) \leq -\gamma\left(\ln y - \ln x\right),$$

and by the rules of the logarithm and the monotonicity of the exponential function

$$\frac{u'_r(y)}{u'_r(x)} \leq \left(\frac{y}{x}\right)^{-\gamma},$$

which is equation (2.3.18) for positive wealth. For $D =]-\infty,0[$ the arguments are similar.

If $x \in (D^u_{\text{eff}})^c$, then $u'_r = \infty$ from which the equation follows, or $\xi \in D^u_{\text{eff}}$ for all $\xi > x$. In the latter case, inequality (2.3.18) also holds for all $y > \xi > x$ by the arguments before, and by the right-continuity of u'_r the inequality then also holds for x.

Let now equation (2.3.18) hold and $R \subset D^u_{\mathrm{ra}}$. Consider the interval $]x, y] \subset R$. We first consider the case where $x, y \in D^u_{\mathrm{eff}}$. On D^u_{eff} the measures $-d \ln u'_r$ and $\mathrm{sgn}(x) d \ln |x|$ are sigma-finite. By taking the logarithm which is monotonic we have

$$\ln u'_r(y) - \ln u'_r(x) \leq \mathrm{sgn}(x)\gamma (\ln y - \ln x)$$

on every half-open interval in $D_{\mathrm{eff}} \cap]a, b[$. The result follows by the following Lemma 2.3.27 for sigma-finite measures.

If $x \in (D^u_{\mathrm{eff}})^c$, then $\rho(\{x\}) = \infty$ but for any constant $\gamma > 0$, $\mathrm{sgn}(y)\gamma d \ln |y|(\{x\}) = 0$, and therefore the relative risk aversion must be bounded from below by γ. □

Lemma 2.3.27. *Let μ_1 and μ_2 two sigma-finite measures on the Borel sets with $\mu_1(I) \leq \mu_2(I)$ for every half-open interval I. Then $\mu_1(B) \leq \mu_2(B)$ for every Borel set B.*

Proof. Let μ_1 and μ_2 be two sigma-finite measures on the Borel sets with $\mu_1(I) \leq \mu_2(I)$ for all half-open intervals I. Sigma-finiteness of the measures μ_j on the Borel subsets of a set D_{eff} means that there are disjoint subsets A_i with $\cup_{i \geq 1} A_i = D_{\mathrm{eff}}$ and $\mu_j(A_i) < \infty$ for all i, for both $j = 1$ and $j = 2$. If there would be a Borel set B on which $\mu_1(B) > \mu_2(B)$, there would at least be one set A_i with $\mu_1(A_i \cap B) > \mu_2(A_i \cap B)$, and on A_i the measures are finite. We may therefore assume that both measures are finite. We then define a signed measure $\lambda := \mu_2 - \mu_1$. This measure is obviously countably additive, positive on all half-open intervals, and negative on B. Because of the additivity of λ, we have that $\lambda \geq 0$ for all finite unions of half-open intervals, which form an algebra. Because this algebra is a subset of the Borel sets, λ is also countably additive on this algebra. By Carathéodory's extension theorem, see [83], one can therefore extend λ (defined on this algebra, where it is positive) to a positive measure $\tilde{\lambda}$, defined on all Borel sets. It follows that $\tilde{\lambda}(B) \geq 0$. The two measures λ and $\tilde{\lambda}$ coincide on a π-system generating the Borel sets, and by the uniqueness lemma, see [83], which also holds for signed measures, it follows that $\lambda = \tilde{\lambda}$ on all Borel sets, therefore $\lambda(B) \geq 0$, a contradiction. □

In the case of constant relative risk aversion, inequality (2.3.18) becomes an equality, and by integrating both sides first with respect to x, then with respect to y, we obtain (we treat here the case $D =]0, \infty[$)

$$u(y)\left(x^{1-\gamma} - y_0^{1-\gamma}\right) - u(y_0)x^{1-\gamma} = u(x)\left(y^{1-\gamma} - y_0^{1-\gamma}\right) - u(y_0)y^{1-\gamma}$$

where y_0 is the lower integration bound. Treating y_0 and x as two different constants in the interval $[a, b]$ where the relative risk aversion is γ, it follows for suitable constants A and B that

$$u(y) = A + By^{1-\gamma}.$$

Hence constant relative risk aversion precisely means the power property, a well-known result. When the relative risk aversion is only bounded but not constant, inequality (2.3.18) gives a super-power property of the utility function.

2.3.5 Relative risk aversion of the conjugate function

The connection between the relative risk aversion of a utility function and its dual stated in Proposition 2.2.2 for smooth utility functions is not so easy any more for the general case. The reason is that u'_r may not be invertible any more, and as a consequence, if we have $rra_{u^*}(y) = \frac{1}{rra_u(x)}$, it is not clear which x one has to take for a specific y. However, it is still possible to make a statement about the bounds of the relative risk aversion. This is the aim of this section.

Definition 2.3.28. Let $u(x)$ be a utility function satisfying Assumptions 2.3.1, 2.3.2 and 2.3.3. We extend the domain in the following way:

$$u_{ext}(x) := \begin{array}{lll} -\infty & \text{on} \quad]-\infty, 0[& \text{if} \quad D =]0, \infty[\\ u(-x) & \text{on} \quad]0, \infty[& \text{if} \quad D =]-\infty, 0[\\ u(x) & \text{on} \quad D & \end{array} \tag{2.3.19}$$

and at 0 such that the function becomes upper semicontinuous. Then its *dual function* is defined as the concave conjugate

$$u^*(y) := \inf_{x \in \mathbb{R}} (xy - u(x)). \tag{2.3.20}$$

Lemma 2.3.29. *Let $u(x)$ satisfy Assumptions 2.3.1, 2.3.2 and 2.3.3, with domain D. Then its dual function $u^*(y)$ also satisfies these assumptions on D. Furthermore, the extension of u^* according to (2.3.19) is $(u^*)_{ext} = (u_{ext})^*$.*

Proof. From Rockafellar [70], it follows that $u^*(y)$ defined on $\mathbb{R}$ is a concave, proper and upper semicontinuous function. (This theorem is proved in [70] for convex instead of concave functions. Note that closedness is equivalent to upper semicontinuity for concave functions). It remains to show that $u^*(y)$ is nondecreasing on D. Let $y_1 < y_2 \in D$. Then for all $x > 0$, it follows that

$$xy_1 - u_{ext}(x) \leq xy_2 - u_{ext}(x).$$

This estimate then also holds for the infimum over all positive x. The result follows by recognizing that for $y \in D$,

$$u^*(y) = \inf_{x \in \mathbb{R}} (xy - u_{ext}(x)) = \inf_{x > 0} (xy - u_{ext}(x)).$$

The final statement is clear for $y \in D$. For $y < 0$ and $D =]0, \infty[$, the expression on the right hand side of (2.3.20) can be made arbitrarily small when $x \to \infty$. For $y > 0$ and $D =]-\infty, 0[$, we have

$$\begin{aligned}(u_{ext})^*(y) &= \inf_{x\in\mathbb{R}} (xy - u_{ext}(x)) = \inf_{x\in\mathbb{R}} ((-x)(-y) - u_{ext}(-x)) \\ &= \inf_{-x\in\mathbb{R}} ((-x)(-y) - u_{ext}(-x)) = u^*(-y),\end{aligned}$$

which is by definition $(u^*)_{ext}$. □

Proposition 2.3.30. *Let u be a utility function satisfying Assumptions 2.3.1, 2.3.2 and 2.3.3. Then its dual function $u^*(y)$ has a relative risk aversion bounded from above (below) by $\frac{1}{\gamma}$ on $]a, b[$ if $u(x)$ has a relative risk aversion bounded from below (above) by γ on $](u^*)'_r(b), (u^*)'_r(a)[$ ($[(u^*)'_r(b), (u^*)'_r(a)]$) if $D =]0, \infty[$, respectively on*
$]-(u^*)'_r(a), -(u^*)'_r(b)[$ ($[-(u^*)'_r(a), -(u^*)'_r(b)]$) if $D =]-\infty, 0[$, provided this set is nonempty.

Proof. Let $a < x < y < b$ and assume first that $x, y \in D^{u^*}_{\text{eff}}$. We have

$$\int_{]x,y]} -d\ln(u^*)'_r = -\ln(u^*)'_r(y) + \ln(u^*)'_r(x)$$

and

$$\int_{](u^*)'_r(y),(u^*)'_r(x)]} d\ln|\xi| = \ln(u^*)'_r(x) - \ln(u^*)'_r(y)$$

for $D =]0, \infty[$. For $D =]-\infty, 0[$, we have

$$\int_{]-(u^*)'_r(x),-(u^*)'_r(y)]} -d\ln|\xi| = \ln(u^*)'_r(x) - \ln(u^*)'_r(y),$$

and therefore the integrals are the same. Because on D the measure $d\ln|x|$ is absolutely continuous with respect to the Lebesgue measure, we may exclude the point at the right end of the integration interval without changing the value of the integral. Let first $D =]0, \infty[$. Then, by assumption, if $](u^*)'_r(y), (u^*)'_r(x)[\subset D^u_{\text{eff}}$,

$$-\int_{]x,y]} d\ln(u^*)'_r = \int_{](u^*)'_r(y),(u^*)'_r(x)[} d\ln x \leq \frac{1}{\gamma}\int_{](u^*)'_r(y),(u^*)'_r(x)[} -d\ln u'_r,$$

and the right hand side is then equal to

$$-\frac{1}{\gamma}\left(\ln u'_r((u^*)'_r(x)_-) - \ln u'_r((u^*)'_r(y))\right) \leq \frac{1}{\gamma}(\ln y - \ln x) = \frac{1}{\gamma}\int_{]x,y]} d\ln|\xi|$$

which is the required result. The first inequality follows from the fact that

$$(u^*)'_r(x) \in \delta u^*(x) \Rightarrow x \in \delta u((u^*)'_r(x))$$

by the general duality rules of superdifferentials, and $u'_r((u^*)'_r(x)_-)$ is the supremum of those superdifferentials, and therefore larger than x. This holds by the general rule that $u'_r(z_-) = \sup\{\delta u(z)\}$. On the other hand, by the same argument, we have that $y \in \delta u((u^*)'_r(y))$ and therefore larger than or equal than the infimum of the superdifferential, which is $u'_r((u^*)'_r(y))$. If $](u^*)'_r(y), (u^*)'_r(x)[\not\subset D^u_{\text{eff}}$, then $u'_r((u^*)'_r(y)) = \infty$, and by the same rules as before, $y = \infty$, a contradiction to the assumption that $y \in D$. The other case, where $D =]-\infty, 0[$, goes essentially the same way, where one has to apply u_{ext} and u^*_{ext}, and the fact that by the symmetry from Definition 2.3.28, $y \in \delta u(z) \Rightarrow -y \in \delta u(-z)$.

We will show now that $x, y \in D^{u^*}_{\text{eff}}$ is always true if u has a relative risk aversion bounded from below, and the other assumptions of Proposition 2.3.30 are satisfied. We have in general $D_{\text{eff}} =]d_{min}, d_{max}[$. Let us first consider the case when $D =]-\infty, 0[$. If $d_{max} < b \leq 0$, we have that $0 \in \delta u^*(d_{max})$ and thus $\pm d_{max} \in \delta u(0)$. It follows that $u(x) \leq u(0) - d_{max}x$, and because u'_r is monotonically decreasing, we have $u'_r \geq -d_{max}$. For $\epsilon > 0$, it follows that

$$\int_{]\epsilon,0[} -d \ln u'_r \leq \frac{1}{-d_{max}} \int_{]\epsilon,0[} -du'_r = \frac{1}{-d_{max}} \left(u'_r(-\epsilon) - u'_r(0_-)\right)$$

which must be bounded if $\epsilon > 0$ is small enough due to the fact that u is proper. On the other hand, $d \ln x(]-\epsilon, 0[) = \infty$ for every $\epsilon > 0$. It follows that the relative risk aversion of u cannot be bounded from below.

On the other side, let $0 > d_{min} > a$. It follows that $(u^*)'_r(a) = \infty$, and therefore $a \in \delta u(\infty)$, or by symmetry $-a \in \delta u(-\infty)$. It follows that $u'_r(x) \leq -a$ for all $x \in D$. By the fact that u'_r is nonincreasing, it follows that $u'_r(x)$ converges as $x \to -\infty$. For all $\epsilon > 0$, there exists an $x \in]-\infty, -(u^*)'_r(b)[$ with

$$\int_{x-1}^{x} -d \ln u'_r \leq \frac{1}{u'_r(c)} \int_{x-1}^{x} -du'_r \leq \frac{u'_r(x-1) - u'_r(x)}{u'_r(c)} < \frac{\epsilon}{u'_r(c)}$$

for all $x < c$, with c such that $x < c < -(u^*)'_r(b)$ and $u'_r(c) > 0$. The choice of such a c is possible because otherwise D^u_{ra} is empty, and then the relative risk aversion could not be bounded anywhere. On the other hand, the measure $-d \ln |\xi|(]x-1, x])$ as a function of x is bounded from below for $x < c$. Again, it follows that the relative risk aversion of u cannot be bounded from below.

The case $D =]0, \infty[$ follows similar arguments, but for completeness we will give them too. Let us first assume that $a < d_{min}$. Then we have $(u^*)'_r(a) = \infty$ and therefore $a \in \delta u(\infty)$. It follows that $](u^*)'_r(b), (u^*)'_r(a_-)[$ is an interval of the form $]c, \infty[$ with $c < \infty$, because otherwise the interval would be empty. By the fact that u'_r is nonincreasing, we must have $u'_r(x) \geq a$ for all $x > 0$, and by the monotonicity and the boundedness $u'_r(x)$ converges as $x \to \infty$. This means that for all $\epsilon > 0$ there must be an $x > c$ with $|u'_r(x) - u'_r(x+1)| < \epsilon$. It follows that

$$-\int_x^{x+1} d\ln u'_r \leq \frac{1}{-a}\int_x^{x+1} du'_r < \frac{\epsilon}{a}.$$

On the other hand, $d\ln|\xi|(]x,x+1]) \geq d\ln|\xi|(]c,c+1]) > 0$, and $d\ln|\xi|(]x,x+1])$ is therefore bounded from below. It follows that the relative risk aversion of u cannot be bounded from below.

For the last case, let us assume that $d_{max} < b$. Then $(u^*)'_r(b) = 0$ and $b \in \delta u(0)$. The interval $](u^*)'_r(b), (u^*)'_r(a_-)[$ is of the form $]0,c[$ with $c > 0$, because otherwise it would be empty. It follows by the rules of superdifferentials that $u(x) - u(0) \leq bx$ and by the monotonicity $u'_r(x) \leq b \ \forall x$. This means that $u'_r(x)$ is monotonically decreasing and bounded from above, and therefore

$$-\int_{\epsilon_1}^{\epsilon_2} d\ln u'_r \leq -\frac{1}{u'_r(c_1)}\int_{\epsilon_1}^{\epsilon_2} du'_r \leq \frac{u'_r(\epsilon_1) - u'_r(\epsilon_2)}{u'_r(c_1)},$$

where $0 < c_1 < c$ is a constant with $u'_r(c_1) > 0$. Such a constant exists if u is not constant, which is impossible because then $D^u_{\text{ra}} = \emptyset$. The right hand side of the above inequality tends to 0 for all $0 < \epsilon_1 < \epsilon_2 < c_1$ as $c_1 \to 0$, but the measure $d\ln|x|(]\epsilon_1,\epsilon_2[)$ remains bounded from below for suitable sequences of ϵ_1, ϵ_2. Again, it follows that the relative risk aversion of u cannot be bounded from below.

We therefore conclude that the boundedness from above holds on any half-open interval, and because $]a,b[\subset D^u_{\text{eff}}$, the measures $d\ln x$ and $d\ln(u^*)'_r$ are sigma-finite on $]a,b[$. The result now follows by Lemma 2.3.27. □

2.4 Essential bounds for the risk aversion

2.4.1 Definition

Definitions 2.3.14 and 2.3.24 would imply that there cannot be jumps in u'_r if the risk aversion is bounded from above. It is clear that the (absolute or relative) risk aversion is not bounded from above at the jumps. On the other hand, Definitions 2.3.7, 2.3.14 and 2.3.24 are too strict. They would still imply that the risk aversion cannot be bounded by a constant either from above or from below for all piecewise linear utility functions. Furthermore, for having "essentially" a super-power property in the sense that a slight modification of inequality (2.3.18) holds, it is only necessary that the relative risk aversion is bounded from below (or bounded from above) up to a certain tolerance.

A first idea for a weaker ordering of risk aversions would consist of requiring that a utility function u is essentially not more risk averse than v on $]a,b[\subset D$ if

$]a, b[\cap D^u_{\text{ra}} =]a, b[\cap D^v_{\text{ra}}$ and there exists a constant $C < \infty$ such that

$$\rho_u(B') \leq \rho_v(B') + C$$

for all $B' \subset \mathcal{B}]a, b[$.

For again obtaining an if and only if statement analogous to Proposition 2.3.26 (super-power or super-exponential property, respectively), as well as for being able to have essentially bounded risk aversion also for piecewise linear functions, we again define it slightly more generally.

Definition 2.4.1. [Weak comparison of risk aversions] A utility function $u(x)$ is *essentially not more risk averse* than a utility function $v(x)$ on an interval $]a, b[\subset D$ if $]a, b[\cap D^u_{\text{ra}} =]a, b[\cap D^v_{\text{ra}}$ and there exists a constant $0 < C < \infty$ such that

$$\rho_u([x, y]) \leq \rho_v([x, y]) + C \tag{2.4.21}$$

for all intervals $[x, y] \subset]a, b[$. Analogously to the open interval $]a, b[$, weak comparison for risk aversions is also defined on closed or half-open intervals.

Remark 2.4.2. From Definition 2.4.1, it follows that inequality (2.4.21) holds for all (open, closed, half-open) intervals, by the fact that one can create those intervals as a countable union of closed intervals, and by the monotone convergence of measures, see [83].

Remark 2.4.3. In order to show that u is essentially not more risk averse than v, one can also show inequality (2.4.21) for half-open intervals $]x, y]$ and, in the case of the closed interval $[a, b]$, additionally for the sets $\{a\}$ and $\{b\}$. The reason is that a closed interval can be written as a countable intersection of half-open intervals. The result then follows by the monotone convergence of measures, see [83], and the fact that each closed interval in D^u_{eff} with finite endpoints is compact.

Indeed, given an open interval $]a, b[$ and $x, y \in D^u_{\text{eff}}$ with $]x, y] \subset]a, b[$, it follows that $[x, y] \subset D^u_{\text{eff}}$. First assume that $[x, y] \subset D^v_{\text{eff}}$, then there exists an N such that $]x - \frac{1}{n}, y] \subset]a, b[$ for all $n \geq N$ and furthermore $]x - \frac{1}{n}, y] \subset D^u_{\text{eff}} \cap D^v_{\text{eff}}$. The monotone convergence of measures yields that the constant C for closed sets is bounded by the one for half-open sets. If $[x, y]$ would not be a subset of D^v_{eff}, by the fact that $\rho_u([x, y]) < \infty$, inequality (2.4.21) would trivially be satisfied.

If x or y would not be contained in D^u_{eff}, then either $\rho_u(]x, y]) = \infty$ or $\rho_u(\{x\}) = \infty$, by the ρ-finiteness of the measure ρ_u (Definition 2.3.11). It follows that in the first case, $\rho_v(]x, y]) = \infty$, and the inequality (2.4.21) still holds. In the second case, if x would be in D^v_{eff}, there must be a $\xi < x$ with $[\xi, x] \subset D^v_{\text{eff}}$, and by the ρ-finiteness of the measure ρ_v, it follows that $\rho_v(]\xi, x]) < \infty$, a contradiction to $\rho_u(]\xi, x]) = \infty$

and the fact that (2.4.21) holds for half-open intervals. It follows that $\rho_v(\{x\}) = \infty$, and therefore inequality (2.4.21) still holds.

Given a closed interval $[a, b]$, the constant C in (2.4.21) for intervals $[x, y]$ must be smaller than or equal to the sum of the one for intervals $]a, b[$ and the ones for the sets $\{a\}$ and $\{b\}$.

Analogous to Definition 2.3.24, we also state what we mean by essential bounds for the risk aversion.

Definition 2.4.4. [Essential bounds for risk aversion] A utility function $u(x)$ satisfying Assumptions 2.3.1 and 2.3.2 *has a relative risk aversion essentially bounded from above by* γ, $0 < \gamma < \infty$, on an interval $]a, b[\subset D$, if $]a, b[\subset D^u_{\mathrm{ra}}$ and there exists a constant $0 < C < \infty$ such that

$$\int_{]x,y]} d\rho_u \leq \int_{]x,y]} \operatorname{sgn}(x)\gamma d \ln |x| + C \tag{2.4.22}$$

holds for all intervals $]x, y] \subset]a, b[$. It has *a relative risk aversion essentially bounded from below by* γ if $]a, b[\subset D^u_{\mathrm{ra}}$ and there exists a constant $0 < C < \infty$ such that

$$\int_{]x,y]} d\rho_u \geq \int_{]x,y]} \operatorname{sgn}(x)\gamma d \ln |x| - C \tag{2.4.23}$$

holds for all intervals $]x, y] \subset]a, b[$. It has *essentially a relative risk aversion of* γ if both (2.4.22) and (2.4.23) are valid. Analogously, essential bounds for the absolute risk aversion are defined, with the measure $\operatorname{sgn}(x)\gamma d \ln |x|$ replaced by the measure dx.

Remark 2.4.5. We have defined the essential bounds here by the use of half-open intervals, because this will be more convenient for the proofs later. This is possible by the Remarks 2.4.2 and 2.4.3. If we look at the interval $[a, b]$ instead of the open one, we would have to add, as in Remark 2.4.3, the requirement that $\rho_u(\{a\}) < \infty$ in the case of boundedness from above.

Example 2.4.6. For $n \geq 0$ define

$$d\rho_u = -d \ln u'_r(x) = \begin{array}{ll} \ln 2 & \text{if} \quad x = \pm\frac{3}{4}2^{-n} \\ 0 & \quad\text{otherwise} \end{array}$$

where the plus and minus signs depend on the domain. Then we have that $\int_B (d\rho_u - \operatorname{sgn}(x)\gamma d \ln |x|)$ is unbounded from above as well as from below with respect to all $B \in \mathcal{B}(D)$. But looking only at intervals, one can see that the function $u(x)$ has essentially a relative risk aversion of 1 in the sense of Definition 2.4.4. Looking only at intervals, the positive and negative parts of the 'measure' $-d \ln u'_r - \operatorname{sgn}(x)\gamma d \ln |x|$

cancel out to a uniformly bounded number, even if this (signed) 'measure' maps some Borel sets to $+\infty$ and some others to $-\infty$ and therefore does not define a signed measure in the strict sense.

Remark 2.4.7. A sufficient condition for u having a risk aversion essentially bounded from above is that there exists a Borel set $C \subset]a, b[$ with

$$\int_C d\rho_u < \infty \tag{2.4.24}$$

such that

$$\int_{B'} d\rho_u \leq \int_{B'} \operatorname{sgn}(x)\gamma d \ln |x| \tag{2.4.25}$$

(relative risk aversion case), respectively

$$\int_{B'} d\rho_u \leq \int_{B'} \gamma dx \tag{2.4.26}$$

(absolute risk aversion case) for every Borel set $B' \subset (]a, b[\backslash C)$.

Example 2.4.8. If u'_r has a finite amount of jumps in $a < x_1 < ... < x_n < b$ it satisfies assumption (2.4.24) if $u'_r(x_n) > 0$ and $u'_r(x_1^-) < \infty$. By the transformation of variables formula for finite variation processes equation (2.4.24) then gives for $C = \{x_1, ..., x_n\}$

$$\int_C d\rho_u = \int_C -d \ln u'_r(x) = \sum_i \ln \left(\frac{u'_r(x_{i-})}{u'_r(x_i)} \right) < \infty.$$

Example 2.4.9. The function $u(x) := -|x|^{\frac{1}{-x}}$ is concave in a region around 0, for both cases of positive and negative wealth. But the relative risk aversion according to Definition 2.4.4 is not essentially bounded from above on the intervals $]-\epsilon, 0[$ or $]0, \epsilon[$ for any $\epsilon > 0$.

Remark 2.4.10. A sufficient condition for u having a risk aversion essentially bounded from below is that there exists a Borel set $C \subset]a, b[$ such that

$$\int_{B'} d\rho_u \geq \operatorname{sgn}(x)\gamma \int_{B'} d \ln |x| \tag{2.4.27}$$

(relative risk aversion case), respectively

$$\int_{B'} d\rho_u \geq \gamma \int_{B'} dx \tag{2.4.28}$$

(absolute risk aversion case) for every Borel set $B' \subset (]a, b[\backslash C)$, where

$$\int_C \operatorname{sgn}(x) d \ln |x| < \infty \tag{2.4.29}$$

for the relative risk aversion case, and $\mu(C) < \infty$ for the absolute risk aversion case, where μ is the Lebesgue measure.

In particular, from the Remarks 2.4.7 and 2.4.10, it follows that the risk aversion is always essentially bounded if it is bounded.

Example 2.4.11. Condition (2.4.29) is satisfied for a finite union of closed intervals in D_{eff}: $C = [x_1, y_1] \cup ... \cup [x_n, y_n]$ with $a < x_1 < y_1 < x_2 < y_2 < ... < y_n < b$, and $D_{\text{eff}} =]a, b[$.

Example 2.4.12. The functions $u(x) := -x \ln x$ for $D =]0, \infty[$, and $v(x) := \frac{-x}{\ln |x|}$ for $D =]-\infty, 0[$, are concave if $|x|$ is sufficiently small. The relative risk aversion is not essentially bounded from below by a constant $\gamma > 0$ on the intervals $]0, \epsilon[$ (for positive wealth) or $]-\epsilon, 0[$ (for negative wealth) for any $\epsilon > 0$. The functions u and v are asymptotically linear.

2.4.2 Connection to power utility

We now reformulate Proposition 2.3.26 for the case of essentially bounded relative risk aversion. For the essentially bounded absolute risk aversion, a similar statement holds. We will omit the proof of the latter case and focus on the former one.

Proposition 2.4.13. *Let u be a utility function satisfying Assumptions 2.3.1 and 2.3.2. Then the relative risk aversion of u is essentially bounded from above (below) by a constant $0 < \gamma < \infty$ for $x \in]a, b[\subset D$, if and only if $]a, b[\subset D^u_{\text{eff}}$ and there exists a constant $K > 0$ such that for all $0 < x < y < b$, we have the inequality*

$$u'_r(y) \geq \begin{array}{ll} K u'_r(x) \left(\frac{y}{x}\right)^{-\gamma} & \text{if} \quad D =]0, \infty[, \\ K u'_r(x) \left(\frac{y}{x}\right)^{\gamma} & \text{if} \quad D =]-\infty, 0[, \end{array} \tag{2.4.30}$$

or the same inequality in the other direction in the case of boundedness from below.

Proof. Let $u(x)$ satisfy equation (2.4.22) of Definition 2.4.4 on $]a, b[$. Let $a < x < y < b$. Because $]x, y] \subset]a, b[$ is an interval, we have

$$\int_{]x,y]} d\rho_u \leq \gamma \operatorname{sgn}(x) \int_{]x,y]} d \ln |x| + C,$$

where C is the constant from equation (2.4.22). It follows by the rules of the logarithm that if $x, y \in D^u_{\text{eff}}$,

$$-\ln u'_r(y) \leq -\ln u'_r(x) \pm \gamma \ln \left(\frac{y}{x}\right) + C,$$

where the plus sign refers to the case $D =]0, \infty[$ and the minus sign to the other one. By the monotonicity of the exponential function, equation (2.4.30) follows with the constant $K = \exp(-C)$. If there would be an $x \in]a, b[$ with $x \in (D^u_{\text{eff}})^c$, then, by the fact that $]a, b[$ is open, there exists a half-open interval $]\xi_1, \xi_2] \subset]a, b[$ with $\rho_u(\{\xi\}) = \infty$ for all $\xi \in]\xi_1, \xi_2]$. Because $\operatorname{sgn}(x) d \ln |x|(\{]\xi_1, \xi_2]\})$ is finite, this contradicts definition (2.4.22) of the essential upper bound. It follows that $]a, b[\subset D^u_{\text{eff}}$.

On the other hand, let there be a constant $K > 0$ such that for all $a < x < y < b$, equation (2.4.30) is satisfied. Because $]a, b[\subset D^u_{\text{eff}}$, we can always take the logarithm. Then, by doing this, we have

$$\ln u_r'(y) - \ln u_r'(x) \geq \mp\gamma(\ln y - \ln x) + \ln K.$$

It follows that

$$-\ln K \geq \int_{]x,y]} -d \ln u_r' - \gamma \int_{]x,y]} \operatorname{sgn}(x) d \ln |x| = \int_{]x,y]} d\rho_u - \gamma \int_{]x,y]} \operatorname{sgn}(x) d \ln |x|.$$

Therefore, equation (2.4.22) is satisfied, with the constant $C = -\ln K$. We have only proved it for intervals $]x, y]$ with $a < x$, but the bound for intervals of the form $]a, y]$ is the same, by the monotone convergence of measures [83].

The statement for the case with boundedness from below is proved in the same way. □

As for smooth utility functions (Corollary 2.2.4), one can see how Proposition 2.4.13 connects essential bounds of the relative risk aversion to an essential superpower property.

Corollary 2.4.14. *Let u be a utility function satisfying Assumptions 2.3.1 and 2.3.2 with relative risk aversion essentially bounded from above by a constant γ on $]a, b[$.*

1. *Assume that $D =]0, \infty[$, $\gamma < 1$ and $a = 0$, and $u(0^+) = 0$. Then for all $0 < x < y < b$ we have the inequality*

$$u(y) \geq K u(x) \left(\frac{y}{x}\right)^{1-\gamma}. \tag{2.4.31}$$

2. *Assume that $D =]0, \infty[$, $\gamma > 1$ and $b = \infty$, and $u(\infty^-) = 0$. Then for all $a < x < y < \infty$ we have the inequality*

$$-u(y) \geq -u(x) K \left(\frac{y}{x}\right)^{1-\gamma}. \tag{2.4.32}$$

3. *Assume that $D =]-\infty, 0[$ and $b = 0$, and $u(0^-) = 0$. Then for all $a < x < y < 0$ we have the inequality*

$$-u(y) \geq -K u(x) \left(\frac{y}{x}\right)^{\gamma+1}. \tag{2.4.33}$$

Proof. Because $u'_r(x)$ is nonincreasing and $]a,b[\in D^u_{\text{eff}}$, it is continuous with the exception of at most countably many points, which are a Lebesgue null set. It follows that almost surely (with respect to the Lebesgue measure), the function u is differentiable and $u' = u'_r$. One can therefore apply the fundamental theorem of calculus even if u'_r is not continuous, yielding

$$\int_{x_0}^{x} u'_r(\xi)d\xi = u(x) - u(x_0), \tag{2.4.34}$$

see e.g. Berberian [8], Theorem 5.10.1.

Applying this fundamental theorem, the proofs are almost the same as the ones for smooth utility functions (Corollary 2.2.4). □

Remark 2.4.15. For $\gamma < 1$, we have already discussed in section 2.2 under which conditions a normalization of u to $u(0) = 0$ is possible. For $\gamma > 1$, a normalization is possible if u is bounded from above, which is the case of the relative risk aversion being essentially bounded from below. Indeed, by equation (2.4.30), and with an $x_0 \in D^u_{\text{eff}}$, it follows that $u'_r(y) \le Cy^{-\gamma}$ for a constant $C > 0$, and by the fundamental theorem of calculus, see [8], this means $u(y) \le u(x_0) + \frac{C}{\gamma-1}\left(x_0^{1-\gamma} - y^{1-\gamma}\right)$ for $y > x_0$.

In the case $D =]-\infty, 0[$, a normalization to $u(0) = 0$ is always possible. This follows directly from the concavity of u.

If the utility function is invertible, one can also deduce some relations for the inverse function from Corollary 2.4.14.

Proposition 2.4.16. *Let $]a,b[\subset D^u_{\text{eff}}$. Then the function $u :]a,b[\to]u(a),u(b)[$ is invertible. Furthermore, let the assumptions of Corollary 2.4.14 be satisfied, in particular the inequalities (2.4.31)-(2.4.33) under the corresponding assumptions. Then for all $\xi < \eta$ with $\xi,\eta \in]u(a),u(b)[$, the inverse function $u^{-1} :]u(a),u(b)[\to]a,b[$ satisfies the following inequalities:*

1. *In the case (2.4.31)*
$$\frac{u^{-1}(\eta)}{u^{-1}(\xi)} \le \frac{1}{K'}\left(\frac{\eta}{\xi}\right)^{\frac{1}{1-\gamma}}. \tag{2.4.35}$$
2. *In the case (2.4.32)*
$$\frac{u^{-1}(\eta)}{u^{-1}(\xi)} \ge \frac{1}{K'}\left(\frac{\eta}{\xi}\right)^{\frac{1}{1-\gamma}}. \tag{2.4.36}$$
3. *In the case (2.4.33)*
$$\frac{u^{-1}(\eta)}{u^{-1}(\xi)} \le \frac{1}{K'}\left(\frac{\eta}{\xi}\right)^{\frac{1}{\gamma+1}}. \tag{2.4.37}$$

Proof. Because $]a,b[\subset D^u_{\text{eff}}$, the function u is continuous on $]a,b[$, and therefore surjective. Furthermore, $u'_r > 0$, from which it follows that u must be injective. The inequalities (2.4.35)-(2.4.37) then follow from equations (2.4.31)-(2.4.33) by setting $x = u^{-1}(\xi)$ and $y = u^{-1}(\eta)$. By the assumptions about the normalization of u in Corollary 2.4.14, $\xi, \eta > 0$ in case 1 and $\xi, \eta < 0$ in cases 2 and 3. □

Remark 2.4.17. In case 3 in Proposition 2.4.16, but with risk loving investors and positive capital, we have from $\tilde{u}(x) = -u(-x)$ that $\tilde{u}^{-1}(\xi) = -u^{-1}(-\xi)$ is the inverse.

2.4.3 Essential bounds for the conjugate function

Proposition 2.4.18. *The relative risk aversion of a utility function u satisfying Assumptions 2.3.1, 2.3.2, and 2.3.3 is essentially bounded from below by a constant γ on a set $]a,b[$ if its conjugate function $u^*(y)$ has a relative risk aversion essentially bounded from above by the constant $\frac{1}{\gamma}$ on the set $[u'_r(b), u'_r(a)]$ or the set $[-u'_r(a), -u'_r(b)]$ for $D =]0,\infty[$ or $D =]-\infty, 0[$, respectively. Furthermore, the statement also holds if we exchange the words 'above' and 'below', then even with the intervals $]u'_r(b), u'_r(a)[$ and $]-u'_r(a), -u'_r(b)[$, respectively, provided these sets are nonempty.*

Proof. Let $a < x < y < b$ be given, and first let $x, y \in D^u_{\text{eff}}$. We define the interval

$$\operatorname{sgn}(x)]u'_r(y), u'_r(x)] := \begin{matrix}]u'_r(y), u'_r(x)] & \text{if} \quad D =]0,\infty[\\ ]-u'_r(x), -u'_r(y)] & \text{if} \quad D =]-\infty, 0[\end{matrix}$$

and analogous for closed or open intervals. We have

$$\int_{]x,y]} -d\ln u'_r = \ln u'_r(x) - \ln u'_r(y) = \int_{\operatorname{sgn}(x)]u'_r(y), u'_r(x)]} \operatorname{sgn}(x) d\ln|x|,$$

and on the right hand side we can take the closed interval instead, because the difference is only a Lebesgue null set.

By the fact that u^* has a relative risk aversion essentially bounded from above by $\frac{1}{\gamma} > 0$, we have in particular that the interval considered is a subset of $D^{u^*}_{\text{eff}}$, and therefore $\rho_{u^*} = -d\ln(u^*)'_r$, and

$$\int_{\operatorname{sgn}(x)[u'_r(y), u'_r(x)]} \operatorname{sgn}(x) d\ln|x| + \gamma K \geq \gamma \left(\int_{\operatorname{sgn}(x)[u'_r(y), u'_r(x)]} -d\ln(u^*)'_r \right),$$

where K is the supremum in equation (2.4.22), and therefore independent of the choice of x and y. A similar argument as in the proof of Proposition 2.3.30 gives that

$$\begin{aligned}\gamma K + \int_{]x,y]} -d \ln u_r' &\geq \gamma \int_{\operatorname{sgn}(x)[u_r'(y),u_r'(x)]} -d \ln (u^*)_r' \geq \gamma \operatorname{sgn}(x)(\ln|y| - \ln|x|) \\ &= \gamma \int_{]x,y]} \operatorname{sgn}(\xi) d \ln |\xi|,\end{aligned}$$

from which equation (2.4.23) follows, because this holds uniformly for all intervals $]x, y]$, i.e. independent of K.

For intervals which contain elements in $(D^u_{\text{eff}})^c$, equation (2.4.23) holds trivially, because by definition the measure ρ_u is infinite on this set, whereas the measure $d \ln |x|$ is finite. □

Remark 2.4.19. It is possible that the closed intervals $[u_r'(b), u_r'(a)]$ and $[-u_r'(a), -u_r'(b)]$, respectively, are not subsets of D. This is the case if $u_r'(a) = \infty$ or $u_r'(b) = 0$. In this case, it is enough to assume that the relative risk aversion is essentially bounded from above on the intersection of this interval and D. Indeed, the requirement of being essentially bounded from above is only needed if $x, y \in D^u_{\text{eff}}$. If $u_r'(x) < u_r'(a)$ and $u_r'(y) > u_r'(b)$, the restriction of the assumption to D (that is to the open interval) suffices. But if $u_r'(y) = 0$ or $u_r'(x) = \infty$, x or y are not elements of D^u_{eff}, and therefore equation (2.4.23) trivially holds.

2.5 Conclusion

In this chapter, we have generalized the notion of risk aversion to nonsmooth utility functions. To this end, we have introduced the concept of risk aversion measures, from which the classical absolute as well as the relative risk aversion, here denoted as risk aversion density, is calculated as a Radon-Nikodym derivative, provided it exists. However, the advantage of the risk aversion measure is that it can be defined for all utility functions, in particular for nonsmooth ones.

It has turned out that the one-to-one relationship between equivalence classes of utility functions and risk aversion densities, a well-known result in the smooth case, can be extended to the nonsmooth case considering a suitable class of measures. The equivalence class of utility functions has been defined in the classical way through positive monotone transformations, but without any assumption on differentiability or strict monotonicity.

Using this notion of risk aversion measures, we have defined an ordering for risk aversions of different investors, that is of different nonsmooth utility functions. For the case where the utility functions are smooth, this ordering coincides with the classical one for densities. Furthermore, we have proved an extension of a classical result of Pratt for nonsmooth utility functions. The connection between u_1 being more risk averse than u_2 and the existence of a concave function T with $u_1(x) =$

$T(u_2(x))$ does still hold for nonsmooth utility functions, if we express "more risk averse" in terms of risk aversion measures. Furthermore, we have shown that the notion of more risk averse in terms of risk aversion measures is also consistent with a reasonable alternative definition which does not necessarily use utility functions.

Typically, relative risk aversion makes sense for utility functions which are defined on the positive real line, that is for positive wealth. We did not try to extend this concept to utility functions which are defined for positive as well as for negative wealth. However, we also treated, alternatively to the typical case, the case where the wealth is always negative. We could give a unifying definition for both cases. However, for some specific proofs, we had to treat both cases separately. For absolute risk aversion, one can take the whole real line as domain, and the restriction mentioned above does not apply.

We have proposed a weaker ordering which we called essential bounds for the risk aversion, and which requires only that the risk aversion is bounded up to a certain tolerance. We have formalized this informal statement, and have shown that a strict bound is always an essential bound. We have given examples where there is no essential bound, as well as where even a piecewise linear function has essentially constant relative risk aversion.

Similarly to the fact the constant relative risk aversion property is equivalent to a power property of the utility function, we have shown that a bound of the relative risk aversion is equivalent to a super-power property of the function. The same holds for absolute risk aversion and exponential property. We have also shown that this equivalence continues to hold for nonsmooth utility functions using our definition, and holds not only for strict but also for essential bounds of the risk aversion, when the super-power property is appropriately relaxed.

Finally, it has been shown how the relative risk aversion transforms into the one of the concave conjugate function, for strict bounds as well as for essential bounds.

Chapter 3

Equivalence of the minimax martingale measure for satiated utility functions

3.1 Introduction

In an incomplete market model, option prices cannot be determined from arbitrage considerations alone. A well known technique to deal with this situation is the utility indifference argument stated e.g. in Schachermayer [76] and in Bellini and Frittelli [7], namely that the option price should be such that an investor investing optimally with respect to his utility function u should be indifferent of first order between whether or not to invest a small amount in the option. Following [7], this means that

$$\sup_{X \in C(x)} E[u(X)] = \sup_{X \in L^\infty : p(X) \leq x} E[u(X)],$$

where $C(x)$ is the set of superreplicable claims at initial portfolio value x, whereas on the right hand side, the optimization is done over all claims with price less than or equal to x.

The minimax martingale measure [7] for a given initial wealth x is the probability measure $\hat{Q}_x$ which minimizes the maximal attainable utility at $U(Q, x)$ at a price x,

$$U(\hat{Q}_x, x) \leq U(Q, x) \; \forall Q \in M_1,$$

where M_1 is the set of absolutely continuous separating measures. Here a separating measure is, as in [7], an absolutely continuous probability measure under which all claims that are replicable at initial portfolio value 0 have a nonpositive expectation, and

$$U(Q, x) := \sup\{E[u(X)] : X \in L^\infty, E^Q[X] \leq x\}.$$

The existence of such a minimax measure has been proved in a very general setup in Bellini and Frittelli [7]. As already mentioned there, this minimax measure is exactly the pricing rule required in order to guarantee that the supremum of the attainable expected utilities with respect to all (not necessarily replicable) claims with price smaller than or equal to the initial wealth is not larger than the maximal expected utility when considering only replicable claims. Furthermore, by the duality theory, the minimax measure can be thought as dual minimizer and can therefore, following e.g. Schachermayer [76] or earlier papers such as Karatzas et al. [53], be used for finding the optimal terminal value of a portfolio.

In general, even if the minimax measure exists, is it not necessarily equivalent to the objective probability measure P. Using the quadratic utility function, it has already in Delbaen and Schachermayer [25] been stated that even for a model with only three states, this measure is not necessarily equivalent.

Why is the question of equivalence an interesting one? When assuming a market which is free of arbitrage and thinking about a representative investor with utility function u and initial wealth x, then, when completing the market by the minimax martingale measure, one would like to have that the completed market is free of arbitrage as well. This is not the case when the minimax measure is only absolutely continuous with respect to P. Indeed, let Q_x be the minimax measure. If it is not equivalent, then there is a measurable set A with $P(A) > 0$ and $Q_x(A) = 0$. It follows that the L^∞-claim 1_A is nonnegative and positive with strictly positive probability. However, due to the pricing rule $p(1_A) = E^{\hat{Q}_x}[1_A] = 0$, this claim has zero price and is therefore an arbitrage opportunity. Such a price system therefore does not seem to be very reasonable. In this sense, one can see the question of equivalence as a test of the model concerning its reasonability.

Having motivated why the question of equivalence may be interesting, we now turn to its answer. For the case of strictly increasing utility functions, including those which satisfy the Inada conditions, this question has essentially been answered positively in the literature, provided there exists an equivalent martingale measure which is L^ϕ-integrable where Φ is the convex conjugate function. This follows essentially from Bouchard et al. [14] or Kabanov and Stricker [52], as well as Bellini and Frittelli [7]. We will refer to this case of strictly increasing utility functions as the classical case. We will review the equivalence result of this class in section 2 of the chapter.

An alternative class of utility functions is the one of satiated utility functions, i.e. the class of nondecreasing functions such that there exists a satiation point above which the function remains constant. Examples for this class are one-sided risk functions, where an investor is only interested in the risk of not achieving a certain benchmark, but not in the opportunity of exceeding it. Related to this is also the expected shortfall in the sense of Föllmer and Leukert [33], except that here we do not deal with state-dependent utility functions. Another example is the class of power utility functions with power larger than 1. Indeed, even if those functions are decreasing after their satiation point, it follows from the definition that the minimax martingale measure is the same as for the function which remains constant after the satiation point. The minimax martingale measures of this class are the q-optimal measures with $q > 1$. In particular, for continuous price processes, the variance-optimal measure coincides with the 2-optimal measure, as pointed out in Bellini and Frittelli [7].

For the specific case of the variance-optimal measure, it has already been shown in Delbaen and Schachermayer [25] that this measure is equivalent if the price processes are continuous and if there exists an equivalent martingale measure which is square integrable. However, if the price processes are discontinuous, simple counterexamples already exist in finite state models.

It is not clear how to construct a unifying theory which includes both satiated and unsatiated utility functions. The underlying assumptions are quite different, and we will also give a proof which is quite different from the one for the standard case which uses strictly increasing utility functions. On the one hand, for strictly increasing utility functions, a linear interpolation between the existing equivalent measure and the minimax martingale measure is done, an argument which cannot be used if the utility function is not strictly increasing. On the other hand, for satiated utility functions, we need the continuity of the processes and use an announcing sequence of stopping times, an argument which does not hold for discontinuous processes.

Instead of trying to unify the two situations, we treat both cases separately. Because the classical case of strictly increasing utility functions, including those satisfying the Inada conditions, essentially already follows from earlier research, the main contribution of our chapter is to prove equivalence for the alternative class of satiated utility functions. This is an extension of the results of Delbaen and Schachermayer [25] from the variance-optimal measure to a general minimax martingale measure for utility functions of this class. For this, we need some assumptions on the relative risk aversion which are trivially satisfied in the case of the variance-optimal measure. Furthermore we define a conditional form of the Luxemburg norm for Young functions and prove a Hölder inequality for this case, which replaces the Cauchy-Schwarz inequality of the proof in [25]. Because simple counterexamples exist for the case of discontinuous processes, we restrict the discussion to the case where the filtration is continuous.

The idea of using Orlicz spaces and Luxemburg norms in Mathematical Finance has perhaps first been applied in Haezendonck and Goovaerts [40] for the calculation of insurance premiums. Subsequently, in Delbaen [23], this concept has been used in the context of coherent risk measures. In Biagini and Frittelli [9], Orlicz spaces and Luxemburg norms have been used for the problem of utility maximization in an incomplete market and its dual problem. In this paper, the authors obtained a unified framework for all cases of utility maximization. However, this problem is not the same as the one in our chapter, and even if we apply the idea of using Orlicz spaces, it is not obvious how to obtain a unified framework for our problem. The reason is that for the considerations in [9], only the left end of the utility function is relevant, corresponding to the behavior of the dual function at infinity. In contrast, for our problem, the behavior of the dual function near 0 has to be considered as well.

The outline of this chapter is as follows. In section 2, we again state the definition of the minimax martingale measure and prove the classical case of strictly increasing utility functions. In section 3, we state some technical issues about Young functions and relative risk aversion which will be needed in the sequel. In section 4, we define a conditional form of the Luxemburg norm for Young functions and prove the Hölder inequality in this case. The main proof concerning the equivalence of the minimax martingale measure for the situation of satiated investors is presented in section 5. Counterexamples showing why some key considerations do not work without the assumptions are given in section 6. Applications to q-optimal measures are given in section 7. Section 8 concludes.

3.2 Classical case: strictly increasing utility functions

Here we recall the definition of the minimax martingale measure from Bellini and Frittelli [7]. As a notational convention, throughout this whole chapter, by the expectation $E[\cdot]$ we mean the expectation under the measure P, unless something else is explicitly mentioned.

Definition 3.2.1. [Minimax martingale measure] An absolutely continuous separating measure $\hat{Q}_x$ is a *minimax martingale measure* if it satisfies

$$\sup_{w \in L^\infty : E^{\hat{Q}_x}[w] \leq 0} \{E[u(x+w)]\} = \min_{Q \in M_1} \sup_{w \in L^\infty : E^{Q}[w] \leq 0} \{E[u(x+w)]\},$$

where M_1 is the set of all absolutely continuous separating measures, i.e.

$$M_1 := \{z \in L^1_+(P) : E^P[zw] \leq 0 \ \forall w \in C, E^P[z] = 1\},$$

where C is the convex cone of superreplicable claims at zero initial portfolio value.

Remark 3.2.2. If the filtration is continuous, it follows by Lemma 1.1 of Bellini and Frittelli [7] that the set of absolutely continuous separating measures corresponds to the set of absolutely continuous local martingale measures. We will therefore not distinguish between those two notions if we work in an environment with continuous filtration.

Remark 3.2.3. By Corollary 2.1 of Bellini and Frittelli [7], it follows that if $\hat{Q}_x$ is a minimax martingale measure, it must satisfy

$$\min_{\lambda \geq 0} \left[\lambda x - E^P \left[u^* \left(\lambda \frac{d\hat{Q}_x}{dP} \right) \right] \right] \leq \min_{\lambda \geq 0} \left[\lambda x - E^P \left[u^* \left(\lambda \frac{dQ}{dP} \right) \right] \right], \tag{3.2.1}$$

where Q is any absolutely continuous separating measure, and $u^*(y)$ is the concave conjugate function of u, i.e.

$$u^*(y) := \inf_x \left(xy - u(x) \right). \tag{3.2.2}$$

Next, we present Theorem 3.2.4, which summarizes the well-known equivalence result for strictly increasing utility functions. It is essentially a combination of the results of Bellini and Frittelli [7] and Bouchard et al. [14], even if the authors of the latter paper formulated it in a slightly different way, in particular without the notion of minimax martingale measure. The proof is also based on those papers. An alternative idea of proof would be to use a similar idea as in Proposition 3.1 in Kabanov and Stricker [52] (see also Frittelli [36] for the specific case of exponential utility functions).

Theorem 3.2.4. *Let Z_0 be a minimax measure for a strictly increasing and concave utility function. Let $\Phi = -u^*$ be the convex conjugate function. Furthermore let there be an equivalent separating measure Z_1 and a constant $\lambda > 0$ such that*

$$E[\Phi(\lambda Z_1)] < \infty.$$

Then the minimax measure Z_0 is equivalent.

Proof. The case where the utility function is unbounded from above has already been shown in Bellini and Frittelli [7]. Let us therefore assume that the utility function is bounded. The set $S := \{(\lambda, \lambda Z) : \lambda \geq 0, Z \in M_1\}$ coincides with the set $\tilde{Y}_+$ from Bouchard et al. [14], as will be shown below. Then the dual problem

$$W(x) := \inf_{(y,Y)\in\tilde{Y}_+} E[\tilde{U}(Y) + xy - YB]$$

from [14] is precisely the same as the minimum from inequality (3.2.1) (with the convex conjugate $\tilde{U} = -u^*$, and $B = 0$). By the assumptions of the theorem, the assumptions for Proposition 3.1 in [14] are satisfied. It follows that the minimum satisfies $\lambda^* Z^* > 0$ almost surely. Because the minimax martingale measure must satisfy inequality (3.2.1), the result follows.

It remains to show that S and $\tilde{Y}_+$ coincide. For the first variable of the pair this is obvious, therefore it is enough to show that M_1 is the same as $Y_1 := \{Z \in L^1_+ : E[ZX] \leq x \ \forall\ x \in \mathbb{R}_+, X \in \mathcal{X}_+(x)\}$, with $\mathcal{X}_+(x)$ the set of all nonnegative random variables replicable at initial capital x. First let $Z \in Y_1$, and let $W \in C$. Then there exists a random variable $W_1 \geq -a$ with $W \leq W_1$ and W_1 replicable at 0. It follows that $W_1 + a \in \mathcal{X}_+(a)$, and because $Z \in Y_1$, $E[Z(W_1 + a)] \leq a$, from which it follows that $E[ZW] \leq E[ZW_1] \leq 0$, so that $Z \in M_1$. On the other hand, if $Z \in M_1$, let $W \in \mathcal{X}_+(x)$ for an $x \in \mathbb{R}_+$. Then $X - x$ is a claim attainable at 0 and bounded from below, and therefore in C. It follows that $E[Z(X - x)] \leq 0$, from which $Z \in Y_1$ follows. □

3.3 Generalized Young functions and relative risk aversion

A specific equivalence result has been proved in Delbaen and Schachermayer [25] for quadratic utility functions, that is for an optimization in L^2. It is easy to generalize those results to L^p. A natural generalization of L^p-spaces are Orlicz spaces with Luxemburg norm, where an integrability with respect to a Young function is required instead of integrability with respect to $|x|^p$.

We will show later that in the case where the utility function has a satiation point c such that $u(x) = u(c)$ for $x > c$, the convex conjugate $\Phi = -u^*$ from inequality (3.2.1) is, up to trivial transformations, a Young function. To be precise, we have to generalize the notion of Young functions slightly, but the ideas are the same. It therefore makes sense to study this class of functions. The later results are connected to the boundedness of the relative risk aversion. Because we want to apply the theory not only to functions which are twice differentiable, we apply a generalized definition of relative risk aversion which has been developed in Chapter 2.

3.3.1 Generalized Young functions

Definition 3.3.1. [Generalized Young function] A function $\Phi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is a *generalized Young function* if it satisfies the following properties:

1. $\Phi(0) = 0$.
2. $\Phi(x)$ is convex and lower semicontinuous for $x \in \mathbb{R}$.
3. There exists a constant $c > 0$ such that $\Phi(x) < \infty$ for all $|x| \leq c$.
4. $\Phi(x) \to \infty$ for $x \to \infty$.
5. $\Phi(x) = \Phi(-x)$ for any $x \in \mathbb{R}$.

Remark 3.3.2. Biagini and Frittelli [9] actually denoted this class of functions simply as Young functions.

Definition 3.3.3. [Conjugate function] A function $\Phi^*(y)$ is *conjugate* to a generalized Young function $\Phi(x)$ if it satisfies for any $x \in \mathbb{R}$

$$\Phi^*(y) = \sup_{x \in \mathbb{R}} \left(xy - \Phi(x) \right). \tag{3.3.3}$$

Lemma 3.3.4. *Let Φ be a generalized Young function. Then the following statements hold:*

- *$\Phi(x)$ is continuous on the interior of $\{\Phi(x) < \infty\}$.*
- *$\Phi(x)$ is nonnegative.*
- *If Φ^* is defined by equation (3.3.3), then Φ^* is also a generalized Young function.*
- *$(\Phi^*)^* = \Phi$.*

Proof. The first and second statement are obvious, the third has been stated also in [9], and the fourth one follows from the Fenchel-Moreau theorem [5]. □

Definition 3.3.5. [Orlicz space] Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\Phi(x)$ be a generalized Young function. The *Orlicz space* L^Φ defined by Φ is the space of all $\mathcal{F}$-measurable random variables X for which there exists a $\lambda > 0$ such that

$$E\left[\Phi\left(\frac{X}{\lambda}\right)\right] < \infty.$$

The Orlicz space can then be endowed with the *Luxemburg norm*, which is defined, using the notion of Biagini and Frittelli [9], by

$$N_\Phi(X) := \inf\{\lambda > 0 : E\left[\Phi\left(\frac{X}{\lambda}\right)\right] \leq 1\}. \tag{3.3.4}$$

We now prove a technical proposition for generalized Young functions which we will need later in the chapter.

Proposition 3.3.6. *Let the generalized Young function $\Phi(x)$ be smooth at 0. Then there exists an $y_0 > 0$ such that the conjugate function $\Phi^*(y)$ is invertible for $0 \leq y \leq y_0$, and there exists a constant $x_0 \leq (\Phi^*)^{-1}(y_0)$ such that*

$$\frac{1}{2} \leq \frac{\Phi(x)}{x(\Phi^*)^{-1}(\Phi(x))} \leq 1 \quad \forall\, 0 \leq x \leq x_0 : \Phi(x) > 0. \tag{3.3.5}$$

Proof. By the Young inequality, we have for any x, y that

$$xy \leq \Phi(x) + \Phi^*(y).$$

In particular, this inequality holds for $y := (\Phi^*)^{-1}(\Phi(x))$. The inverse exists for x sufficiently small, first by the continuity of $\Phi(x)$ for small values of x, and second because $\Phi^*(y) > 0$ for all $y > 0$ and therefore strictly increasing, by the fact that $\Phi(x)$ is smooth at 0. It follows that

$$x(\Phi^*)^{-1}(\Phi(x)) \leq \Phi(x) + \Phi^*((\Phi^*)^{-1}(\Phi(x))) = 2\Phi(x).$$

Dividing by $2x(\Phi^*)^{-1}(\Phi(x))$ one obtains the inequality on the left hand side of (3.3.5).

On the other hand, there exists for every $y \geq 0$ a $0 \leq x^{opt} \leq \infty$ such that

$$x^{opt} y = \Phi(x^{opt}) + \Phi^*(y). \tag{3.3.6}$$

This follows from the fact that

$$\Phi^*(y) = \sup_{x \geq 0} \left(xy - \Phi(x)\right).$$

If there is an $x_{max} > 0$ with $x_{max} y \leq \Phi(x_{max})$ and $\Phi(x_{max}) < \infty$, then $xy - \Phi(x)$ is a continuous function on a compact interval $[0, x_{max}]$ and therefore attains its maximum at x^{opt}. Because of the convexity of Φ, $\Phi^*(y)$ cannot be larger than $x^{opt} y - \Phi(x^{opt})$.

If $x_{max} y \leq \Phi(x)$ only holds when $\Phi(x_{max}) = \infty$, but still $x_{max} < \infty$, then there is, by the convexity and the lower semicontinuity of Φ, an $\tilde{x}$ such that $\Phi(\tilde{x}) < \infty$ and $\Phi(x) = \infty$ for $x > \tilde{x}$. Again, $xy - \Phi(x)$ is a continuous function on the compact interval $[0, \tilde{x}]$ and therefore attains its maximum at x^{opt}. Because for any larger x, $xy - \Phi(x) = -\infty$, the maximum at x^{opt} is also the value of $\Phi^*(y)$. If there exists no $x_{max} < \infty$ with $x_{max}y \leq \Phi(x_{max})$, then, by the convexity of $\Phi(x)$, the function $xy - \Phi(x)$ is increasing for all $x \geq 0$. This means that $x^{opt} = \infty$.

Now let $y = (\Phi^*)^{-1}(\Phi(x))$. Again, this number is well defined and small if x is sufficiently small. If $x < x^{opt}$, then $\tilde{x}y - \Phi(\tilde{x})$ is monotonically increasing on $[0, x^{opt}[$, and therefore $\tilde{x}y \geq \Phi(\tilde{x})$ for all $\tilde{x} < x^{opt}$. This holds in particular for $\tilde{x} = x$, and therefore

$$x(\Phi^*)^{-1}(\Phi(x)) = xy \geq \Phi(x),$$

and the inequality on the right hand side of (3.3.5) is proved. If $x \geq x^{opt}$, then $x^{opt} < \infty$ and

$$x(\Phi^*)^{-1}(\Phi(x)) \geq x^{opt}(\Phi^*)^{-1}(\Phi(x)) = \Phi(x^{opt}) + \Phi^*((\Phi^*)^{-1}(\Phi(x))) \geq \Phi(x),$$

because by the nonnegativity of Φ we have $\Phi(x^{opt}) \geq 0$, and for the equality equation (3.3.6) has been applied. □

3.3.2 Relative risk aversion

Here we briefly present the definition of essential bounds for the relative risk aversion. For a more detailed introduction and discussion we refer to Chapter 2. As stated in this chapter, if $\Phi(x)$ is a generalized Young function, we have that for $x < 0$, the function $x \mapsto -\Phi(-x)$ is a utility function satisfying the main assumptions in Chapter 2. The relative risk aversion is then defined according those rules.

Definition 3.3.7. [Essential bounds for relative risk aversion] A generalized Young function $\Phi(x)$ has *essentially a relative risk aversion bounded from above by* γ, $0 < \gamma < \infty$, on a set $]0, b[$, if $\Phi(x)$ only vanishes at 0 and there exists a constant $0 < K < \infty$ such that

$$\int_{]x,y]} d \ln \Phi'_l \leq \int_{]x,y]} \gamma d \ln x + K \tag{3.3.7}$$

for all intervals $]x, y] \subset]0, b[$. It has *a relative risk aversion essentially bounded from below by* γ if $\Phi(x)$ only vanishes at 0 and there exists a constant $0 < K < \infty$ such that

$$\int_{]x,y]} d \ln \Phi'_l \geq \int_{]x,y]} \gamma d \ln x - K. \tag{3.3.8}$$

It has *essentially a risk aversion of* γ if both (3.3.7) and (3.3.8) are valid.

Remark 3.3.8. By Remark 3.4 of Chapter 2, $\Phi(x) = -u(-x)$, where u is defined on the negative real line and Φ on the positive real line, and therefore the right derivative u'_r used in Chapter 2 translates into the left derivative Φ'_l.

Finally, we prove a technical proposition for the relative risk aversion, which will be needed when proving equivalence of the minimax martingale measure.

Proposition 3.3.9. *Let the generalized Young function $\Phi(x)$ have a relative risk aversion which is essentially bounded from below by $\gamma > 0$ in a region around 0. Then we have for all sequences $x_n \to 0$ and $p_n \to 0$ that*

$$\frac{(\Phi^*)^{-1}(p_n x_n)}{(\Phi^*)^{-1}(x_n)} \to 0.$$

Proof. By the fact that the relative risk aversion of Φ is essentially bounded from below around 0, it follows that there exists an $\epsilon > 0$ such that Φ is essentially bounded from below on the nonempty interval $]0 = (\Phi^*)'(0), (\Phi^*)'_l(\epsilon)[$. From Chapter 2, we have for the conjugate function that $\Phi^*(y)$ has a relative risk aversion which is essentially bounded from above by $\frac{1}{\gamma}$ if $y > 0$ is sufficiently small. Furthermore, because $0 < \Phi^*(y) < \infty$ for $y > 0$ sufficiently small, the inverse $(\Phi^*)^{-1}(x)$ exists and satisfies the inequality

$$\frac{y}{x} \leq \left(K \frac{(\Phi^*)^{-1}(y)}{(\Phi^*)^{-1}(x)} \right)^{\frac{\gamma+1}{\gamma}}$$

according to the considerations in Chapter 2. Applying this with $x = p_n x_n$ and $y = x_n$ we get

$$\frac{(\Phi^*)^{-1}(p_n x_n)}{(\Phi^*)^{-1}(x_n)} \leq K p_n^{\frac{\gamma}{\gamma+1}}.$$

Because $\gamma > 0$, the right hand side tends to 0 as $p_n \to 0$. □

3.4 Conditional Luxemburg norm and Hölder inequality

The aim of this section is to get a conditional version of the Hölder inequality for general Luxemburg norms. We also have to make a further generalization with respect to the definition (3.3.4) of the Luxemburg norm. Indeed, the norm defined there in principle gives a comparison of random variables at the point where $E[\Phi(\lambda X)] = 1$. In L^p spaces, this does not matter, because $E[\Phi(\lambda X)] = \lambda^p E[\Phi(X)]$, so that if we would replace the constant 1 in (3.3.4) by another constant c, we would get an equivalent norm. However, this is no longer true for a general function $\Phi(x)$. The fact that a random variable X has a larger Luxemburg norm than a random variable Y does not imply that $E[\Phi(X)] \geq E[\Phi(Y)]$, if we are not in the L^p case. For later finding an equivalent martingale measure $\tilde{Z}$ which shows that an absolutely continuous measure Z^{opt} cannot be a minimax measure, we have to compare the measures at the right point. The point 1 is completely arbitrary. In general, for the conditional version of the Luxemburg norm, we only take a strictly positive random variable which is measurable with respect to the sub-sigma-algebra.

A final issue is that a power function $\Phi(x) = x^p$ always satisfies the so-called Δ_2 condition stated for example in Bloom and Kerman [11] or Biagini and Frittelli [9], which in particular implies for any $t_0 > 0$ that

$$E[\Phi(t_0 X)] < \infty \Rightarrow E[\Phi(tX)] < \infty \ \forall t \geq 0.$$

This is in general not satisfied by a generalized Young function. Therefore, on some subsets of Ω, the condition in the set in equation (3.3.4) may still be satisfied for a specific λ, even if the expectation is infinite on the total Ω for this λ. Therefore, for a definition of a conditional Luxemburg norm, one would like to have a conditional expectation even if the random variable is not integrable. On the other hand, by the definition of the generalized Young functions, the random variables are always nonnegative. We therefore have to use the nonnegative version of the conditional expectation, which has been discussed in Shiryaev [82].

It is quite obvious that the main results on the existence of a right- or left-continuous version of conditioned processes also hold if we take the conditional expectation for nonnegative processes instead of the integrability condition. Because we did not find a reference for this, we put the results needed, including proofs, into an appendix.

3.4.1 Conditional Luxemburg norm

For having uniqueness, the definition of the conditional Luxemburg norm uses a right-continuous version of a conditional expectation. In order to guarantee that such a version exists, we first prove the following lemma.

Lemma 3.4.1. *Let Φ be a generalized Young function, $(\Omega, \mathcal{F}, P)$ a probability space, X a random variable in the Orlicz space L^Φ, and $\mathcal{G}$ a sub-sigma-algebra. Then a version of the conditional expectation*

$$E\left[\Phi\left(\frac{X}{\lambda}\right)|\mathcal{G}\right]$$

which is right-continuous and monotonically decreasing in λ exists. We will denote this version in the sequel as $E^{rc}[\cdot]$ for emphasizing that the right-continuous version is meant.

Proof. Define with $t := \frac{1}{\lambda}$

$$Z_t := E\left[\Phi\left(tX\right)|\mathcal{G}\right] = E\left[\Phi\left(\frac{X}{\lambda}\right)|\mathcal{G}\right]. \tag{3.4.9}$$

Then Z_t satisfies the assumptions of Proposition 3.A.3, and therefore has a left-continuous version, which is furthermore monotonically increasing. Therefore $Z_\lambda = Z_{\frac{1}{t}}$ has a right-continuous version which is monotonically decreasing in λ. □

Definition 3.4.2. [Conditional Luxemburg norm] For a generalized Young function Φ, a sub-sigma-algebra $\mathcal{G} \subset \mathcal{F}$, a random variable X in the Orlicz space L^Φ, and for a $\mathcal{G}$-measurable, nonnegative and integrable random variable ξ, the *conditional Luxemburg norm* is the $\mathcal{G}$-measurable random variable $\Omega \to [0,\infty]$ given by

$$N_\Phi^\xi(X|\mathcal{G})(\omega) := \begin{matrix} \inf \Lambda(\omega) & \text{if} \quad \Lambda(\omega) \neq \emptyset, \\ \infty & \text{if} \quad \Lambda(\omega) = \emptyset, \end{matrix} \tag{3.4.10}$$

where the set $\Lambda(\omega)$ is defined as

$$\Lambda(\omega) := \{\lambda > 0, E^{rc}\left[\Phi\left(\frac{X}{\lambda}\right)|\mathcal{G}\right] \leq \xi\} \tag{3.4.11}$$

and the notion of E^{rc} means a version of the conditional expectation for nonnegative random variables which is right-continuous in λ.

Theorem 3.4.3. *The conditional Luxemburg norm as defined above exists and is unique in the sense that if λ and $\tilde{\lambda}$ are two versions of the conditional Luxemburg norm, then*

$$P[\lambda = \tilde{\lambda}] = 1.$$

Furthermore, if λ is the conditional Luxemburg norm defined in (3.4.10) and (3.4.11), we have on $\{\lambda > 0\}$

$$E\left[\Phi\left(\frac{X}{\lambda}\right)|\mathcal{G}\right] \leq \xi \ a.s., \tag{3.4.12}$$

whereas on $\{\lambda = 0\}$, we have $E[\Phi(X)|\mathcal{G}] = 0$.

Moreover, if ξ is strictly positive, then the conditional Luxemburg norm is almost surely finite.

Proof. Consider for the moment a specific right-continuous version of the conditional expectation. Then for each $\omega \in \Omega^*$, $P[\Omega^*] = 1$, the set $\Lambda(\omega)$ is uniquely defined as a subset in $\mathbb{R}$ which is bounded by 0 from below. Therefore the infimum of $\Lambda(\omega)$ exists for all $\omega \in \Omega^*$ for which it is nonempty. We may therefore define the conditional Luxemburg norm by (3.4.10) without ambiguity on Ω^* .

We now show that this definition is independent of the choice of the version of the right-continuous conditional expectation. Let Y_1 and Y_2 be two right-continuous versions of the conditional expectation. It follows that they are indistinguishable, i.e. there exists a set $\Omega^{**} \subset \Omega$ with $P[\Omega^{**} = 1]$ and on which $Y_1 = Y_2$. The set

$\Omega^* \cap \Omega^{**}$ has probability 1, and the sets $\Lambda(\omega)$ from (3.4.11) are the same for every $\omega \in \Omega^* \cap \Omega^{**}$. It follows that the definition is unique on a set with probability 1.

Because $X \in L^\Phi$, we have by definition that X is almost surely finite, and $\frac{X}{\lambda} \to 0$ as $\lambda \to \infty$, and the same property holds for $\Phi(\frac{X}{\lambda})$, by the continuity of generalized Young functions at 0. It follows that the left-continuous modification of the process in (3.4.9) satisfies the assumptions of Corollary 3.A.5 and therefore converges to 0 almost surely. If $\xi > 0$ strictly, it follows that $\Lambda(\omega)$ is almost surely nonempty, and therefore the infimum in (3.4.10) exists and is finite.

Now let the random variable defined in (3.4.10) and (3.4.11) be denoted by $\lambda^*(\omega)$. We have seen that this random variable exists and is unique. We need to show that $\lambda^*(\omega)$ is $\mathcal{G}$-measurable. Let $\gamma \in \mathbb{R}$. We need to show that $\{\omega : \lambda^*(\omega) \leq \gamma\} \in \mathcal{G}$. But for $\gamma < \infty$, we have

$$\{\lambda^* \leq \gamma\} = \{Y_\gamma := E^{rc}\left[\Phi\left(\frac{X}{\gamma}\right)|\mathcal{G}\right] \leq \xi\}.$$

To prove this, let ω be in the set on the left. Then the set $\Lambda(\omega)$ is nonempty, and there exists a sequence λ_n converging to λ^* from above. Because $\lambda_n \in \Lambda(\omega)$, it satisfies equation (3.4.11), and $Y_{\lambda_n} \leq \xi$. Because Y_λ is right-continuous, $Y_\gamma \leq \xi$ as well. If on the other hand $Y_\gamma \leq \xi$, then by definition $\gamma \in \Lambda(\omega)$, and its infimum λ^* is always smaller than or equal to γ. Because Y_γ is obviously $\mathcal{G}$-measurable, the result follows.

It remains to prove equation (3.4.12). On the set $\{\lambda^* = \infty\}$, we have $\Phi(\frac{X}{\lambda^*}) = \Phi(0) = 0$, and therefore equation (3.4.12) is satisfied, because ξ is nonnegative. Now consider the set $\{\lambda^* < \infty\}$. It is suffices to show the inequality for the sets $S := \{\lambda^* \leq \lambda_{max}\}$ for all $\lambda_{max} > 0$. Because $\lambda^* \in \mathcal{G}$ and λ^* is bounded from above on S, we may approximate it from above by step functions $\lambda_n := \sum_k \lambda_{nk} 1_{A_{nk}}$ with $A_{nk} \in \mathcal{G}$. On the set A_{nk}, we therefore have that $\lambda(\omega) \leq \lambda_{nk}$, and by the monotonicity and right-continuity of the conditional expectation in (3.4.11) it follows that $\lambda_{nk} \in \Lambda(\omega)$ on A_{nk}. This implies by definition (3.4.11)

$$E\left[\Phi\left(\frac{X}{\lambda_{nk}}\right)|\mathcal{G}\right] \leq \xi$$

on the set A_{nk}, and therefore

$$E\left[\Phi\left(\frac{X}{\lambda_n}\right)|\mathcal{G}\right] = \sum_k E\left[\Phi\left(\frac{X}{\lambda_{nk}}\right)|\mathcal{G}\right] 1_{A_{nk}} \leq \xi$$

on the whole set S. Because this holds for every step function on S, we conclude that inequality (3.4.12) holds by the monotone convergence theorem. Finally, on $\{\lambda^* = 0\}$, we may take the sequence $\lambda_{nk} = \frac{1}{n}$, and it follows that

$$E[\Phi(nX)|\mathcal{G}] \leq \xi$$

for all $n \in \mathbb{N}$. But because $\Phi(nX)$ tends to infinity on $\{X > 0\}$ and ξ is almost surely finite, this can only happen if $X = 0$ and therefore $E[\Phi(X)|\mathcal{G}] = 0$ on the set $\{\lambda^* = 0\}$. This proves Theorem 3.4.3. □

3.4.2 Hölder inequality

We now prove a generalization of the Hölder inequality in conditional form, which will be important for the development below:

Theorem 3.4.4. *Let Φ and Φ^* be complementary Young functions which vanish only at 0, and let ξ be a strictly positive random variable. Then*

$$E[|XY||\mathcal{G}] \leq 2\xi N^{\xi}_{\Phi}(X|\mathcal{G}) N^{\xi}_{\Phi^*}(Y|\mathcal{G}). \tag{3.4.13}$$

Furthermore, if $X \in L^{\Phi}$, $Y \in L^{\Phi^*}$, then XY is integrable.

Proof. It is enough to prove the statement for nonnegative X and Y. Let us first assume that the conditional Luxemburg norms are almost surely strictly positive, and let λ and μ be two strictly positive $\mathcal{G}$-measurable random variables such that

$$\begin{array}{rcll} E\left[\Phi\left(\frac{X}{\lambda}\right)|\mathcal{G}\right] & \leq & \xi & \text{a.s.} \\ E\left[\Phi^*\left(\frac{Y}{\mu}\right)|\mathcal{G}\right] & \leq & \xi & \text{a.s.} \end{array} \tag{3.4.14}$$

By the Young inequality, we have that

$$\frac{XY}{\lambda\mu} \leq \Phi\left(\frac{X}{\lambda}\right) + \Phi^*\left(\frac{Y}{\mu}\right) \quad \text{a.s.}$$

By the monotonicity of the conditional expectation, this yields

$$E\left[\frac{XY}{\lambda\mu}|\mathcal{G}\right] \leq E\left[\Phi\left(\frac{X}{\lambda}\right) + \Phi^*\left(\frac{Y}{\mu}\right)|\mathcal{G}\right] \quad \text{a.s.}$$

But by (3.4.14) the right hand side is almost surely smaller than or equal to 2ξ. Because λ and μ are $\mathcal{G}$-measurable, we get

$$E[XY|\mathcal{G}] \leq 2\xi\lambda\mu \quad \text{a.s.}$$

By (3.4.12), inequality (3.4.14) holds in particular for $\lambda = N^{\xi}_{\Phi}(X|\mathcal{G})$ and $\mu = N^{\xi}_{\Phi^*}(Y|\mathcal{G})$, which yields (3.4.13).

If on a set with nonzero probability at least one of the conditional Luxemburg norms is 0, then, by Theorem 3.4.3, either $E[\Phi(X)|\mathcal{G}] = 0$ or $E[\Phi^*(Y)|\mathcal{G}] = 0$ on this set. Because Φ and Φ^* only vanish at 0, we have $E[XY|\mathcal{G}] = 0$ on this set, and (3.4.13) is still satisfied. □

3.5 Minimax measures for satiated utility functions

3.5.1 Definitions and assumptions

Throughout this section, we work in an environment with continuous filtration. It is known that without this assumption, the equivalence of the minimax martingale measure can already be violated in a three-state model (see Delbaen and Schachermayer [25]).

Assumption 3.5.1. The probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ has a continuous filtration $\mathcal{F}_t$.

It follows that every price process, as well as every density process, has to be continuous.

Remark 3.5.2. In [25], the authors used the weaker assumption that only the price processes are continuous. It is a specific feature of the variance-optimal measure that its density process under an equivalent local martingale measure is continuous if the price processes are, which cannot be expected to hold in general.

Here we work in the setting of Bellini and Frittelli [7] and therefore use the same assumptions on the utility functions, with the additional requirement that there is a satiation point.

Assumption 3.5.3. The utility function $u : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ is upper semicontinuous and concave on $\mathbb{R}$, and nondecreasing in its effective domain, which is assumed to have a nonempty interior. Furthermore u has a satiation point in the interior of its effective domain, i.e. a point c with $u(x) < u(c)$ for $x < c$ and $u(x) = u(c)$ for $x \geq c$.

Remark 3.5.4. Because c is in the interior of the effective domain of u, it follows that there exists a point $x < c$ with $u(x) > -\infty$.

Remark 3.5.5. A power utility function of the form $u(x) = -|x|^p$, $p > 1$, does not satisfy Assumption 3.5.3, because it is decreasing for $x > 0$. However, it can be replaced by the function $\tilde{u}(x) = -|\min(x, 0)|^p$, which satisfies this assumption. Because $E^Q[\min(X, 0)] \leq E^Q[X]$ for all measures Q and $E[u(\min(X, 0))] = E[\tilde{u}(X)]$, the minimax martingale measures according to Definition 3.2.1 coincide.

Assumption 3.5.6. There exists at least one equivalent separating measure $Z_\infty^{(0)}$ contained in L^1.

Assumption 3.5.7. There exists a minimax martingale measure Z_∞^{opt} according to Definition 3.2.1. For conditions which guarantee the existence of such a measure we refer to Bellini and Frittelli [7].

Because Z_∞^{opt} is integrable, it follows that the martingales

$$\begin{array}{rcl} Z_t^{(0)} & := & E[Z_\infty^{(0)}|\mathcal{F}_t] \\ Z_t^{opt} & := & E[Z_\infty^{opt}|\mathcal{F}_t] \end{array} \tag{3.5.15}$$

are uniformly integrable, and with the assumption that the filtration $\mathcal{F}_t$ is continuous, they are also continuous. The stopping times T and T_n are defined through

$$\begin{array}{rcl} T & := & \inf\{t \geq 0 : Z_t^{opt} = 0\}, \\ T_n & := & \inf\{t \geq 0 : Z_t^{opt} = \frac{1}{n}\} \wedge n. \end{array} \tag{3.5.16}$$

It follows that $T_n < T_{n+1} < T$ and $T_n \to T$ almost surely.

3.5.2 Utility functions and generalized Young functions

If $u(x)$ is a utility function satisfying Assumption 3.5.3, it is easy to see that

$$\Phi^*(x) := u(c) - u(c - |x|) \tag{3.5.17}$$

is a generalized Young function (Definition 3.3.1).

The next lemma shows why we can restrict our considerations to generalized Young functions:

Lemma 3.5.8. *Let the utility function $u(x)$ satisfy Assumption 3.5.3. Then*

1. *For all $y \geq 0$, the concave conjugate function as defined in (3.2.2) satisfies*

$$u^*(y) = yc - u(c) - \Phi(y), \tag{3.5.18}$$

where $\Phi(y) = \Phi^{**}(y)$ is the conjugate function to $\Phi^*(x)$, defined in (3.5.17).

2. *If $\hat{Q}_x$ is a minimax martingale measure (Definition 3.2.1), it satisfies*

$$\hat{\lambda} x_0 + E[\Phi(\hat{\lambda} Z_\infty^{opt})] \leq \lambda x_0 + E[\Phi(\lambda Z_\infty)] \quad \forall Z_\infty \in \mathcal{M},\ \forall \lambda > 0, \tag{3.5.19}$$

where $\mathcal{M}$ is the set of all absolutely continuous local martingale measures, and $\hat{\lambda}$ is the minimizing value.

Proof. For $y \geq 0$ we have

$$\begin{array}{rcl} \Phi(y) & = & \sup_x (xy - \Phi^*(x)) = \sup_x (xy - u(c) + u(c - |x|)) \\ & = & \sup_{x \geq 0} (xy - u(c) + u(c - x)) \\ & = & \sup_{x' \leq c} ((c - x')y - u(c) + u(x')) \\ & = & cy - u(c) + \sup_{x' \leq c} (u(x') - x'y) \\ & = & cy - u(c) - \inf_{x' \leq c} (x'y - u(x')) \\ & = & cy - u(c) - \inf_{x'} (x'y - u(x')) \\ & = & cy - u(c) - u^*(y). \end{array}$$

Statement 1 now follows.

For any $\lambda \geq 0$ and for any local martingale measure Q, it follows that

$$\lambda x - E\left[u^*\left(\lambda \frac{dQ}{dP}\right)\right] = \lambda x - \lambda c + u(c) + E\left[\Phi\left(\lambda \frac{dQ}{dP}\right)\right].$$

Hence, if $\hat{Q}_x$ is a minimax martingale measure, we have from Remark 3.2.3 that

$$\min_{\lambda \geq 0}\left[\lambda x_0 + E\left[\Phi\left(\lambda \frac{d\hat{Q}_x}{dP}\right)\right]\right] \leq \min_{\lambda \geq 0}\left[\lambda x_0 + E\left[\Phi\left(\lambda \frac{dQ}{dP}\right)\right]\right]$$

for all absolutely continuous local martingale measures Q, where $x_0 := x - c$. Performing the minimization over λ on the left hand side yields statement 2. □

3.5.3 Boundedness of the relative risk process

We now turn to the key reasons of the above considerations. We would like to prove an extension of the argument stated in Delbaen and Schachermayer [25]. Indeed we would like to show that if the relative risk aversion near 0 is essentially bounded from below away from 0, then the relative risk, as defined below, converges to infinity for a continuous martingale on the set where it converges to 0. On the other hand, if the equivalent martingale measure has enough integrability properties, its relative risk will remain bounded.

Definition 3.5.9. [Relative risk] Let X be a L^{Φ}-integrable random variable and let c be a constant. Then the *relative risk of X at the point c* is defined as

$$RR(c) := \frac{E[\Phi(cX)]}{\Phi(c)} \leq \infty. \tag{3.5.20}$$

Proposition 3.5.10. *Let Φ be a generalized Young function and let the relative risk aversion of $\Phi(x)$ be essentially bounded from above by $0 \leq \gamma \leq \infty$ for all $x \geq 0$. Furthermore, let X be a random variable which is $L^{\gamma+1}$-integrable, and let $\mathcal{F}_n$ be a filtration. If $\gamma < \infty$, then for every sequence $a_n \in \mathcal{F}_n$ with $a_n > 0$ which is uniformly bounded from below away from 0 for all ω, and for every sequence $c_n \in \mathcal{F}_n$ such that $c_n \to 0$ almost surely, the relative risk remains almost surely bounded, i.e.*

$$\frac{E\left[\Phi(c_n X)\mathcal{F}_n\right]}{\Phi(a_n c_n)} \leq K(\omega) < \infty \;\; a.s. \quad \forall \quad n \in \mathbb{N}. \tag{3.5.21}$$

If $\gamma = \infty$, then there exists for every ω a lower bound $\beta(\omega) > 0$ such that for every sequence $a_n > 0$, $a_n \in \mathcal{F}_n$, uniformly bounded from below by $\beta(\omega)$, the statement still holds for n large enough.

Proof. By the fact that the relative risk aversion is essentially bounded from above, it is clear that $\Phi(x) > 0$ for all $x > 0$. We first assume that $\gamma < \infty$. Then, from Chapter 2, we have

$$\Phi(c_n X) \leq K\Phi(a_n c_n) \left(\frac{X}{a_n}\right)^{\gamma+1} 1_{\frac{X}{a_n} \geq 1} + \Phi(a_n c_n) 1_{\frac{X}{a_n} \leq 1},$$

and taking the expectations we conclude that

$$E\left[\Phi(c_n X)|\mathcal{F}_n\right] \leq \Phi(a_n c_n) \left(KE\left[\left(\frac{X}{a_n}\right)^{\gamma+1} |\mathcal{F}_n\right] + 1\right).$$

By the assumption that X is contained in $L^{\gamma+1}$ and $\gamma \geq 0$ and that a_n is almost surely bounded from below away from 0, the expression within the brackets on the right hand side of the above estimate is almost surely bounded, and the result follows.

If $\gamma = \infty$ and $\Phi(x) > 0$ for all $x > 0$, then, by the monotonicity of Φ, we have

$$\Phi(c_n X) \leq \Phi(c_n ||X||_\infty)$$

and therefore

$$E\left[\Phi(c_n X)|\mathcal{F}_n\right] \leq \Phi(c_n ||X||_\infty).$$

Because Φ is a generalized Young function, there exists a constant $1 > b > 0$ with $\Phi(b) < \infty$. Because $c_n \to 0$, we must have an $N \in \mathbb{N}$ such that $c_n ||X||_\infty \leq b < 1$ for all $n \geq N$. Since Φ is convex, it follows that

$$\Phi(c_n ||X||_\infty) \leq c_N ||X||_\infty \Phi\left(\frac{c_n}{c_N}\right)$$

for $n \geq N$. Hence, with a lower bound of $a_n \geq \frac{1}{c_N}$ for $n \geq N$, we get, for $n \geq N$,

$$\frac{E[\Phi(c_n X)|\mathcal{F}_n]}{\Phi(a_n c_n)} \leq c_N ||X||_\infty.$$

□

We now prove a generalization of Lemma 3.4 of Delbaen and Schachermayer [25]. The idea of the proof is the same, with the Hölder inequality for Orlicz spaces instead of the Cauchy-Schwarz inequality, and some additional arguments.

Proposition 3.5.11. *Let X_t be a continuous uniformly integrable martingale with stopping times as in (3.5.16) for the process X_t instead of Z_t^{opt}. If the relative risk aversion of $\Phi(x)$ is essentially bounded from below away from 0 in a region around 0, then for all $\mathcal{F}_{T_n}$-measurable sequences $a_n > 0$ which are bounded from above in n for every ω, we have that the relative risk satisfies*

$$\frac{E[\phi(X_\infty)|\mathcal{F}_{T_n}]}{\phi(a_n X_{T_n})} \to \infty$$

on the set $\{X_T = 0\}$.

Proof. We take $\xi_n = \Phi(a_n X_{T_n})$ and apply the Hölder inequality, which is possible because X_{T_n} and therefore ξ_n are strictly positive, and $\Phi(x)$ as well as $\Phi^*(y)$ vanish only at 0, because of the essential bound from below. Then

$$E[X_\infty 1_{X_T \neq 0}|\mathcal{F}_{T_n}] \leq 2\Phi(a_n X_{T_n}) N_\Phi^{\Phi(a_n X_{T_n})}(X_\infty|\mathcal{F}_{T_n}) N_{\Phi^*}^{\Phi(a_n X_{T_n})}(1_{T\neq 0}|\mathcal{F}_{T_n}).$$

We have that Φ^* is invertible in a region around 0 and

$$\begin{aligned} N_{\Phi^*}^{\phi(a_n X_{T_n})}(1_{T\neq 0}|\mathcal{F}_{T_n}) &= \inf\{\lambda > 0 : \Phi^*\left(\tfrac{1}{\lambda}\right) p_n \leq \Phi(a_n X_{T_n})\} \\ &= \left((\Phi^*)^{-1}\left(\tfrac{\phi(a_n X_{T_n})}{p_n}\right)\right)^{-1} \end{aligned}$$

with $p_n = P[X_T \neq 0|\mathcal{F}_{T_n}] \to 0$ on the set $\{X_\infty = 0\}$, where the second equality follows if we define

$$(\Phi^*)^{-1}(x) := \inf\{y \geq 0 : \Phi^*(y) \geq x\}.$$

By the fact that $\Phi^*(y)$ is strictly increasing, this coincides with the usual inverse as long as $\Phi^*(y)$ is finite, and it remains constant after the point where $\Phi^*(y)$ jumps to ∞.

It follows that

$$\frac{1}{2} \leq N_\Phi^{\Phi(a_n X_{T_n})}(X_\infty|\mathcal{F}_{T_n}) \frac{\Phi(a_n X_{T_n})}{X_{T_n}(\Phi^*)^{-1}(\Phi(a_n X_{T_n}))} \frac{(\Phi^*)^{-1}(\Phi(a_n X_{T_n}))}{(\Phi^*)^{-1}(\frac{\Phi(a_n X_{T_n})}{p_n})}. \tag{3.5.22}$$

By Proposition 3.3.6, we have that the expression

$$\frac{\Phi(a_n X_{T_n})}{a_n X_{T_n}(\Phi^*)^{-1}(\Phi(a_n X_{T_n}))}$$

converges to a finite constant away from 0 if $a_n X_{T_n} \to 0$. Because the sequence a_n is bounded in n, this is always satisfied if $X_{T_n} \to 0$. By the fact that $(\Phi^*)^{-1}(x) \to 0$ as $x \to 0$ and $\Phi(a_n X_{T_n}) \to 0$ as well, the last fraction of equation (3.5.22) can only converge to a value different from 0 (or not converge) if $\frac{\Phi(a_n X_{T_n})}{p_n} \to 0$ as well. But in this case, Proposition 3.3.9 guarantees that this last fraction of equation (3.5.22) still converges to 0 if the relative risk aversion is essentially bounded from below away from 0 in a region of 0. It follows, again by the boundedness of the sequence a_n from above, that

$$N_\Phi^{\Phi(a_n X_{T_n})}(X_\infty|\mathcal{F}_{T_n}) \to \infty$$

on $\{X_T = 0\}$. This means that, almost surely, for every $\lambda > 0$ we may find an $n \in \mathbb{N}$ with

$$E[\Phi(\frac{X_\infty}{\lambda})|\mathcal{F}_{T_n}] \geq \Phi(a_n X_{T_n}).$$

By the convexity of Φ and $\Phi(0) = 0$, we have that $\Phi(\frac{X_\infty}{\lambda}) \leq \frac{1}{\lambda}\Phi(a_n X_\infty)$ for $\lambda \geq 1$, and thus

$$\frac{E[\Phi(X_\infty)|\mathcal{F}_{T_n}]}{\Phi(a_n X_{T_n})} \geq \lambda.$$

Because λ can be made arbitrarily large, the relative risk converges to infinity. □

3.5.4 Equivalence of minimax martingale measures

Theorem 3.5.12. *Let Φ be a generalized Young function, and let Assumptions 3.5.1, 3.5.3, 3.5.6 and 3.5.7 be satisfied. Further assume that the following holds:*

- *The relative risk aversion of $\Phi(x)$ is essentially bounded from below away from 0 in a neighborhood of 0.*
- *The relative risk aversion of $\Phi(x)$ is essentially bounded from above by a constant $\gamma \leq \infty$.*
- *There exists an equivalent martingale measure $Z_\infty^{(0)}$ which is an element of $L^{\gamma+1}$.*

Then any minimal martingale measure satisfying equation (3.5.19) is equivalent to the original measure P.

Proof. Here we follow the arguments of Delbaen and Schachermayer [25], which we can extend to this general situation using Propositions 3.5.10 and 3.5.11. Assume that Z_∞^{opt} is not equivalent, but satisfies equation (3.5.19). We therefore have, with T_n and T given as in section 3.5.1, that $\{Z_T^{opt} = 0\}$ has nonzero probability. Define, as in [25], the process

$$Z_t := \begin{array}{lll} Z_t^{opt} & \text{on} & A_n^c \cup \{T_n > t\} \\ Z_t^{(0)} \frac{Z_{T_n}^{opt}}{Z_{T_n}^{(0)}} & \text{on} & A_n \cap \{T_n \leq t\} \end{array} \tag{3.5.23}$$

for any $\mathcal{F}_{T_n}$-measurable set A_n. Because the martingale $Z_\infty^{(0)}$ is strictly positive, it follows that the sequence $Z_{T_n}^{(0)}$ is uniformly bounded from below away from 0, almost surely. Therefore the $\mathcal{F}_{T_n}$-measurable random variables

$$c_n := \hat{\lambda}\frac{Z_{T_n}^{opt}}{Z_{T_n}^{(0)}}$$

satisfy $c_n \to 0$ a.s. Because $Z_\infty^{(0)} \in L^{\gamma+1}$, it follows by Proposition 3.5.10 that the relative risk

$$\frac{E\left[\Phi(c_n Z_\infty^{(0)})|\mathcal{F}_{T_n}\right]}{\Phi(a_n c_n)}$$

remains bounded as $n \to \infty$, for every sequence $a_n \in \mathcal{F}_{T_n}$ with a sufficiently large lower bound. We may therefore choose a sequence which is also bounded from above.

On the other hand, the assumption that $\Phi(x)$ has bounded relative risk aversion implies that the assumptions of Proposition 3.5.11 are satisfied, and the relative risk converges to ∞ as $n \to \infty$ on the set $\{Z_T^{opt} = 0\}$, i.e.

$$\frac{E\left[\Phi(\hat{\lambda} Z_\infty^{opt})|\mathcal{F}_{T_n}\right]}{\Phi(a_n c_n)} \to \infty$$

with the same choice of a_n and c_n as above, because $\frac{a_n}{Z_{T_n}^{(0)}}$ is almost surely bounded from above. For almost every $\omega \in \{Z_T^{opt} = 0\}$ we may therefore find an $N \in \mathbb{N}$ with

$$E[\Phi(\hat{\lambda} Z_\infty^{opt})|\mathcal{F}_{T_n}] > E[\Phi(\hat{\lambda} Z_\infty^{(0)})|\mathcal{F}_{T_n}]$$

for all $n \geq N$. But this means that for n large enough, we have a set A_n which is $\mathcal{F}_{T_n}$-measurable and has strictly positive probability, on which the estimate

$$E[\Phi(\hat{\lambda} Z_\infty^{opt})|\mathcal{F}_{T_n}] > E[\Phi(\hat{\lambda} Z_\infty^{(0)})|\mathcal{F}_{T_n}]$$

holds. Taking this set for equation (3.5.23), we conclude that the martingale measure Z_t defined in (3.5.23) satisfies

$$\hat{\lambda} x_0 + E[\Phi(\hat{\lambda} Z_\infty^{opt})] > \hat{\lambda} x_0 + E[\Phi(\hat{\lambda} Z_\infty)].$$

But this means that Z_∞^{opt} cannot satisfy equation (3.5.19). □

Corollary 3.5.13. *Let the Assumptions 3.5.1, 3.5.3, 3.5.6 and 3.5.7 be satisfied. Furthermore, let the utility function $u(x)$ satisfy the following properties:*

- *u is smooth at the satiation point c.*
- *There is an $\epsilon > 0$ such that $u(x)$ has a relative risk aversion essentially bounded from above for $x \in (c - \epsilon, c)$.*
- *For $x < c$, $u(x)$ has a relative risk aversion essentially bounded from below by $\frac{1}{\gamma}$ for some $\gamma > 0$.*
- *There exists an equivalent local martingale measure $Z_\infty^{(0)}$ which is contained in $L^{\gamma+1}$.*

Then the minimax martingale measure, as defined in Definition 3.2.1, is equivalent.

Proof. Consider the generalized Young function $\Phi^*(x)$ defined in equation (3.5.17). By Assumption 3.5.3 and the assumptions of the corollary it follows that $\Phi^*(x)$ and $\Phi(y)$ are smooth at 0, and there exists an $\epsilon > 0$ such that $\Phi^*(x)$ is essentially bounded from above on $]0 = \Phi'(0), \Phi_l'(\epsilon)]$. From Chapter 2, it is then obvious that its conjugate function $\Phi(y)$ has a relative risk aversion essentially bounded from below on $]0, \epsilon[$.

Furthermore, by the assumptions of the corollary, $\Phi^*(x)$ has a relative risk aversion which essentially bounded from below by $\frac{1}{\gamma}$. From Chapter 2, it follows that $\Phi(y)$ has a relative risk aversion essentially bounded from above by γ. It follows that the assumptions of Theorem 3.5.12 are satisfied.

By Lemma 3.5.8, the minimax martingale measure satisfies the estimate (3.5.19). It follows from Theorem 3.5.12 that this measure must be equivalent. □

3.6 Counterexamples

The aim of this section is to provide examples for what may happen if the assumptions on the relative risk aversion are not satisfied.

3.6.1 An L^Φ-integrable random variable for which the relative risk does not remain bounded

We consider the following generalized Young function:

$$\Phi(x) := \begin{array}{ll} x^2 & \text{if} \quad |x| \leq 1, \\ 2|x| - 1 & \text{if} \quad |x| > 1. \end{array} \tag{3.6.24}$$

This is a Young function which even has a continuous derivative. The relative risk aversion as $x \to 0$ is 1. As $x \to \infty$, the function behaves as a linear function, and therefore every integrable random variable is contained in L^Φ. The relative risk aversion is therefore uniformly bounded from above by 1, and from below by 1 as $x \to 0$. Consider now the random variable

$$X := \frac{1}{U^{\frac{2}{3}}} - 1, \tag{3.6.25}$$

where U is a uniformly distributed random variable. Using a Brownian filtration, this random variable may be generated for example by

$$U = \tilde{\Phi}(W_n - W_{n-1}),$$

where $\tilde{\Phi}$ is the cumulative standard normal distribution, n is an integer (indeed it may be any real number), and W_t a Wiener process.

We have that $X > 0$ with probability 1 and

$$E[X] = \int_0^1 \frac{du}{u^{-\frac{2}{3}}} - 1 = 3u^{\frac{1}{3}}|_{u=1} - 1 = 2,$$

and therefore X is integrable. But looking at the second moment, we have

$$E[X^2] = \int_0^1 \left(u^{-\frac{2}{3}} - 1\right)^2 du = \int_0^1 u^{-\frac{4}{3}} du - 2\int_0^1 u^{-\frac{2}{3}} du + 1 = -3u^{-\frac{1}{3}}|_0^1 - 5 = \infty.$$

It follows that X is not square-integrable and therefore does not satisfy the assumption of Proposition 3.5.10.

Let now $c_n := \frac{1}{n}$ be a sequence converging to 0. Then we have

$$E[\Phi(c_n X)] = E[\frac{1}{n^2} X^2 1_{X \le n}] + E[(\frac{2X}{n} - 1)1_{X>n}].$$

To evaluate this expression, we recognize that $\{X \le n\} = \{U \ge \frac{1}{(n+1)^{\frac{3}{2}}}\}$ and thus with $u_0 := \frac{1}{(n+1)^{\frac{3}{2}}}$,

$$\begin{aligned} E[\Phi(\tfrac{1}{n}X)] &= \tfrac{1}{n^2}\int_{u_0}^1 \left(u^{-\frac{2}{3}} - 1\right)^2 du + \int_0^{u_0} \left(\tfrac{2}{n}(u^{-\frac{2}{3}} - 1) - 1\right) du \\ &= \tfrac{1}{n^2}\left(3(u_0^{-\frac{1}{3}} - 1) - 6(1 - u_0^{\frac{1}{3}}) + 1 - u_0\right) \\ &\quad + \tfrac{1}{n}\left(6u_0^{\frac{1}{3}} - 2u_0\right) - u_0 \\ &= \tfrac{1}{n^2}\left(3\sqrt{n+1} + \tfrac{6}{\sqrt{n+1}} - \tfrac{1}{(n+1)^{\frac{3}{2}}} + 4\right) \\ &\quad + \tfrac{1}{n}\left(\tfrac{6}{\sqrt{n+1}} - \tfrac{2}{(n+1)^{\frac{3}{2}}}\right) - \tfrac{1}{(n+1)^{\frac{3}{2}}}. \end{aligned}$$

The dominating terms behave as $\frac{8}{(n+1)^{\frac{3}{2}}}$, and therefore, as $n \to \infty$, the relative risk satisfies

$$\frac{E[\Phi(\frac{X}{n})]}{\Phi(\frac{1}{n})} \sim 8\frac{n^2}{(n+1)^{\frac{3}{2}}} \to \infty.$$

This example shows that, if X is not contained in $L^{\gamma+1}$ where γ is the essential upper bound of the relative risk aversion of Φ, the statement of Proposition 3.5.10 is not necessarily true.

3.6.2 A uniformly integrable martingale for which the relative risk does not converge to infinity

We start with the following discrete-time martingale.

$$Z_n := \begin{array}{lll} 1 & \text{if} & n = 1 \\ Z_{n-1} & \text{if} & Z_{n-1} \ge Z_{n-2} \text{ and } n > 2 \\ C_n & \text{otherwise} & \text{with probability } p_n = \frac{1}{n^2} \\ \frac{1}{n^4} & \text{otherwise} & \text{with probability } (1 - p_n) = \frac{n^2-1}{n^2} \end{array} \qquad (3.6.26)$$

and C_n chosen in such a way that the process is indeed a martingale. Obviously this discrete-time martingale is bounded, nonnegative, and converges to 0 as $n \to \infty$ with nonzero probability. To see this, not that

$$\ln\left(\prod_n \frac{n^2-1}{n^2}\right) = \sum_n \ln\left(1 - \frac{1}{n^2}\right) \sim \sum_n -\frac{1}{n^2} > -\infty,$$

which shows that the product of the probabilities that this martingale is decreasing converges to a number strictly larger than 0.

With an underlying Brownian motion, we define

$$p_n := P\left[W_{n^4} - W_{(n-1)^4} \leq a_n\right] \tag{3.6.27}$$

where a_n is chosen in a way that the probabilities fit. With

$$Z_\infty := \lim_{n\to\infty} Z_n$$

the process

$$X_t := E[Z_\infty|\mathcal{F}_t] \tag{3.6.28}$$

with the filtration generated by the Brownian motion W_t therefore defines a bounded continuous nonnegative martingale which converges to 0 on a set with nonzero probability.

For $t = n^4$, we furthermore have that $X_t = Z_n$. If T_n is the announcing sequence of stopping times, i.e. $X_{T_n} = \frac{1}{n}$, we would like to show that the supremum

$$y_n := \sup_\omega\{X_\infty(\omega) : X_{T_n}(\omega) = \frac{1}{n}\}$$

converges to 0 with order $\sqrt{X_{T_n}}$. First of all, this is true if $n = k^4$ for a $k \in \mathbb{N}$, because then, by the definition (3.6.26) of the martingale $X_{T_n} = Z_k$,

$$C_k = \frac{k^2}{(k-1)^4} - \frac{k^2-1}{k^4} \leq \frac{K}{k^2}$$

for a constant $K > 0$. But also if $n \in \left](k-1)^4, k^4\right]$, by construction the maximum that the random variable X_∞ can attain must be bounded by C_{k-1}. Therefore, for $n \in \left((k-1)^4, k^4\right]$,

$$C_n \leq C_{k-1} \leq \frac{K}{(k-1)^2} \leq \frac{K_2}{k^2} \leq \frac{K_2}{\sqrt{n}} = K_2\sqrt{X_{T_n}},$$

where K_2 is another constant. This shows that y_n converges to 0 with order $\sqrt{X_{T_n}}$.

Now we take a dual utility function ϕ which has a relative risk aversion which is not essentially bounded from below, see Chapter 2, namely

$$\phi(x) := \frac{x}{\ln(\frac{1}{x})}$$

for sufficiently small x. We want to calculate the fraction

$$\frac{E[\phi(Z_\infty)|\mathcal{F}_{T_n}]}{\phi(Z_{T_n})}.$$

On the set where $Z_T = 0$, conditional on $\mathcal{F}_{T_n}$, we have $Z_{T_n} = \frac{1}{n}$, and the supremum the random variable can attain if $Z_{T_n} = \frac{1}{n}$ is $K_2 \frac{1}{\sqrt{n}}$. We therefore have

$$E\left[\frac{Z_\infty}{\ln(\frac{1}{Z_\infty})}|\mathcal{F}_{T_n}\right] \leq E\left[\frac{Z_\infty}{\ln(\frac{\sqrt{n}}{K_2})}|\mathcal{F}_{T_n}\right] = \frac{1}{n\ln(\frac{\sqrt{n}}{K_2})}. \tag{3.6.29}$$

On the other hand,

$$\phi(Z_{T_n}) = \frac{1}{n\ln(n)},$$

and the relative risk becomes

$$\frac{E[\phi(Z_\infty)|\mathcal{F}_{T_n}]}{\phi(Z_{T_n})} \leq \frac{\ln n}{\ln(\frac{\sqrt{n}}{K_2})} = \frac{\ln n}{0.5\ln n - \ln K_2},$$

which is obviously bounded as $n \to \infty$. We therefore have an example which shows that if the relative risk aversion of ϕ is not essentially bounded from below away from 0, the conclusion of Proposition 3.5.11 does not necessarily hold, and the relative risk does not need to converge to ∞.

3.6.3 A continuous market with a non-equivalent q-optimal martingale measure

Let W^1 and W^2 be two independent Brownian motions, and let the stock price process until time $t = 1$ satisfy

$$dS = dW^1. \tag{3.6.30}$$

The filtration is generated by W^1 and W^2 until time $t = 1$. It is clear that $Z_t = 1$ is the density process of an equivalent martingale measure for $t \leq 1$ and that the market admits an absolutely continuous martingale measure which is not equivalent. Precisely, there exists a density process Z_t^{abs} of an absolutely continuous martingale measure as well as a set A with strictly positive probability and with $Z_1^{abs}(\omega) = 0$ for $\omega \in A$. We may even choose Z_1^{abs} in such a way that it is bounded.

For $1 \leq t \leq 2$, the price process is constructed in the following way. Let $F(x)$ be the cumulative distribution function of a strictly positive random variable which is integrable but has bad integrability properties in the sense that it is not, say, p-integrable. Let ψ be its inverse, and

$$X := \psi(\Phi(W_2^2 - W_1^2)). \tag{3.6.31}$$

Furthermore, the filtration for $t \geq 1$ is now only the one generated by W_1^1 and W_t^2, i.e. for $t > 1$, the first Brownian motion does not play a role any more. It is clear that X is independent of $\mathcal{F}_1$, and that X is integrable but not p-integrable. Because $A \in \mathcal{F}_1$, we also have that $X1_A$ is integrable but not p-integrable,

$$E[(X1_A)^p] = E[E[X^p|\mathcal{F}_1]1_A] = E[X^p]E[1_A] = \infty.$$

Because $X1_A$ is integrable, we may choose it in a way that $E[X] = 1$. Now we define for $1 \leq t \leq 2$ the martingale

$$X_t := E[X1_A|\mathcal{F}_t] + 1_{A^c}. \tag{3.6.32}$$

It is clear that X_t is a strictly positive uniformly integrable martingale for all $1 \leq t \leq 2$. We may therefore take the stochastic logarithm of X_t and define the process $\lambda(t)$ in such a way that

$$X_t = \mathcal{E}\left(-\int_1^t \lambda(s)dW_s^2\right)$$

where we also applied the martingale representation theorem.

Now we define the stock price process for $1 \leq t \leq 2$ by

$$dS := \lambda(t)dt + dW^2. \tag{3.6.33}$$

With this construction, we have the following properties:

1. For any choice of the martingale measure Z_1 for $t \leq 1$, we have that the measure $Z := Z_1X_2$ is a uniformly integrable martingale measure.

2. The martingale measure $\hat{Z} := Z_1^{abs}X_2$ is not equivalent but bounded and therefore contained in L^q for any $q \geq 1$.

3. All equivalent local martingale measures in this market are of the form $Z = Z_1X_2$.

4. The market does not admit an equivalent martingale measure which is p-integrable.

These properties may be proved as follows. Because $Z_1 \in \mathcal{F}_1$ and X is independent of $\mathcal{F}_1$, we have for $t \geq 1$

$$E[Z|\mathcal{F}_t] = Z_1\left(E[X1_A + 1_{A^c}|\mathcal{F}_t]\right) = Z_1X_t,$$

and for $t \leq 1$ we have

$$E[Z|\mathcal{F}_t] = E[Z_1E[X_2|\mathcal{F}_1]|\mathcal{F}_t] = E[X_2]E[Z_1|\mathcal{F}_t] = Z_t,$$

and therefore Z_t is a martingale. For $t \leq 1$, the construction of Z_t already guarantees that Z_t is a martingale measure. For $1 \leq t \leq 2$, we have

$$dZ_t = Z_1 dX_t = -\lambda(t) Z_1 X_t dW_t^2 = -\lambda(t) Z_t dW_t^2,$$

and thus

$$d(SZ) = \lambda(t) Z_t dt - \lambda(t) Z_t d\langle W^2, W^2 \rangle_t + \text{loc. martingale},$$

which is obviously a local martingale.

From the above proof, it is clear that $\hat{Z}$ is a martingale measure. Because $\hat{Z} = 0$ on A and $P[A] > 0$, it is also clear that this martingale measure is not equivalent. By construction, Z_1^{abs} is bounded. On A^c, $X_2 = 1$, and therefore $\hat{Z}$ remains bounded on this subset. But on A, $\hat{Z}$ is 0 and therefore bounded too. It follows that $\hat{Z}$ is bounded.

Let now be Z any absolutely continuous local martingale measure. Because the density process is a martingale, we must have $E[Z|\mathcal{F}_1] = Z_1$. For the economy up to time $t = 1$, S must be a local martingale under the measure with density Z_1, and therefore Z_1 must be one of the martingale measures chosen in the economy for $t \leq 1$.

For $t \geq 1$, we will follow a market completeness argument. First note that X_2 is a strictly positive integrable random variable with expectation 1. Therefore, the measure Q defined by $\frac{dQ}{dP} = X_2$ is equivalent. By construction, the density process Z_t follows the stochastic differential equation

$$dZ = -\lambda(t) Z_t dW_t^2$$

with $\lambda(t) = 0$ for $t < 1$ and $\lambda(t)$ as above for $t \geq 1$. By the Girsanov theorem, the process $\tilde{W}_t^2$ defined by

$$d\tilde{W}^2 := dW^2 + \lambda dt$$

is a Brownian motion under Q. We therefore have that for $t \geq 1$, S follows under Q the dynamics

$$dS = \lambda(t)dt + dW_t^2 = \lambda(t)dt + (d\tilde{W}_t^2 - \lambda(2)dt) = d\tilde{W}_t^2.$$

Let now $\tilde{Q}$ be any equivalent local martingale measure, and let $\tilde{Z} := \frac{d\tilde{Q}}{dQ}$ be its density. Then the process $\tilde{Z}_t$ is a martingale under Q, and by the martingale representation theorem it can be represented for $t \geq 1$ by

$$\tilde{Z}_t = \tilde{Z}_1 + \int_0^t H_s d\tilde{W}_s^2 = \tilde{Z}_1 + \int_0^t H_s dS$$

for a predictable process H_s. Because $\tilde{Q}$ is also a local martingale measure, $\tilde{Z}S$ must be a local martingale under Q. This implies that the quadratic variation of $\tilde{Z}S$ must vanish, because both $\tilde{Z}$ and S are martingales under Q. It follows for $t \geq 1$ that

$$0 = d\langle \tilde{Z}, S\rangle = d\langle \int HdS, S\rangle = Hd\langle S\rangle = Hd\langle W^2\rangle = Hdt,$$

and therefore $H = 0$. Consequently, any density process of an equivalent local martingale measure is $\frac{d\tilde{Q}}{dQ} = Z_1$, and from this we conclude that

$$\frac{d\tilde{Q}}{dP} = \frac{d\tilde{Q}}{dQ}\frac{dQ}{dP} = Z_1 X_2.$$

Let Z now be an equivalent martingale measure. It follows that $Z_1 > 0$, and therefore there exists a set $A_n \subset A$, $A_n \in \mathcal{F}_t$, with $P[A_n] > 0$ and on which $Z_1 \geq \frac{1}{n}$. Because $A_n \in \mathcal{F}_1$, it is independent of X. Because $Z = Z_1 X_2$, we have

$$E[Z^p] = E\left[Z_1^p E\left[(1_A X + 1_{A^c})^p|\mathcal{F}_1\right]\right] \geq \frac{1}{n^p} P[A_n]E[X^p] = \infty.$$

This completes the proof of the properties 1 to 4.

The financial market defined before shows that, if the assumption of the existence of a square integrable local martingale measure in Theorem 1.3 in Delbaen and Schachermayer [25] is dropped, the variance-optimal martingale measure does not need to be equivalent. This can be seen as follows. With the example before, it is clear that there is an absolutely continuous local martingale measure which is bounded and therefore square-integrable, but for which there does not exist any equivalent local martingale measure which is p-integrable. With the choice $p = 2$ (and therefore the appropriate choice of the distribution function $F(x)$), no equivalent local martingale measure is square-integrable. It follows that an equivalent local martingale measure cannot be variance-optimal.

3.7 Application: $q-$optimal measures

From section 3.5, it follows that q-optimal local martingale measures, i.e. martingale measures $\hat{Q}$ which minimize the expression

$$E\left[\left(\frac{dQ}{dP}\right)^q\right] \tag{3.7.34}$$

for $q > 1$ over all absolutely continuous local martingale measures Q, are always equivalent, provided there exists an equivalent local martingale measure which is bounded in L^q. By Bellini and Frittelli [7], this q-optimal measure always exists if $q > 1$. Because the function $\Phi(x) = x^q$ has constant relative risk-aversion of $q - 1$

which is obviously bounded from below away from 0, the assumptions of Theorem 3.5.12 are clearly satisfied.

On the other hand, section 3.6.3 shows how to construct a financial market for which there does not exist an equivalent local martingale measure which is contained in L^q, and for which the q-optimal measure is only absolutely continuous and not equivalent.

From section 3.2, it follows that if there exists an equivalent local martingale measure for which the expectation (3.7.34) remains bounded, for $q < 1$ that the q-optimal local martingale measure is equivalent provided it exists. But by Bellini and Frittelli [7], this existence is not guaranteed any more.

3.8 Conclusion

In this chapter, we have shown that the minimax martingale measure in the sense of Bellini and Frittelli [7] is equivalent to the objective probability measure for satiated utility functions, under some conditions on the utility function as well as on the existence of an equivalent local martingale measure which has sufficiently strong integrability properties, if the filtration is continuous. Whereas for the case with strictly increasing utility functions equivalence has essentially been shown already in Bouchard et al. [14], the main contribution of this chapter is to prove equivalence for the alternative situation with a utility function that has a maximum point. This situation has before only been treated in the specific case of the variance-optimal martingale measure in Delbaen and Schachermayer [25]. In our chapter, we essentially use the same method as in [25] for proving a generalization of this result. Furthermore, we provide an example which shows that the condition of the existence of an equivalent local martingale measure which is square-integrable cannot be dropped without possibly additional assumptions on the financial market.

For further research, one could try to find a sharper distinction whether or not the minimax martingale measure is equivalent for situations where the relative risk either remains bounded or converges to infinity for the absolutely continuous as well as for the equivalent local martingale measure. Furthermore, for finding counterexamples, we had to assume quite specific market situations, which are different from the models that are normally used. It may therefore be advantageous to find conditions on the market rather than on the utility function which guarantee that the minimax martingale measure is equivalent.

One main question remains, namely if we really need the stronger condition of the existence of an equivalent local martingale measure which is contained in $L^{\gamma+1}$, where γ is the upper bound of the essential relative risk aversion, rather than only the weaker one, namely the existence of an equivalent local martingale measure which is contained in L^Φ. From the counterexamples, it becomes clear that we will not be able to prove the stronger result using this method of proof. Combining the counterexamples, it would even be possible to construct a situation where the relative risk of every equivalent local martingale measure tends to infinity faster than the one for the absolutely continuous one on the set where the absolutely continuous local martingale measure tends to zero. But the question then still is what happens on the set where the absolutely continuous local martingale measure does not converge to zero.

A further issue would be, as mentioned in the introduction, to construct a unifying framework about the conditions for equivalence, which includes all cases which here are treated separately. Because the result is not true in general for satiated utility functions unless the filtration is continuous, it may be interesting how such a condition, or perhaps a weaker one, fits into the general framework.

Finally, we had, as in Bellini and Frittelli [7], always the assumption of the existence of a risk-free asset, or equivalently, that the investor optimizes his terminal wealth by discounting everything by a numéraire. If this assumption is dropped, the optimal portfolio may more easily hit the maximum point of the utility function, which mostly implies by the duality results that the dual minimizer is zero with nonzero probability.

3.A Appendix: Conditional expectation for nonnegative processes

The aim of this appendix is to prove a statement on the existence of a left-continuous as well as monotonically increasing version of the conditional expectation for nonnegative random variables, which is needed in order to prove the existence of the conditional Luxemburg norm. As soon as we assume integrability, the statement is a consequence of the standard result for supermartingales. One may expect that this also holds in the case of nonnegative random variables. Here we give the proofs for completeness.

We recall the definition of the conditional expectation for nonnegative random variables from Shiryaev [82].

Definition 3.A.1. [Conditional expectation] Let $X \geq 0$ be a random variable which may attain the value ∞, and let $\mathcal{G}$ be a sub-sigma-algebra. Then a random variable $Y : \Omega \to [0, \infty]$ is said to be a version of the conditional expectation of X conditionally upon $\mathcal{G}$, denoted by $E[X|\mathcal{G}]$, if

1. Y is nonnegative
2. Y is $\mathcal{G}$-measurable
3. For every subset $G \in \mathcal{G}$ we have

$$\int_G X dP = \int_G Y dP \tag{3.A.35}$$

where equality in $[0, \infty]$ means that either both are finite and equal or both are infinite.

It has been stated in [82] that all rules about the conditional expectation with integrability condition can also been used for the nonnegative version of it.

With the existence of the conditional expectation for nonnegative random variables, we can define, for a nondecreasing, nonnegative, and left-continuous process X_t, a process Y_t by

$$Y_t := E[X_t|\mathcal{G}]. \tag{3.A.36}$$

We now show that Y_t has a nonnegative, nondecreasing, left-continuous modification. To this end, we would like to proceed as in Karatzas and Shreve [54], but we cannot directly apply their results, because for these results integrability has been assumed. For doing so, we first need a new upcrossing lemma.

Lemma 3.A.2. *Let X_t be a nonnegative and nondecreasing process, $\mathcal{G}$ a sub-sigma-algebra, and Y_t as in (3.A.36). Then the following holds:*

1. *For any increasing sequence t_n, the process Y_{t_n} is almost surely nondecreasing.*
2. *The process Y_t is almost surely nondecreasing for $t \in \mathbb{Q}$.*
3. *For any $a < b$, the amount of upcrossings $U_N[a, b]$ is almost surely bounded by 1.*

Proof. Let $t_1 < t_2$. Then, by the monotonicity of the conditional expectation, $E[X_{t_1}|\mathcal{G}] \leq E[X_{t_2}|\mathcal{G}]$ almost surely, and therefore $P[Y_{t_1} > Y_{t_2}] = 0$. It follows that

$$P[\exists t_n < t_m : Y_{t_n} > Y_{t_m}] = P\left[\cup_{n<m}\{Y_{t_n} > Y_{t_m}\}\right] = 0,$$

and therefore almost every process Y_{t_n} is nondecreasing, which shows item 1.

For item 2, define the set

$$A := \{\omega \in \Omega : \exists t_1 < t_2 \in \mathbb{Q} : Y_{t_1} > Y_{t_2}\}.$$

Then, if we define for fixed $t_1 < t_2$ the sets $A_{t_1,t_2} := \{Y_{t_1} > Y_{t_2}\}$ which all have probability 0, A is a countable union of A_{t_1,t_2}. Therefore A has probability 0.

Finally, note that Y_{t_n} is a nonnegative nondecreasing process, and therefore for any $a < b$, we have

$$U_N[a, b] \leq 1.$$

This proves item 3. □

From Lemma 3.A.2 it follows also that $U_\infty[a, b] \leq 1$ almost surely, and therefore, for any $a < b$,

$$E[U_\infty[a, b]] \leq 1 < \infty.$$

We now proceed in a similar way as in Karatzas and Shreve [54]. First we define from Y_t a process $\tilde{Y}_t$ which is nonnegative, nondecreasing, and left-continuous, and subsequently we prove that this process is a modification of Y_t.

Proposition 3.A.3. *For any version of the process Y_t defined above, define the process*

$$\tilde{Y}_t := \sup_{s<t; s\in\mathbb{Q}} Y_s. \tag{3.A.37}$$

Then there exists a subset $\Omega^* \subset \Omega$ with $P[\Omega^*] = 1$, such that for all $\omega \in \Omega^*$ the following is true:

1. *$\tilde{Y}_t(\omega)$ is nonnegative for all t,*
2. *$\tilde{Y}_t(\omega)$ is nondecreasing,*
3. *$\tilde{Y}_t(\omega)$ is left-continuous,*
4. *$\tilde{Y}_t$ is a modification of Y_t.*

Proof. We define $\Omega^* := \Omega \setminus A$, with the set A from the proof of the previous Lemma. It follows that $P[\Omega^*] = 1$. Nonnegativity is due to the definition of the conditional expectation for nonnegative random variables. That $\tilde{Y}_t$ is nondecreasing follows from the definition of the supremum.

We now show the left-continuity. Let $t_n < t$ be any sequence converging monotonically to t. Because $\tilde{Y}_t$ is nondecreasing it follows that $\tilde{Y}_{t_n}$ is a nondecreasing sequence, bounded by $\tilde{Y}_t$, or $\tilde{Y}_t = \infty$. Let the limit $\lim_{n\to\infty} \tilde{Y}_{t_n}$ be strictly smaller than $\tilde{Y}_t$. Then there exists a sequence $s_m < t, s_m \in \mathbb{Q}$, such that $Y_{s_m} > \lim_{n\to\infty} \tilde{Y}_{t_n}$ for all m and n. Because $t_n \to t$, we may choose a subsequence $t_{n_m} =: t_m$ such that $t_m > s_m$, and a sequence $q_m \in \mathbb{Q}$ with $t_m > q_m > s_m$. By the definition of the supremum, it follows that

$$Y_{q_m} \leq \tilde{Y}_{t_m} < Y_{s_m}.$$

Because $q_m > s_m$, this realization is not nondecreasing on $\mathbb{Q}$, and can therefore not be an element of Ω^*. It follows that $\tilde{Y}_t$ is left-continuous. That $\tilde{Y}_t$ is a modification of Y_t is shown as follows. By definition, there exists a sequence $t_n \in \mathbb{Q}$ such that $\tilde{Y}_t = \lim_{n\to\infty} Y_{t_n}$, where $t_n < t$. By the fact that Y_t is nondecreasing on $\mathbb{Q}$ (Lemma 3.A.2), we may choose this sequence in such a way that $t_n \to t$. Furthermore, we may choose a subsequence of t_n which is increasing. Then Y_{t_n} is nondecreasing. Therefore, by the monotone convergence theorem, for $G \in \mathcal{G}$

$$\int_G \tilde{Y}_t dP = \lim_{n\to\infty} \int_G Y_{t_n} dP = \lim_{n\to\infty} \int_G X_{t_n} dP = \int_G X_t dP,$$

where the second equality follows from the fact that Y_{t_n} is a version of the conditional expectation of X_{t_n}, and the last one by the fact that X_t is left-continuous and again by the monotone convergence theorem. □

Proposition 3.A.4. *Let the process X_t be continuous at a point t_0, and integrable for a point $t_1 > t_0$. Then $\tilde{Y}_{t_0} = \hat{Y}_{t_0}$ almost surely, and $\tilde{Y}_t$ is continuous at t_0 as well, where*

$$\hat{Y}_t := \inf_{s>t; s\in\mathbb{Q}} Y_s.$$

Proof. With the same arguments as in Proposition 3.A.3, it follows that $\hat{Y}_t$ is nonnegative, nondecreasing and right-continuous at t_0, for all $\omega \in \Omega^*$. Furthermore, for every $s, q \in \mathbb{Q}$ with $s < t < q$, we have $Y_s \le Y_q$ on Ω^*. Taking the supremum on the left hand side and the infimum on the right hand side yields

$$\tilde{Y}_t \le \hat{Y}_t \ \forall \omega \in \Omega^*, \ \forall\, t \ge 0.$$

Because of the right-continuity of $\hat{Y}_t$ at t_0 and the fact that $\tilde{Y}_t$ is nondecreasing, for any sequence $t_n \to t_0$, $t_n > t_0$, we conclude that

$$\tilde{Y}_{t_0} \le \tilde{Y}_{t_n} \le \hat{Y}_{t_n} \to \hat{Y}_{t_0}$$

for all $\omega \in \Omega^*$. If there is a $t > t_0$ for which X_t is integrable, then, by the dominated convergence theorem, $\hat{Y}_{t_0}$ is a version of the conditional expectation as well, and therefore $\hat{Y}_{t_0} = \tilde{Y}_{t_0}$ almost surely. The result follows. □

Corollary 3.A.5. *Let the process X_t be left-continuous, integrable at $t = t_0 > 0$, and continuous at 0 with $X_0 = 0$. Then the process $\tilde{Y}_t$ converges to 0 almost surely as $t \to 0$.*

Proof. By the dominated convergence theorem, setting $Y_0 = 0$ gives a modification of $\hat{Y}_0$. The result follows by the right-continuity of $\hat{Y}_t$ at $t = 0$. □

Chapter 4

A nonlinear PDE approach for indifference pricing in a stochastic volatility model

4.1 Introduction

Models extending the classical Black-Scholes model become more and more important, in industry as well as in academic research. One of the most popular examples of such an extension is a model with stochastic volatility.

In contrast to the classical Black-Scholes model, a model with stochastic volatility is incomplete with respect to the underlying asset. This first means that one cannot hedge every European option only using the underlying asset, and furthermore, the price of such an option cannot be uniquely determined by only considering the underlying stock.

Usually, apart from the underlying, there are also options available in the market. Therefore, the model would be complete, provided one has the true stochastic model for the option price process, or equivalently, by the pricing PDE, a completely calibrated price of volatility risk process. When this calibration is done, in a stochastic volatility model driven by two Brownian motions, a unique option pricing rule as well as a complete hedge of every contingent claim is given, using the risk-free asset, the underlying stock as well as the option.

However, the general case where the price of volatility risk is an adapted process gives still a rather broad class of option pricing rules, and therefore either has to be parametrized or to be calibrated by a nonparametric approach. In the literature, it is often assumed that this price of volatility risk is constant such as in Clarke and Parrot [20], or proportional to the instantaneous volatility such as in the Heston model [44]. However, this assumption is a strong parametric restriction which should be justified.

In our chapter, we consider a completely different parametric approach for the calculation of the price of volatility risk and the option prices. Our indifference pricing approach is more related to a stochastic control problem. We consider an investor with a certain utility function who aims to maximize the expected utility of his terminal wealth at a given time horizon. Assuming that this investor has only invested in the bond and the underlying stock until now, the maximum price such an investor is willing to pay for investing also in options then is the indifference price, which also is the minimum price he needs for selling such options. In this sense, the indifference price is a decision rule for an investor whether or not to invest in options. Furthermore, if there are no options yet on the market, this price gives a minimum reward an investor requires for being willing to sell such options. In this sense, it is also a pricing tool.

By the duality results, it turns out (see Bellini and Frittelli [7], Monoyios [67] for power utility functions) that the indifference price of an option is obtained by the minimax martingale measure. In stochastic volatility models, this measure is essentially obtained by the price of volatility risk as dual minimizer of the primal stochastic control problem.

The numerical literature about indifference pricing is mainly focused on the case of exponential utility functions, such as in Monoyios [66], Lim [60], or Grasselli and Hurd [38], whereas in [60], also the specific case of stochastic volatility has been treated. For their numerical solutions, either dynamic programming or a duality approach is used.

In this chapter, we assume power utility functions as the class of utility functions under consideration, with the addition of the exponential as well as the logarithmic utility functions. Assuming that the whole market behavior can be expressed by a representative utility function of this class of functions, the minimax martingale measure is given by the so-called q-optimal measure, and the price of volatility risk can be parametrized by q, which essentially parametrizes the power that appears in the utility function. This gives an alternative parametric family for the market price of volatility risk, and therefore an option pricing rule. Comparing the option prices calculated in this way with the market option prices, one obtains a new way to calibrate those prices.

The price of volatility risk obtained by this procedure can then be used as a pricing tool, as well as for hedging. Even in complete markets where enough options are available, the hedging strategy obtained from this procedure is different from the one obtained by the assumption that the price of volatility risk is constant, because it leads to a different stochastic model for the option price process.

If there are no options available in the market, an investor with a certain utility function and a certain time horizon for his investment aims to maximize the expected utility of his terminal wealth, trading only with stock and bond. This is a stochastic control problem. Assuming power utility functions, the primal and dual stochastic control problems have been studied in Monoyios [67]. There it has also been shown how the optimal trading strategy can be derived having solved the dual problem, or equivalently, having found the correct price of volatility risk. In this sense, the determination of the price of volatility risk can be used for obtaining the optimal hedging strategy in an incomplete market.

Looking at the numerical method, our approach is different from the one in Monoyios [66], Lim [60], or Grasselli and Hurd [38]. Indeed, we make use of the fact that the price of volatility risk for the class of q-optimal measures can be obtained by the so-called Hobson representation equation [47]. This is a nonlinear partial differential equation, which can be solved for a given q, for example, by the Finite Difference method as well as by Monte Carlo simulation. In our chapter, we study both methods and give corresponding error estimates. We restrict ourselves to those two methods because they are the most frequently used ones in Computational Finance.

Once the price of volatility risk is obtained, it can be used for option valuation by the pricing PDE. In contrast to many other papers such as Clarke and Parrot [20], the price of volatility risk will be a non-constant function. We study the convergence properties of different recently published methods, and also make an estimate of the impact of the numerical error in the price of volatility risk to the option price. Finally, a rough market calibration is performed using the S&P500 daily implied volatilities.

The chapter is organized as follows. In section 4.2, the model is described, and a theoretical background is given. Section 4.3 solves the Hobson representation equation and compares the different methods. In section 4.4, the pricing PDE is solved for a given price of volatility risk following from a q-optimal measure, and different methods are compared. Section 4.5 makes the calibration, whereas section 4.6 concludes.

4.2 Model description and theoretical background

4.2.1 Stochastic Volatility model

The asset price model we deal with in this chapter is a stochastic volatility model satisfying the stochastic differential equation

$$\begin{aligned} dB &= rBdt \\ dS &= \mu Sdt + \sqrt{Y}SdW_1 \\ dY &= \Theta(\omega - Y)dt + \sqrt{2\gamma\Theta}YdW_2 \end{aligned} \qquad (4.2.1)$$

where r is the risk-free rate, B is the risk-free asset, S is the stock and Y is the squared volatility. The sources of uncertainty W_1 and W_2 may be correlated with a coefficient ρ.

Let $G(S_T, T)$ be a European payoff function. By the fundamental theorem of asset pricing there exists an equivalent probability measure Q such that the pricing rule for the option price P is given by

$$P = e^{-r(T-t)}E^Q\left[G(S_T, T)|\mathcal{F}_t\right]. \qquad (4.2.2)$$

Applying the Girsanov theorem for the measure transformation as well as the Itô rule, the option price must satisfy the pricing PDE

$$\begin{aligned} rP &= \tfrac{\partial P}{\partial t} + rS\tfrac{\partial P}{\partial S} + \tfrac{1}{2}YS^2\tfrac{\partial^2 P}{\partial S^2} + \Theta(\omega - Y)\tfrac{\partial P}{\partial Y} + \gamma\Theta Y^2\tfrac{\partial^2 P}{\partial Y^2} \\ &\quad + \rho\sqrt{2\gamma\Theta}SY^{\frac{3}{2}}\tfrac{\partial^2 P}{\partial S\partial Y} - \sqrt{2\gamma\Theta}Y\left(\rho\tfrac{\mu - r}{\sqrt{Y}} + \bar{\rho}\lambda(S, Y, t)\right)\tfrac{\partial P}{\partial Y}, \end{aligned} \qquad (4.2.3)$$

where λ is the price of volatility risk with respect to the source of risk independent of W_1. This may in general be an adapted process. In the sequel, we will assume that it is a function of (S, Y, t). Furthermore, $\bar{\rho} = \sqrt{1 - \rho^2}$.

4.2.2 Investor preferences and q-optimal measures

In general, in an incomplete market with respect to the underlying stocks, such as a stochastic volatility model, the option price is not unique. This means that the price of volatility risk may be any function of the stock price, the volatility and the time, without violation of the no-arbitrage assumption.

One possible choice of a specific equivalent martingale measure is the so-called minimax measure (see Bellini and Frittelli [7]). This measure is a good choice because an investor who wants to maximize the expected utility of his terminal wealth,

$$E\left[U\left(\alpha \cdot \begin{pmatrix} B \\ S \end{pmatrix}_T\right)\right] \to max,$$

with α a self-financing strategy with initial cost smaller than or equal to the initial wealth x, is indifferent whether or not to invest in options if the option prices are given by the minimax measure (see [7] for details).

It follows from [7] that, if the utility function has the form

$$U(x) = \begin{array}{lll} \frac{\gamma}{1-\gamma}x^\gamma & x > 0 & \text{if} \quad \gamma < 0, \\ \frac{\gamma}{1-\gamma}x^\gamma & x \geq 0 & \text{if} \quad 0 < \gamma < 1, \\ \frac{\gamma}{1-\gamma}(-x)^\gamma & x \leq 0 & \text{if} \quad \gamma > 1, \end{array} \tag{4.2.4}$$

the minimax measure is given by the so-called q-optimal measure, i.e. the measure given by

$$\min_{Q \in \mathcal{M}} E\left[\left(\frac{dQ}{dP}\right)^q\right], \tag{4.2.5}$$

where $q = \frac{\gamma}{\gamma-1}$ and where the minimization is done over all local martingale measures which are absolutely continuous with respect to P. Furthermore, the points with $q = 0$ and $q = 1$ are not real singularities, but they are the minimax measures for the utility functions $u_0(x) = \ln(x)$ and $u_1(x) = -e^{-x}$, and the q-optimal measure is then given by

$$\begin{array}{l} \min_{Q_0 \in \mathcal{M}} E\left[\frac{dQ_0}{dP} \ln\left(\frac{dQ_0}{dP}\right)\right], \\ \min_{Q_1 \in \mathcal{M}} E\left[-\ln\left(\frac{dQ_1}{dP}\right)\right]. \end{array} \tag{4.2.6}$$

4.2.3 Hobson representation equation

From Hobson [47], we have that for q-optimal measures with $q \geq 1$ and for a correlation satisfying $\rho^2 < \frac{1}{q}$, the price of volatility risk can be obtained by

$$\lambda(Y, t) = \sqrt{2\gamma\Theta}Y\frac{\partial g}{\partial Y}, \tag{4.2.7}$$

where the function $g(t, Y)$ satisfies the partial differential equation

$$\begin{array}{rcl} -\frac{\partial g}{\partial t} & = & \frac{q}{2}\frac{(\mu-r)^2}{Y} - q\sqrt{2\gamma\Theta Y}(\mu - r)\rho\frac{\partial g}{\partial Y} - (1 - q\rho^2)\gamma\Theta Y^2\left(\frac{\partial g}{\partial Y}\right)^2 \\ & & +\Theta(\omega - Y)\frac{\partial g}{\partial Y} + \gamma\Theta Y^2\frac{\partial^2 g}{\partial Y^2} \end{array} \tag{4.2.8}$$

with the terminal condition that $g(T, Y) = 0$. In Monoyios [67], is has been shown that this equation holds also true for $q < 1$, without restriction on the correlation.

By the Feynman-Kac representation theorem, one can express the solution to (4.2.8) as

$$\exp\left(-(1 - q\rho^2)g(t, y)\right) = \hat{E}\left[\exp\left(-\frac{q}{2}(1 - q\rho^2)\int_t^T \frac{(\mu - r)^2}{Y_s}ds\right)|Y_t = y\right] \tag{4.2.9}$$

where $\hat{E}$ denotes the expectation under an equivalent martingale measure under which the dynamics of Y is given by

$$dY = \Theta(\omega - Y)dt - \sqrt{2\gamma\Theta}\rho q(\mu - r)\sqrt{Y}dt + \sqrt{2\gamma\Theta}Y d\tilde{W}. \tag{4.2.10}$$

If the correlation is 0, this can be replaced by the original measure, as also stated in Hobson [47]. To avoid confusion, we would like to clarify that this is not an option pricing measure, but only a probability measure for calculating the Feynman-Kac representation.

Equation (4.2.8) can be used for a PDE-based solution of the Hobson representation equation, whereas equation (4.2.9) offers an opportunity for a simulation-based solution. Both approaches will be treated in this chapter.

4.2.4 Monotonicity and calibration

It is a remarkable result from Henderson [42] (see also from the same author [43] for the case with correlation), that if the Sharpe ratio is monotonic in the volatility, it follows that the option price is monotonic in the parameter q. In the case of our model (4.2.1) where the Sharpe ratio is decreasing, this means that the option price is increasing in the parameter q.

If market option prices are available, we can therefore find a unique q such that the q-optimal option price corresponds to all market option prices, provided the model is correct and the representative investor indeed has a power utility function. Through the Hobson representation equation, a calibration of q corresponds to a calibration of the market price of volatility risk $\lambda(t, Y)$. This gives an alternative parametric model for calibrating the market price of volatility risk, instead of the assumption that λ is constant or proportional to the instantaneous volatility which often has been considered (Boswijk [13], Fiorentini et al. [31]).

4.3 Solution of the Hobson representation equation

4.3.1 Solution by simulation

Simulation setup and formulas

Equation (4.2.9) is discretized on the basis of a grid (t_k, y_l), linear in the first variable and logarithmic around the mean-reversion point ω in the second variable. With $\Delta t := \frac{T}{K}$, K the amount of time grid points and T the terminal time in (4.2.9), we have the following grid

$$(t_k, y_l) = (k\Delta t, \omega \exp(\eta_l)) = (k\Delta t, \omega \exp(-\eta_{max} + l\Delta\eta)), \tag{4.3.11}$$

where $\Delta\eta = \frac{\eta_{max}}{L}$ and where $2L$ is the amount of grid points in the volatility direction.

Starting from the point t_k at the value y_l, the process $Y_{kl}(s, \tilde{\omega})$ is simulated by an Euler approximation scheme, i.e.

$$\begin{aligned} Y_{kl}(s+\Delta s, \tilde{\omega}) &= Y_{kl}(s, \tilde{\omega}) + \Theta(\omega - Y_{kl}(s, \tilde{\omega}))\Delta s - \sqrt{2\gamma\Theta}\rho q(\mu - r)\sqrt{Y_{kl}(s, \tilde{\omega})}\Delta s \\ &\quad + \sqrt{2\gamma\Theta Y_{kl}(s, \tilde{\omega})}\sqrt{\Delta s} Z_{s+\Delta s}. \end{aligned} \tag{4.3.12}$$

Here we write $\tilde{\omega}$ for the event in order to distinguish it from the mean reverting point ω. Z denotes a standard normally distributed random variable, and the increments $Z_{s+\Delta s}$ are all independent. The choice of the parameters Δs as well as the amount of simulations are important to control the accuracy of the simulation and will later be considered in more detail.

The numerical integration for the expression under the expectation operator in (4.2.9) is a straightforward computation, yielding

$$\hat{g}(\tilde{\omega}, t_k, y_l) = \exp\left(-\frac{q}{2}(1 - q\rho^2) \sum_{s=t_k}^{T-\Delta s} \frac{2(\mu - r)^2}{Y_{kl}(s, \tilde{\omega}) + Y_{kl}(s+\Delta s, \tilde{\omega})} \Delta s \right). \tag{4.3.13}$$

By taking the expectation, we obtain

$$\exp\left(-(1 - q\rho^2) g(t_k, y_l)\right) = \bar{\hat{g}}(\tilde{\omega}, t_k, y_l) = \frac{1}{N} \sum_{\tilde{\omega}=1}^{N} \hat{g}(\tilde{\omega}, t_k, y_l), \tag{4.3.14}$$

where N is the number of simulation steps.

Numerical differentiation

From (4.2.7), we know that we are not interested in g itself, but rather in its derivative with respect to y. With our logarithmic scale, this means

$$\begin{aligned} \lambda(y_l, t_k) &= \sqrt{2\gamma\Theta} y_l \tfrac{\partial g}{\partial y}(t_k, y_l) = \sqrt{2\gamma\Theta} \tfrac{\partial g}{\partial \eta}(t_k, \eta_l) \\ &\approx \sqrt{2\gamma\Theta} \tfrac{1}{2\Delta\eta} \left(g(t_k, \eta_{l+1}) - g(t_k, \eta_{l-1}) \right), \end{aligned} \tag{4.3.15}$$

which converges of order $(\Delta\eta)^2$ by the Taylor rule.

Error estimate

The sample variance of the random variable $\hat{g}$ in (4.3.14) is estimated by the standard rule

$$V(t_k, y_l) = \frac{1}{N} \sum_{\tilde{\omega}=1}^{N} \left(\hat{g}(\tilde{\omega}) - \bar{\hat{g}} \right)^2. \tag{4.3.16}$$

The random variable $\hat{g}$ is an approximation as well, as we still have a discretization error of order Δs when taking the weak convergence order.

Considering only the simulation error and neglecting the error due to the time discretization Δs, we may calculate the 5% confidence intervals

$$E[\hat{g}(t_k, y_l)] = \bar{\hat{g}}(t_k, y_l) \pm \frac{1.96}{\sqrt{N}} \sqrt{V(t_k, y_l)}$$

and from this an error estimate for the function $g(t_k, y_l)$, namely

$$2e(t_k, y_l) = \frac{1}{1 - q\rho^2} \times \left[-\ln\left(\bar{\hat{g}}(t_k, y_l) - 1.96\sqrt{\frac{V(t_k, y_l)}{N}} \right) + \ln\left(\bar{\hat{g}}(t_k, y_l) + 1.96\sqrt{\frac{V(t_k, y_l)}{N}} \right) \right]. \quad (4.3.17)$$

When performing a numerical differentiation, a problem arises which is described in Glasserman [37], namely that the choice of the grid in (y_l) affects the simulation error: the finer the grid, i.e. the lower the step $\Delta\eta$, the higher the simulation error. On the other hand, one should not choose $\Delta\eta$ too large, because this would cause a large error in the numerical differentiation. In [37], it was shown that the optimal $\Delta\eta$ is

$$\left(\frac{18\sigma^2}{g'''(\eta)^2} \right)^{\frac{1}{6}},$$

where σ^2 may be estimated by two times the sample variance from (4.3.16). Because we do not know g, we will not use this formula, but we will see in the numerical experiments that we should not choose $\Delta\eta$ very small, because otherwise there would not be an acceptable simulation error at a reasonable time. This will be the large advantage of the Finite Difference solution, where we can choose a much finer grid.

The simulation error for the price of volatility risk λ then is, from (4.3.15) and (4.3.17),

$$e_{sim}(t_k, y_l) := \Delta\lambda = \sqrt{2\gamma\Theta} \frac{2e(t_k, y_l)}{\Delta\eta}, \quad (4.3.18)$$

where we indeed can see the inverse proportional influence of the discretization step $\Delta\eta$.

Numerical experiments

There is no obvious way how to choose control variates or importance sampling. We will therefore perform the simulations described above without further transformations.

For comparing the results, a standard parameter set is taken with market parameters that seem reasonable. The parameters μ, ω and Θ have been estimated from

S&P500, 20.10.1982-31.12.2004, daily data. For the risk-free return r, a parameter has been taken which seems reasonable. For γ, we take the value given in Melenberg and Werker [61], who use the same model as we do. One has to interpret those parameters as yearly data. The values are the following:

Parameter	Description	Value
r	Risk-free interest rate	0.0252
μ	Mean return of stock	0.0973
ω	Mean reverting point of volatility process	0.0287
Θ	Mean reverting drift	52.92
γ	Relative diffusion of volatility process	0.229

We first look at the effect of the discretization of the stochastic process (4.2.10), i.e. the time step Δs in (4.3.12). We therefore choose the correlation $\rho = 0$, a space discretization $\Delta\eta = 4.0$, an amount of simulations $N = 200$, and a fixed value of $q = 400$. We will see later that for the case with no correlation, such a large value of q has to be taken in order to perform a successful calibration. We choose different values for the time horizon, namely $t = 1$, $t = 10$, and $t = 50$ days. The results are given in Table 4.1. By the symbol $\pm$, the confidence intervals at a 5% level are indicated, due to formulas (4.3.16), (4.3.17), and (4.3.18). With the variable $Time$, the computational time is meant that was necessary for this simulation in C++ with an Intel Pentium M750.

From Table 4.1, one can see that even with this small amount $N = 200$ of simulations, one has quite small error bounds related to the simulation error, especially for small time horizons t. However, for larger time horizons, the simulation error is still quite large. For large values of y, i.e. larger than the mean reverting point $\ln\omega \approx -9$, the solution is almost independent of the choice of time discretization Δs and lies in the confidence intervals. However, for small values of y, the solution is highly dependent on Δs, and clearly lies out of the confidence intervals. As a consequence, for small values of y, the Euler scheme converges to the wrong solution if Δs is too large. An application of the Milstein scheme instead of equation (4.3.12) gives no substantial improvement. On the other hand, a too small choice of Δs means a very large computation time, even for a small amount of simulations and very large space steps $\Delta\eta$. This is especially a problem for larger time horizons, i.e. for larger values of t. From Table 4.1, we can see that we should at least choose a time step of $\Delta s = 10^{-4}$ in order to get sufficiently accurate results for a range of $\ln y \geq -17$. In this case, for small values of t, there is still a small error at $\ln y = -17$, but on the other hand, for larger time horizons $t \geq 50$, we would not be able to obtain small confidence intervals in a reasonable computation time with a smaller choice of Δs.

Table 4.1: Price of volatility risk $\lambda(t, y)$: Impact of time discretization

$t = 1$				
$\ln y$	$\Delta s = 10^{-3}$	$\Delta s = 10^{-4}$	$\Delta s = 10^{-5}$	$\Delta s = 10^{-6}$
-21	-0.065±0.001	-0.125±0.002	-0.179±0.002	-0.209±0.001
-17	-0.165±0.002	-0.203±0.001	-0.210±0.002	-0.210±0.001
-13	-0.168±0.001	-0.168±0.002	-0.168±0.001	-0.168±0.002
-9	-0.069±0.001	-0.070±0.001	-0.070±0.001	-0.069±0.001
-5	-0.006±0.000	-0.006±0.000	-0.006±0.000	-0.006±0.000
-1	-0.000±0.000	-0.000±0.000	-0.000±0.000	-0.000±0.000
3	-0.000±0.000	-0.000±0.000	-0.000±0.000	-0.000±0.000
Time	0.2 min	1.5 min	16 min	150 min
$t = 10$				
$\ln y$	$\Delta s = 10^{-3}$	$\Delta s = 10^{-4}$	$\Delta s = 10^{-5}$	$\Delta s = 10^{-6}$
-21	-0.067±0.005	-0.125±0.005	-0.176±0.005	-0.207±0.005
-17	-0.168±0.005	-0.205±0.005	-0.212±0.005	-0.210±0.005
-13	-0.208±0.006	-0.208±0.005	-0.209±0.005	-0.210±0.005
-9	-0.158±0.005	-0.159±0.005	-0.158±0.005	-0.159±0.005
-5	-0.061±0.001	-0.061±0.001	-0.062±0.001	-0.061±0.001
-1	-0.005±0.001	-0.006±0.000	-0.006±0.001	-0.006±0.000
3	-0.000±0.000	-0.000±0.000	-0.000±0.000	-0.000±0.000
Time	0.2 min	2 min	16 min	150 min
$t = 50$				
$\ln y$	$\Delta s = 10^{-3}$	$\Delta s = 10^{-4}$	$\Delta s = 10^{-5}$	$\Delta s = 10^{-6}$
-21	-0.076±0.029	-0.143±0.019	-0.177±0.023	-0.212±0.018
-17	-0.157±0.024	-0.218±0.018	-0.210±0.020	-0.220±0.031
-13	-0.206±0.020	-0.204±0.030	-0.217±0.022	-0.219±0.026
-9	-0.220±0.017	-0.211±0.020	-0.202±0.024	-0.200±0.039
-5	-0.201±0.017	-0.199±0.023	-0.206±0.023	-0.196±0.023
-1	-0.163±0.013	-0.165±0.015	-0.173±0.016	-0.174±0.019
3	-0.100±0.007	-0.100±0.007	-0.094±0.007	-0.093±0.007
Time	2 min	19 min	180 min	900 min

For further research, an adaptive time step Δs, where very small values of Δs are taken for small values of Y_t, and larger values if Y_t becomes larger, may give a numerical improvement.

After having chosen an appropriate time discretization $\Delta s = 10^{-4}$, we will move to the question of a reasonable tradeoff between the choice of the space step $\Delta\eta$ and the simulation error which may occur, as well as an appropriate choice of the amount of simulations N. We will solve the problem for two sets of parameters, namely the one chosen above, and one with a strong negative correlation $\rho = -0.75$ with a small value of $q = 1.75$. Because the problem of simulation error mainly arises for larger time horizons, we will choose a time horizon of $t = 90$. On the other hand, for small values of t, it is expected that the discretization error in $\Delta\eta$ is more important. Therefore, we will also look at a smaller time horizon $t = 10$. Tables 4.2.1 and 4.2.2 show the results for no correlation, and Table 4.3.1 the results with correlation.

The reference solutions in Tables 4.2.1 to 4.3.1 were obtained by the Finite Difference solution. As will be seen in the following section, this method gives very accurate results in a reasonable computation time.

From Tables 4.2.1 to 4.3.1, one can first see that the error bounds due to the simulation error are quite large, especially for the larger time horizon $t = 90$, where they may be larger than 10%. On the other hand, even for a large space discretization step of $\Delta\eta = 4$, the correct results without correlation lie in the error bound. As predicted in Glasserman [37], the error bounds become larger with a decrease of $\Delta\eta$. For large time horizons and without correlation, it is therefore better to choose a large space discretization step in order to get lower error bounds at a reasonable computation time. According to [37], as the price of volatility risk λ is approximately constant, it is expected that the optimal discretization step tends to infinity.

This fact changes for smaller time horizons such as $t = 10$, as well as for the case with correlation, where the function λ is not constant any more. In the case of small time horizons, because the computation is faster, the error bounds become around 5% even for a fine space grid, so that the amount of simulations is not as important as it is for larger time horizons. On the other hand, the solution with $\Delta\eta = 4$ sometimes lies clearly out of the error bounds. Further calculations which are not reported here show that this effect is in particular strong in the case with correlation. It follows that one has to consider the discretization error too, and therefore the total error is a sum of the simulation and the discretization errors. The tradeoff between these two sources of error may be to choose $\Delta\eta = 2$. The higher the number of simulations, the smaller $\Delta\eta$ should be.

In the situation with correlation, the solution with $\Delta\eta = 4$ is also out of the error

Table 4.2.1: Impact of space discretization and amount of simulations for $t = 90$, without correlation

$\ln y$	$N = 200$ $\Delta\eta = 4$	$N = 500$ $\Delta\eta = 4$	$N = 500$ $\Delta\eta = 2$	$N = 2000$ $\Delta\eta = 4$	$N = 2000$ $\Delta\eta = 2$	Reference solution
-19	–	–	-0.209±0.040	–	-0.206±0.025	-0.212
-17	-0.203±0.030	-0.210±0.021	-0.251±0.061	-0.203±0.019	-0.218±0.038	-0.212
-15	–	–	-0.193±0.090	–	-0.207±0.035	-0.212
-13	-0.240±0.029	-0.232±0.031	-0.197±0.054	-0.209±0.013	-0.202±0.031	-0.212
-11	–	–	-0.215±0.059	–	-0.208±0.025	-0.212
-9	-0.216±0.065	-0.202±0.045	-0.219±0.067	-0.215±0.013	-0.217±0.026	-0.212
-7	–	–	-0.227±0.069	–	-0.213±0.024	-0.212
-5	-0.176±0.038	-0.204±0.027	-0.194±0.067	-0.215±0.013	-0.222±0.023	-0.212
-3	–	–	-0.186±0.055	–	-0.221±0.030	-0.212
-1	-0.207±0.026	-0.212±0.030	-0.208±0.041	-0.209±0.017	-0.209±0.048	-0.212
1	–	–	-0.186±0.040	–	-0.214±0.039	-0.210
time	15 min	40 min	60 min	130 min	250 min	

Table 4.2.2: Impact of space discretization and amount of simulations for $t = 10$, without correlation

$\ln y$	$N = 500$ $\Delta\eta = 4$	$N = 500$ $\Delta\eta = 2$	$N = 500$ $\Delta\eta = 1$	$N = 2000$ $\Delta\eta = 2$	$N = 2000$ $\Delta\eta = 1$	Reference solution
-19	–	-0.199±0.007	-0.216±0.014	-0.197±0.004	-0.206±0.007	-0.212
-17	-0.205±0.004	-0.210±0.006	-0.211±0.014	-0.213±0.004	-0.214±0.007	-0.212
-15	–	-0.217±0.007	-0.213±0.013	-0.213±0.003	-0.213±0.007	-0.212
-13	-0.208±0.003	-0.216±0.007	-0.209±0.015	-0.208±0.004	-0.215±0.007	-0.211
-11	–	-0.201±0.006	-0.201±0.013	-0.204±0.004	-0.202±0.007	-0.207
-9	-0.159±0.003	-0.175±0.006	-0.187±0.012	-0.175±0.003	-0.180±0.006	-0.184
-7	–	-0.112±0.004	-0.110±0.007	-0.113±0.002	-0.115±0.004	-0.114
-5	-0.062±0.001	-0.044±0.001	-0.039±0.003	-0.046±0.001	-0.039±0.002	-0.037
-3	–	-0.011±0.001	-0.008±0.001	-0.011±0.000	-0.008±0.001	-0.007
-1	-0.006±0.000	-0.002±0.000	-0.001±0.000	-0.002±0.000	-0.001±0.000	-0.001
1	–	-0.000±0.000	-0.000±0.000	-0.000±0.000	-0.000±0.000	-0.000
time	2 min	6 min	13 min	25 min	49 min	

Table 4.3.1: Impact of space discretization and amount of simulations for $t = 90$, with correlation $\rho = -0.75$, results multiplied by 10^3

$\ln y$	$N = 200$ $\Delta\eta = 4$	$N = 500$ $\Delta\eta = 4$	$N = 500$ $\Delta\eta = 2$	$N = 2000$ $\Delta\eta = 4$	$N = 2000$ $\Delta\eta = 2$	Reference solution
-19	–	–	-0.921±0.113	–	-0.864±0.056	-0.927
-17	–	-0.526±0.054	-0.981±0.115	-0.878±0.028	-0.933±0.056	-0.926
-15	–	–	-0.940±0.113	–	-0.917±0.056	-0.925
-13	–	-0.945±0.057	-0.871±0.111	-0.917±0.028	-0.900±0.055	-0.922
-11	–	–	-0.855±0.115	–	-0.912±0.056	-0.915
-9	–	-0.933±0.057	-0.984±0.110	-0.909±0.027	-0.931±0.057	-0.905
-7	–	–	-0.937±0.110	–	-0.891±0.055	-0.895
-5	–	-0.899±0.056	-0.841±0.118	-0.888±0.028	-0.875±0.057	-0.890
-3	–	–	-0.834±0.107	–	-0.901±0.058	-0.887
-1	–	-0.861±0.055	-0.861±0.115	-0.876±0.027	-0.936±0.057	-0.886
1	–	–	-0.858±0.108	–	-0.867±0.056	-0.886

bounds for the larger time horizon $t = 90$, but lies within the error bounds with the choice $\Delta\eta = 2$. Nevertheless the choice $\Delta\eta = 4$ gives a more precise result, but one should consider both the simulation error and the discretization error.

All in all, especially for larger time horizons, the Monte Carlo method requires a lot of computational effort, even for results with a large remaining total error.

4.3.2 Solution by Finite Difference Method

Transformation and boundary conditions

We apply a first transformation of equation (4.2.8) into the logarithmic scale $Y = e^y$ and a change in time $t \mapsto T - t$ to obtain the initial value problem

$$\begin{aligned} \frac{\partial g}{\partial t} &= \frac{q}{2}(\mu - r)^2 e^{-y} - q\sqrt{2\gamma\Theta}\rho(\mu - r)e^{-\frac{y}{2}}\frac{\partial g}{\partial y} - (1 - q\rho^2)\gamma\Theta\left(\frac{\partial g}{\partial y}\right)^2 \\ &\quad + \Theta(\omega e^{-y} - 1 - \gamma)\frac{\partial g}{\partial y} + \gamma\Theta\frac{\partial^2 g}{\partial y^2} \end{aligned} \tag{4.3.19}$$

with the initial condition

$$g(0, y) = 0. \tag{4.3.20}$$

Equation (4.3.19), together with (4.3.20), is a nonlinear initial value problem. For numerical purposes, we need also boundary conditions. Research which we do not describe here indicates that the solution behaves linearly at the boundaries. We will therefore choose linear boundary conditions, which means no diffusion at the boundaries. We will consider a broad range for the choices of the boundaries, denoted by y_{min} and y_{max}.

Localization and discretization

Because the equation is nonlinear, the implementation of any implicit method would require an iterative solution of a system of nonlinear equations. To avoid this problem, we implement an explicit method, and we will show that this method works very well.

For the boundary, we choose in the logarithmic scale a symmetric region around $\ln\omega$, i.e. we solve the differential equation in the spatial region

$$\Omega = (\ln\omega - y_{max}, \ln\omega + y_{max}),$$

where for y_{max} different values are chosen in order to determine the influence of the boundaries.

As stated above, the explicit finite difference method has been used for the time discretization. For the discretization in space, the central difference method has been used for the first as well as the second derivative, with the exception of the boundaries. There, when applying linear conditions, the forward and backward difference method have been used for the first derivative.

Numerical experiments

First we will look at the impact of the localization, i.e. we look at different values of y_{max}. For this we will again choose the case with no correlation and $q = 400$. The case with a correlation of $\rho = -0.75$ and for $q = 1.75$ also exhibits almost no dependency on y_{max} and is not reported here. For the discretization in space, we choose for this analysis a step width of $y = 1.0$. This discretization step will be used for the numerical solution of equation (4.3.19) as well as for the numerical differentiation for obtaining λ by (4.2.7), or in the discretized version by (4.3.15). The results are contained in Table 4.4.

The results in Table 4.4 show that the results are almost independent of the choice of boundaries. On the other hand, the time for the computation strongly increases when y_{max} is large. The reason is that we have to choose a very small time step for the explicit Finite Difference method in order to guarantee stability. Therefore, as long as we are not interested in values of λ far away from the mean reverting point ω, it makes sense to choose y_{max} small. In order to obtain at least a slightly larger region, we choose $y_{max} = 12$ for the further analysis.

Tables 4.5.1 and 4.5.2 show the convergence of the Finite Difference scheme to the solution of (4.3.19) and (4.3.20).

By inspecting Tables 4.5.1 and 4.5.2, we observe that the results are already quite accurate with a spatial discretization of $\Delta y = 0.5$, at least for the case of no correlation. For the case of a strong correlation, the convergence is slower but still acceptable. Between $\Delta y = 0.2$ and $\Delta y = 0.1$, there is almost no difference any more in the case of no correlation, and only a small difference in the case of a strong correlation. A further calculation with an even finer space step showed no difference to the case of $\Delta y = 0.1$ at our 4 significant numbers.

Further numerical experiments which are not reported here have shown that the explicit Finite Difference method remains stable and shows a similar convergence as in Tables 4.5.1 and 4.5.2 for a broad range of parameters, in particular when increasing the value of γ or when taking a positive instead of a negative correlation ρ, but also for larger time horizons.

Because the solutions are much more accurate than the ones obtained by Monte Carlo simulation, one is able to observe their general behavior. First of all, for small values of the volatility, the solution does not change in time, whereas for large values, it decreases monotonically from 0 to a limit value if $t \to \infty$. The convergence to this limit value is fast for small values of y and slower for larger values. Figures 4.1

Table 4.4: Impact of boundary truncation

Without correlation			
$t = 1$			
$\ln y$	$y_{max} = 10$	$y_{max} = 15$	$y_{max} = 20$
-21	–	-0.2119	-0.2119
-17	-0.2116	-0.2116	-0.2116
-13	-0.1963	-0.1963	-0.1978
-9	-0.0441	-0.0441	-0.0478
-5	-0.0011	-0.0011	-0.0013
-1	-0.0000	-0.0000	-0.0000
3	–	-0.0000	-0.0000
$t = 10$			
-21	–	-0.2119	-0.2119
-17	-0.2119	-0.2119	-0.2119
-13	-0.2113	-0.2113	-0.2113
-9	-0.1832	-0.1832	-0.1837
-5	-0.0396	-0.0396	-0.0404
-1	-0.0015	-0.0015	-0.0016
3	–	-0.0000	-0.0000
$t = 90$			
-21	–	-0.2119	-0.2119
-17	-0.2119	-0.2119	-0.2119
-13	-0.2119	-0.2119	-0.2119
-9	-0.2119	-0.2119	-0.2119
-5	-0.2119	-0.2119	-0.2119
-1	-0.2119	-0.2119	-0.2119
3	–	-0.2010	-0.2018
Time	0.05 min	1.5 min	200 min
Δt	1E-4	2E-6	2E-8

Table 4.5.1: Convergence of the numerical solution of the Hobson representation equation, without correlation

$\ln y$	$\Delta y = 1.0$	$\Delta y = 0.5$	$\Delta y = 0.2$	$\Delta y = 0.1$
$t = 1$				
-20	-0.2119	-0.2119	-0.2119	-0.2119
-17	-0.2116	-0.2116	-0.2116	-0.2116
-13	-0.1963	-0.1964	-0.1965	-0.1965
-11	-0.1339	-0.1346	-0.1348	-0.1349
-9	-0.0441	-0.0421	-0.0415	-0.0414
-7	-0.0079	-0.0071	-0.0069	-0.0068
-5	-0.0011	-0.0010	-0.0010	-0.0010
-3	-0.0002	-0.0001	-0.0001	-0.0001
1	-0.0000	-0.0000	-0.0000	-0.0000
$t = 10$				
-20	-0.2119	-0.2119	-0.2119	-0.2119
-17	-0.2119	-0.2119	-0.2119	-0.2119
-13	-0.2113	-0.2113	-0.2113	-0.2113
-11	-0.2074	-0.2073	-0.2073	-0.2073
-9	-0.1832	-0.1840	-0.1842	-0.1842
-7	-0.1114	-0.1131	-0.1137	-0.1138
-5	-0.0396	-0.0377	-0.0371	-0.0370
-3	-0.0090	-0.0074	-0.0069	-0.0068
1	-0.0002	-0.0002	-0.0001	-0.0001
$t = 90$				
-20	-0.2119	-0.2119	-0.2119	-0.2119
-17	-0.2119	-0.2119	-0.2119	-0.2119
-13	-0.2119	-0.2119	-0.2119	-0.2119
-11	-0.2119	-0.2119	-0.2119	-0.2119
-9	-0.2119	-0.2119	-0.2119	-0.2119
-7	-0.2119	-0.2119	-0.2119	-0.2119
-5	-0.2119	-0.2119	-0.2119	-0.2119
-3	-0.2119	-0.2119	-0.2119	-0.2119
1	-0.2111	-0.2103	-0.2101	-0.2101
time	0.1 min	0.2 min	4 min	95 min
Δt	5E-5	2E-5	5E-6	1E-6

Table 4.5.2: Convergence of the numerical solution of the Hobson representation equation, with correlation, results multiplied by 10^3

$\ln y$	$\Delta y = 1.0$	$\Delta y = 0.5$	$\Delta y = 0.2$	$\Delta y = 0.1$
$t = 1$				
-20	-0.9269	-0.9269	-0.9269	-0.9269
-17	-0.9250	-0.9250	-0.9250	-0.9250
-13	-0.8554	-0.8560	-0.8561	-0.8561
-11	-0.5833	-0.5865	-0.5875	-0.5876
-9	-0.1923	-0.1836	-0.1810	-0.1807
-7	-0.0347	-0.0309	-0.0299	-0.0298
-5	-0.0050	-0.0043	-0.0042	-0.0042
-3	-0.0007	-0.0006	-0.0006	-0.0006
1	-0.0000	-0.0000	-0.0000	-0.0000
$t = 10$				
-20	-0.9270	-0.9270	-0.9270	-0.9270
-17	-0.9263	-0.9263	-0.9263	-0.9263
-13	-0.9195	-0.9195	-0.9194	-0.9194
-11	-0.9003	-0.8999	-0.8998	-0.8998
-9	-0.8069	-0.8102	-0.8107	-0.8108
-7	-0.5047	-0.5152	-0.5187	-0.5192
-5	-0.1789	-0.1703	-0.1675	-0.1671
-3	-0.0396	-0.0325	-0.0303	-0.0300
1	-0.0009	-0.0007	-0.0006	-0.0006
$t = 90$				
-20	-0.9270	-0.9270	-0.9270	-0.9270
-17	-0.9263	-0.9263	-0.9263	-0.9263
-13	-0.9215	-0.9215	-0.9215	-0.9215
-11	-0.9146	-0.9146	-0.9146	-0.9146
-9	-0.9046	-0.9047	-0.9046	-0.9046
-7	-0.8954	-0.8954	-0.8954	-0.8954
-5	-0.8899	-0.8899	-0.8899	-0.8899
-3	-0.8874	-0.8873	-0.8873	-0.8873
1	-0.8858	-0.8859	-0.8858	-0.8858
time	0.1 min	0.3 min	5 min	55 min
Δt	5E-5	2E-5	5E-6	1E-6

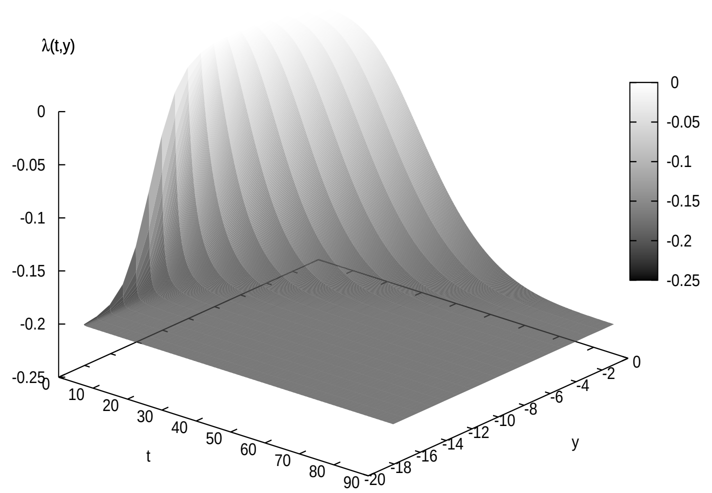


Figure 4.1: Price of volatility risk, without correlation

and 4.2 show the shape of the price of volatility risk $\lambda(t, y)$ as a function of the time horizon t and the logarithm of the squared volatility y, for the case without as well as with correlation.
A remarkable feature of the solution without correlation is that for large time horizons, the price of volatility risk tends to a constant, not only in time, but also in volatility. This justifies, from our approach based on investor preference, the choice of a constant price of volatility risk which has often been made. This fact is no longer valid for smaller time horizons, nor for the situation where the two sources of uncertainty are correlated.

4.3.3 Comparison of the two approaches

Comparing Tables 4.5.1 and 4.5.2 of the Finite Difference solution with Tables 4.2.1 and 4.3.1 of the solution by simulation, one can see that the Finite Difference solution always lies in the error bounds of the simulation solution, provided the space and time discretization steps of the simulation solution are small enough in order to have convergence to the right solution.

However, in the simulation solution, a large computational time is needed while the error bounds are still large, whereas the solution by the explicit Finite Difference

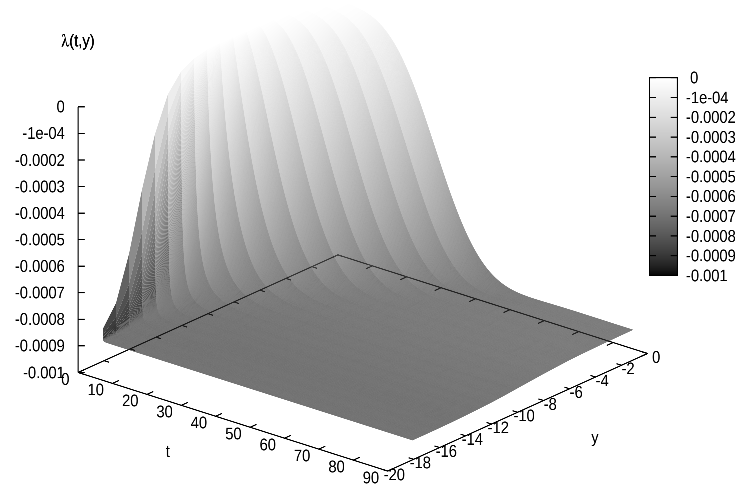


Figure 4.2: Price of volatility risk, with correlation

method gives very accurate results in a reasonable time. Even when setting the space discretization large in order to get fast convergence, the solution is already quite near to the correct solution, as Tables 4.5.1 and 4.5.2 show. These observations hold regardless of whether the correlation is zero or nonzero. However, the convergence is slower in the case of a strong correlation.

One single advantage the solution by simulation has is that in the case of no correlation, by formula (4.3.14), it is not necessary to perform the whole simulation procedure for each value of q separately; one can apply the same simulated random variables for different values of q. Solving the equation by Finite Difference method, one has to carry out the calculation for every q separately. But this advantage disappears when there is a correlation between stock and volatility process, because in this case one has to adjust the volatility process for each q by the right measure, as given by formula (4.2.10).

As already mentioned, a major problem of the Monte Carlo method is that one has to choose the time step Δs in formula (4.3.12) very small in order to get convergence to the right solution for small values of the volatility process Y. An obvious possible improvement of applying the Milstein scheme instead of the approximation (4.3.12) has turned out not to help with respect to this problem.

Further possible improvements of the simulation method may be:

- Applying an adaptive time step, i.e. choosing the Δs very small for small values of the volatility process Y_s.

- Finding an analytical form of the transition matrix for the process Y in (4.2.10) and performing an exact simulation.

Without a substantial improvement of the Monte Carlo simulation e.g. by the methods mentioned above, the Finite Difference solution gives by far the better results.

4.4 Solution of the pricing PDE

Once the price of volatility risk λ is determined, one is able to solve equation (4.2.3) by the Finite Difference method. Together with the terminal condition, the problem reads

$$\begin{array}{rcl} \frac{\partial P}{\partial t} + LP & = & 0 \\ P(T,S,Y) & = & f(S), \end{array} \tag{4.4.21}$$

where L is the spatial differential operator of equation (4.2.3) and $f(S)$ is the option payoff.

Different transformation methods have been proposed in the literature in order to improve the numerical properties of the solution method, such as in Hilber et al. [45] or Clarke and Parrot [20]. Furthermore, one knows from [45] or Achdou and Tchou [1] that the solution of a stochastic volatility model is bounded in volatility direction, and for convex options, it has been shown in Romano and Touzi [74] that the option price is monotonically increasing. This suggests that, assuming enough regularity, the first derivative should converge to zero for high values of volatility, a condition that has often been used in the literature, such as in Clarke and Parrot [20]. An alternative idea of a boundary condition, proposed in Hilber et al. [45], is to introduce a transformation in volatility direction in order to get convergence to zero. In the sequel, we will investigate which of these methods is the most efficient one for our problem.

4.4.1 Transformation methods and boundary conditions

As proposed in Hilber et al. [45], the solution may be transformed in stock direction as well as in volatility direction in order to obtain solutions which converge exponentially to zero at infinity.

Transformation in Stock direction

For fixed volatility, we know from Seydel [81] that at the boundaries, the price of a European Call option is

$$\begin{aligned} C(t,S) &= 0 && \text{as} \quad S \to 0 \\ C(t,S) &= S - Ke^{-r(T-t)} && \text{as} \quad S \to \infty, \end{aligned} \tag{4.4.22}$$

where T is the time of expiry.

From (4.4.22), one may expect that the transformed option price

$$\tilde{C}(t,S,Y) := C(t,S,Y) - \left(S - Ke^{-r(T-t)}\right)^+ \tag{4.4.23}$$

converges to zero at the left as well as at the right boundary. By the usual rules of differential calculus, by (4.4.23) and (4.4.21) we get for $\tilde{C}$ the inhomogeneous terminal value problem

$$\begin{aligned} \frac{\partial \tilde{C}}{\partial t} + L\tilde{C} &= \frac{1}{2} Y S^2 \delta\left(S - Ke^{-r(T-t)}\right) \\ \tilde{C}(T,S,Y) &= 0, \end{aligned} \tag{4.4.24}$$

where $\delta(x)$ is the Dirac delta function, and consequently we have transformed the homogeneous terminal value problem into an inhomogeneous one with terminal condition zero. As mentioned in Hilber et al. [45], the solution of (4.4.24) decays to zero as $S \to \infty$ and as $S \to 0$.

Transformation in Volatility direction

Applying the result of Achdou and Tchou [1] stating that the solution of (4.4.21) remains bounded in the Y direction, a further transformation of $\tilde{C}$ has been proposed in Hilber et al. [45] in order to guarantee that the solution also decays exponentially as $Y \to \infty$, namely

$$u(t,S,Y) := e^{-\frac{1}{2}\alpha Y^2} \tilde{C}(t,S,Y), \tag{4.4.25}$$

where α is a constant that remains to be chosen.

Applying differential calculus as in [45], one obtains a new differential operator

$$\begin{aligned} \tilde{L}u \quad := \quad & Lu + 2\alpha\gamma\Theta Y^3 \frac{\partial u}{\partial Y} + \alpha\rho\sqrt{2\gamma\Theta} S Y^{\frac{5}{2}} \frac{\partial u}{\partial S} \\ & + \alpha \left[\Theta Y(\omega - Y) - \sqrt{2\gamma\Theta}\left(\rho \frac{\mu - r}{\sqrt{Y}} + \bar{\rho}\lambda\right) Y^2 + \gamma\Theta Y^2 + \alpha\gamma\Theta Y^4\right] u, \end{aligned} \tag{4.4.26}$$

and the problem (4.4.24) transforms into

$$\begin{aligned} \frac{\partial u}{\partial t} + \tilde{L}u &= \frac{1}{2} Y S^2 e^{-\frac{1}{2}\alpha Y^2} \delta\left(S - e^{-r(T-t)}\right) \\ u(T,S,Y) &= 0. \end{aligned} \tag{4.4.27}$$

Equation (4.4.27) holds for every choice of α. Furthermore, the solution of problem (4.4.27) has, in addition to (4.4.24), the property that it converges exponentially to zero as $Y \to \infty$. However, there is still a question about the behavior of the solution for $Y \to 0$. In Hilber et al. [45], a mirror principle has been applied, i.e. an extension of u to the negative half plane, applying $u(Y) := u(-Y)$ for $Y < 0$. Obviously, in this case, as $Y \to -\infty$, one again gets an exponential decay.

However, it has turned out that for our purposes, a logarithmic transformation in the coordinates, as also done in Clarke and Parrot [20], is very efficient. A mirror principle therefore is not possible. Instead, we always apply the Neumann condition that the first derivative is zero as $Y \to 0$, which is always reasonable, because near zero, the solutions of the transformed and the untransformed equation should be similar.

Coordinate transformations

First, all problems are transformed from a terminal to an initial value problem by the transformation $t \mapsto T - t$. As already mentioned, a coordinate transformation $Y = \omega e^y$ has been performed in all situations. For the stock direction, two possible transformations have been applied:

- A natural choice, also proposed in Hilber et al. [45], is the logarithmic transformation $S = Ke^x$.
- A transformation proposed in Clarke and Parrot [20] is the hyperbolic sine transformation,
$$S = \sinh\big(x + \sinh^{-1}(-K)\big) + K.$$
The idea of this transformation is that most points are needed in a region around $S = K$, whereas away from this region the solution will be almost flat.

The exponential transformation of problem (4.4.24) leads to the partial differential equation

$$\begin{aligned} \frac{\partial \tilde{C}}{\partial t} - L_{LOG}\tilde{C} &= \tfrac{1}{2}\omega e^y K e^{-rt}\delta(x - rt) \\ \tilde{C}(0, x, y) &= 0, \end{aligned} \tag{4.4.28}$$

where L_{LOG} is the spatial differential operator in logarithmic scale, i.e., by transforming L by the usual rules of differential calculus,

$$\begin{aligned} L_{LOG}\tilde{C} = & \left(r - \tfrac{1}{2}\omega e^y\right)\frac{\partial \tilde{C}}{\partial x} + \tfrac{1}{2}\omega e^y \frac{\partial^2 \tilde{C}}{\partial x^2} + \Theta(e^{-y} - 1 - \gamma)\frac{\partial \tilde{C}}{\partial y} + \gamma\Theta\frac{\partial^2 \tilde{C}}{\partial y^2} \\ & + \rho\sqrt{2\gamma\Theta}\sqrt{\omega}e^{\frac{y}{2}}\frac{\partial^2 \tilde{C}}{\partial x \partial y} - \sqrt{2\gamma\Theta}\left(\rho\frac{\mu - r}{\sqrt{\omega}}e^{-\frac{y}{2}} + \bar{\rho}\lambda\right)\frac{\partial \tilde{C}}{\partial y} - r\tilde{C}. \end{aligned} \tag{4.4.29}$$

At the boundaries, for $x \to \pm\infty$, one has, as also obtained in Hilber et al. [45], the exponential convergence $\tilde{C} \to 0$. Thus, one may impose Dirichlet boundary conditions, as in [45], as well as Neumann boundary conditions, as in Clarke and Parrot [20].

As mentioned at the beginning of this section, the monotonicity and boundedness of the solution in volatility direction suggests, in the logarithmic scale, that Neumann boundary conditions are always suitable for problem (4.4.28). On the other hand, as $Y \to \infty$, the option price converges to a constant strictly larger than zero, so that Dirichlet boundary conditions make no sense.

The exponential transformation of problem (4.4.27) gives the partial differential equation

$$\begin{array}{lcl} \frac{\partial u}{\partial t} - \tilde{L}_{LOG} u & = & \frac{1}{2}\omega e^{y} K e^{-rt} e^{-\frac{1}{2}\alpha\omega^2 e^{2y}} \delta(x - rt) \\ \tilde{u}(0,x,y) & = & 0 \end{array} \tag{4.4.30}$$

with the differential operator $\tilde{L}_{LOG}$ from (4.4.26) in logarithmic scale,

$$\begin{array}{rcl} \tilde{L}_{LOG} u & = & L_{LOG} u + 2\alpha\gamma\Theta\omega^2 e^{2y} \frac{\partial u}{\partial y} + \rho\alpha\sqrt{2\gamma\Theta}\omega^{\frac{5}{2}} e^{\frac{5}{2}y} \frac{\partial u}{\partial x} \\ & & +\alpha\omega^2 e^{2y} \left[\Theta(e^{-y} - 1) - \sqrt{2\gamma\Theta}\left(\rho\frac{\mu - r}{\sqrt{\omega}} e^{-\frac{y}{2}} + \bar{\rho}\lambda\right) + \gamma\Theta\left(1 + \alpha\omega^2 e^{2y}\right)\right] u. \end{array} \tag{4.4.31}$$

At the boundaries, one gets $u \to 0$ as $x \to \pm\infty$, and $y \to \infty$, as seen in Hilber et al. [45], with exponential decay. As discussed before, we have that $\frac{\partial u}{\partial y} \to 0$ as $y \to \pm\infty$. As a consequence, one may choose Dirichlet as well as Neumann boundary conditions for $x \to \pm\infty$ and $y \to \infty$, but one should impose Neumann boundary conditions for $y \to -\infty$.

The hyperbolic sine transformation of problem (4.4.24) gives

$$\begin{array}{lcl} \frac{\partial \tilde{C}}{\partial t} + L_{SH}\tilde{C} & = & \frac{1}{2}\omega e^{y} \frac{K^2 e^{-2rt}}{\sqrt{1 + K^2(1 - e^{-rt})^2}} \delta\left(x + \sinh^{-1}(-K) + \sinh^{-1}\left(K(1 - e^{-rt})\right)\right) \\ \tilde{C}(0,x,y) & = & 0 \end{array} \tag{4.4.32}$$

with the transformed differential operator

$$\begin{array}{rcl} L_{SH}\tilde{C} & = & \frac{1}{2}\omega e^{y} \tilde{T}^K(x_K)^2 \frac{\partial^2 \tilde{C}}{\partial x^2} + \gamma\Theta\frac{\partial^2 \tilde{C}}{\partial y^2} \\ & & +\tilde{T}^K(x_K)\left(r - \frac{1}{2}\omega e^{y}\tanh(x_K)\tilde{T}^K(x_K)\right)\frac{\partial \tilde{C}}{\partial x} \\ & & +\rho\sqrt{2\gamma\Theta}\sqrt{\omega} e^{\frac{y}{2}} \tilde{T}^K(x_K)\frac{\partial^2 \tilde{C}}{\partial x \partial y} - r\tilde{C} \\ & & +\left(\Theta(e^{-y} - 1 - \gamma) - \sqrt{2\gamma\Theta}\left(\rho\frac{\mu - r}{\sqrt{\omega}} e^{-\frac{y}{2}} + \bar{\rho}\lambda\right)\right)\frac{\partial \tilde{C}}{\partial y} \end{array} \tag{4.4.33}$$

with

$$\tilde{T}^K(x) := \frac{\sinh(x) + K}{\cosh(x)}$$

and

$$x_K := x + \sinh^{-1}(-K).$$

In Table 4.6, a list of the transformation methods is given which will be compared in the next section by numerical experiments. The methods used for these numerical experiments are combinations of the ones explained above.

4.4.2 Discretization

For the discretization in space, the standard central difference method has been used, which has a convergence of order $(\Delta x)^2$ and $(\Delta y)^2$, respectively. As in Clarke and Parrot [20], this discretization method is only stable if the Peclet condition is satisfied, i.e. if

$$max\left(\frac{|a_x(x_i,y_j)|\Delta x}{2a_{xx}(x_i,y_j)},\frac{|a_y(x_i,y_j)|\Delta y}{2a_{yy}(x_i,y_j)}\right)<1,$$

where (x_i, y_j) are the mesh points, a_{xx} and a_{yy} are the diffusion coefficients, and a_x and a_y the convection coefficients.

As in [20], for matrix entries where this condition is not satisfied, an automatic change is done from the central difference to the forward or backward difference method (see [20] for details) in order to get stability. These difference methods have only first order convergence. But one may expect that it is only necessary for a few elements to change from the central difference to another method, so that this change does not substantially affect the general convergence behavior.

For the discretization in time, the Crank-Nicolson method has been used for all experiments. Even if this method is not L-stable in connection with delta functions, this does not affect our solutions. For the numerical solution of the linear system of equations, the iterative Gauss-Seidel method has been applied with an error tolerance of 10^{-9}. The amount of iteration steps is shown in Tables 4.8.1 and 4.8.2.

Finally, the discretization of the delta function is done in the following way. For a spatial variable x with $x_j < x < x_{j+1}$, where x_j, x_{j+1} are grid points and $\Delta x = x_{j+1} - x_j$, the Dirac delta function is approximated by

$$\begin{array}{lcl}\delta(x_j - x) & \approx & \frac{x_{j+1}-x}{\Delta x}1_{x\in[x_j,x_{j+1}[},\\ \delta(x_{j+1} - x) & \approx & \frac{x-x_j}{\Delta x}1_{x\in[x_j,x_{j+1}[},\end{array} \tag{4.4.34}$$

i.e. a weighted distribution is applied at the neighboring points.

Table 4.6: Transformation and truncation methods

Method	Price transformation	Coordinate transformation	Equation number	Space domain	Boundary conditions
method 1	Stock (4.4.23) Vol. (4.4.25)	$S = Ke^x$ $Y = \omega e^y$	(4.4.30)	$[-x_m, x_m]$ $\times$ $[-y_m, y_m]$	$u(\pm x_m) = 0$ $u(y_m) = 0$ $\frac{\partial u}{\partial y}(-y_m) = 0$
method 2	Stock (4.4.23)	$S = \sinh(x + \kappa)$ $+K$ $\kappa = \sinh^{-1}(-K)$ $Y = \omega e^y$	(4.4.32)	$[0, x_m]$ $\times$ $[-y_m, y_m]$	$\frac{\partial u}{\partial x}(\pm x_m) = 0$ $\frac{\partial u}{\partial y}(\pm y_m) = 0$
method 3	none	$S = Ke^x$ $Y = \omega e^y$	$\frac{\partial u}{\partial t} = L_{LOG}u$ $u(0, x, y) =$ $K(e^x - 1)^+$	$[-x_m, x_m]$ $\times$ $[-y_m, y_m]$	$u(-x_m) = 0$ $u(x_m) =$ $K(e^{x_m} - e^{-rt})$ $\frac{\partial u}{\partial y}(\pm y_m) = 0$

Table 4.7: Method 1

S/K	$x_m = 1.5$ $y_m = 4.0$ $\alpha = 1000$	$x_m = 1.5$ $y_m = 4.0$ $\alpha = 10000$	$x_m = 1.5$ $y_m = 6.0$ $\alpha = 10000$	$x_m = 1.5$ $y_m = 7.0$ $\alpha = 1000$	$x_m = 2.5$ $y_m = 4.0$ $\alpha = 10000$	$x_m = 2.5$ $y_m = 6.0$ $\alpha = 1000$
50%	0.01588	0.01588	0.01626	0.01626	0.01588	0.01625
80%	0.01281	0.01281	0.01281	0.01281	0.01281	0.01281
100%	0.01252	0.01252	0.01252	0.01252	0.01252	0.01252
120%	0.01275	0.01275	0.01276	0.01276	0.01275	0.01276
150%	0.01364	0.01364	0.01366	0.01366	0.01364	0.01366

4.4.3 Comparisons, Numerical Experiments

Impact of boundary truncation

In this section, we consider the impact of the truncation points on the implied volatility for the three different methods described in last section.

Apart from the parameters of the standard parameter set (see section 4.3.1), the parameters for this analysis are the following:

$$\Delta x = 0.01, \quad \Delta y = 0.1, \quad \Delta t = 0.1.$$

From Table 4.7, one can see that in *method 1* there is no impact of the choice of the boundary truncation x_m nor of the choice of the value of α. On the other hand, the computation time as well as the amount of iteration steps is much smaller in *method 1* when $\alpha = 10,000$ is chosen. Only a slight impact occurs on y_m, and only in regions far out of the money and far in the money. From $y_m = 6.0$ on, no change is visible any more. The other methods show a similar dependency on x_m and y_m and are not reported here. Altogether, one can say that the choice $x_m = 1.5$ and $y_m = 6.0$ is completely sufficient for getting exact results.

It may be interesting if things change for the case with correlation, because the price of volatility risk then is not constant in volatility any more for large time horizons, but decreases with increasing volatility, as shown in Table 4.5.2. Further numerical experiments which are not reported here show that the choice of boundaries is still good for the case with correlation.

Table 4.8.1: Convergence of method 1, without correlation

S/K	$\Delta x = 0.05$ $\Delta y = 0.5$	$\Delta x = 0.02$ $\Delta y = 0.2$	$\Delta x = 0.01$ $\Delta y = 0.1$	$\Delta x = 0.005$ $\Delta y = 0.05$
50%	0.01571	0.01662	0.01626	0.01629
80%	0.01291	0.01283	0.01281	0.01281
100%	0.01257	0.01258	0.01252	0.01255
120%	0.01274	0.01276	0.01276	0.01275
150%	0.01381	0.01369	0.01366	0.01366
Δt	0.1	0.1	0.1	0.1
iter	5	15	30	115
time	0.2 min	1 min	7 min	80 min

Impact of discretization error

In this section, we describe the convergence of the numerical methods to the solution of the differential equation (4.4.21) when the grid becomes finer and finer, still with the boundary conditions of Table 4.6, with the values of $x_m = 1.5$ (*method 2*: $x_m = 2.0$), $y_m = 6.0$ and $\alpha = 10,000$, as discussed in the last section, with the standard parameter set and $q = 400$. The variable *iter* denotes the average amount of iterations per time step and the variable *time* the total calculation time.

In Table 4.8.1 the convergence of *method 1* is shown. At the finest grid, the other methods which are not reported here give precisely the same results at the values of x_m, y_m, α and q just mentioned, rounded to four significant figures. Thus we can think of these results as the true solution, at least at the accuracy considered here, i.e. 0.05%. The deviations from this solution we may view as the discretization errors.

For the best accuracy reported here, a rather long computation time is necessary. However, using the grid $\Delta x = 0.02$ and $\Delta y = 0.2$, one still gets an accuracy of 1%, with a computation time of 1-2 minutes, as can be seen in the table. The same also holds for the other methods which are not reported here, as well as for a parameter value of $q = 0$ instead of $q = 400$.
In the case with correlation, Table 4.8.2 shows the convergence for the correlation $\rho = -0.75$ and $q = 1.75$ with *method 2*. Further numerical experiments with the other methods which we do not describe here show a similar convergence behavior, and the result of the calculation with the finest grid is still almost the same for all methods. We consider this result therefore as the true solution.

The calculation with a strong correlation is numerically more delicate than the

Table 4.8.2: Convergence of method 2, with correlation $\rho = -0.75$

S/K	$\Delta x = 0.05$ $\Delta y = 0.2$	$\Delta x = 0.02$ $\Delta y = 0.1$	$\Delta x = 0.01$ $\Delta y = 0.05$	$\Delta x = 0.005$ $\Delta y = 0.02$
50%	0.01170	0.00886	0.00879	0.00878
80%	0.01035	0.00967	0.00959	0.00957
100%	0.01076	0.01078	0.01075	0.01076
120%	0.01211	0.01197	0.01197	0.01197
150%	0.01418	0.01389	0.01385	0.01384
Δt	0.1	0.1	0.1	0.1
iter	11	27	66	287
time	0.5 min	2.8 min	31 min	900 min

one without correlation. We have to be careful in considering the quotient of the step sizes in x- and y-direction. Furthermore, one has to choose the right 7-point stencil for the space discretization. The 9-point stencil gave worse results than the ones we present here.

Note that we observe these features only in the case of a strong correlation. Further experiments with only a small positive correlation show similar convergence properties as for the case without correlation.

Again, for a nearly perfect solution, the computation time is rather long. Applying a grid of $\Delta x = 0.02$ and $\Delta y = 0.1$, one needs a time of less than 3 minutes, and only an error of 1% or less.

It turns out that with respect to the computational time, *method 1* with $\alpha = 10000$ is the best one, although for a smaller number of grid points, the methods become comparable. With respect to the accuracy, the methods are comparable, although *method 2* seems to exhibit the best convergence for far out of the money options, with exception of the coarsest grid.

It turns out that *method 3* also shows a very good efficiency. This is a rather surprising result, because *method 3* is the method with the fewest transformations. As a consequence, at least for the parameter set and for the values of stock prices considered here, at an instantaneous volatility of $Y = \omega$, the transformation methods presented above do not appear to lead to a considerable gain in numerical efficiency, in spite of their sophistication.

In particular, *method 2*, which uses a very sophisticated coordinate transformation, shows a very good convergence until a grid of $\Delta x = 0.02$ and $\Delta y = 0.1$, but needs a higher computational time for the same amount of grid points than *method*

1. However, for fewer grid points, this disadvantage disappears. Furthermore, the method may be improved by using a multigrid instead of the Gauss-Seidel method, as also done in Clarke and Parrot [20].

4.4.4 Alternative: Fast mean-reversion approximation

In Fouque et al. [35], an alternative method for the calculation of option prices with stochastic volatility has been presented, using a fast mean-reversion approximation. It would be interesting to see whether this method can be applied for the solution of the pricing PDE as an alternative to the Finite Difference method.

First, as already mentioned in [35], in order to have the simple formula for calculating the option prices obtained there it is necessary to have a price of volatility risk which depends only on the volatility y but not on the stock price nor on the time. Looking at formulas (4.2.7) and (4.2.8) or at Tables 4.5.1 and 4.5.2, this requirement is not satisfied except for large values of t.

With a price of volatility risk depending on t, the constants obtained in [35] become time-dependent, and one would have to calculate them for each t on a grid. The solution cannot be obtained in closed form any more, but still has to be obtained by the Finite Difference method, even with only one spatial variable. Performing for each t on a grid a numerical integration and subsequently solving a Finite Difference equation again needs a large computational effort. Therefore, the advantage of the fast mean-reversion approximation, namely the numerical efficiency, is not valid any more, even when one continues to apply only the first-order approximation.

In our calculations we tried to approximate the price of volatility risk by its long-run value, which is then time constant, in the case with as well as without correlation. Because by Tables 4.5.1 and 4.5.2 the price of volatility risk is decreasing in time, a replacement by its long-run value leads to an underestimation of the price of volatility risk, and therefore, by the Henderson comparison theorem (see section 4.2.4), to an overestimation of the implied volatility. Performing the calculation for the case without correlation and for a time horizon $t = 90$, we obtain an overestimation of the implied volatility which is much larger than the one with the fewest grid points in Tables 4.8.1 and 4.8.2, especially for high values of q. As a consequence, this method cannot give reasonable results in the case of time dependency. Perhaps for investors with a very large time horizon ($t \to \infty$) this may be different.

Another issue is that in the case without correlation one has for $q = 0$ that the implied volatility calculated by fast mean-reversion approximation is constant over all fractions S/K, so that it cannot be true for all values of S/K. Thus, it becomes clear that when applying fast mean-reversion approximation for the calibration, higher-order terms would have to be considered.

Altogether, for the given parameter set, either the numerical accuracy of the fast mean-reversion approximation with the additional approximation of a time constant price of volatility risk is much less than the one of the Finite Difference approximation, or one would have to perform a better approximation which again would lead to a large computational effort.

4.4.5 Conclusion

In the cases of a small correlation or no correlation, all methods in general show a good convergence in a reasonable time. In the case of a strong correlation, one still has a good convergence, but it needs slightly more time. The question of which method is the most accurate one strongly depends on the points that are considered and cannot be answered in general. *Method 1* is the fastest if a value of $\alpha = 10000$ is chosen; with a lower value, the computation takes more time than with other methods. *Method 3* is the slowest, but the accuracy is comparable to *method 1* or *method 2*.

For the following calibration, we will take *method 1* for the case without correlation, because it is the fastest one, with $\alpha = 10000$, $\Delta x = 0.01$, and $\Delta y = 0.1$.

4.5 Calibration

In the previous sections, we have seen how to calculate a price of a European Call option under the assumption that the correct pricing measure is a q-optimal measure. We therefore obtain the implied volatility as a function of the parameter q, with a numerical error of $\Delta\sigma_{impl} = 0.00003$ for *method 1*, which arrives from the following sources (we focus in this section on the case with no correlation):

- An error arises from the error $\Delta\lambda(q)$, i.e. from solving the Hobson representation equation. Using the Finite Difference solution and considering the last column of Table 4.5.1 as the true results ($\Delta y = 0.1$), we calculated the implied volatilities using this exact solution of λ and compared this result with the one obtained from the implied volatility calculated with $\Delta y = 0.2$ from Table 4.5.1. It turned out that there is no difference in implied volatility at those 4 significant numbers. Therefore this source of error is negligible compared to the other one.

- An error arises from the Finite Difference approximation of the pricing PDE. From Table 4.8.1, considering the last column as the true solution, one can see that the error in *method 1* is 0.00003 around $q \approx 400$. We assume that this error remains the same in a region around $q = 400$.

On the other hand, if options are available in the market, one can find a specific value of q, say $\hat{q}$, such that the option prices are given by the $\hat{q}$-optimal measure.

In this chapter, we only performed a rough analysis, using the average daily implied volatilities of at the money options from S&P500, which is $\sigma_{impl}^{emp} = 0.01260$, or about 20% on a yearly basis. This has been compared with the implied volatility calculated in the former sections, for at the money options, with average instantaneous volatility $\sqrt{\omega}$.

The results for the case without correlation are presented in the following table:

q	σ_{impl}	$\Delta\sigma_{impl}$
410	0.0125847	0.00003
413	0.0125999	0.00003
414	0.0126076	0.00003
415	0.0126152	0.00003

Here the error has been taken from Table 4.8.1, assuming that there is no large difference in error in this small region of q.

From the inverse function theorem, we know that in a region around $\hat{q}$ we have that

$$\Delta q \approx \frac{\Delta\sigma_{impl}(q)}{\sigma'_{impl}(q)}.$$

An approximation of $\sigma'_{impl}(q)$ from the table above is

$$\sigma'_{impl}(413) \approx \frac{\sigma_{impl}(415) - \sigma_{impl}(410)}{5} = 0.0000061,$$

and therefore

$$\hat{q} = 413 \pm 5,$$

which implies an error of slightly more than 1%. This corresponds to a power of the representative utility function of $\hat{\gamma} = \frac{\hat{q}}{\hat{q}-1} = 1.0024$. Therefore the price of volatility risk is $\lambda(\hat{q})$, calculated by the Finite Difference method described in section 4.3.

When doing the calibration for the model with the correlation $\rho = -0.75$, a huge $q > 1000$ would have to be taken. It follows that the stochastic volatility model 4.2.1, in combination with power utility functions, does not produce risk aversions which are typically observed in the market. However, the calibration procedure described here can be applied to any other stochastic volatility model which is covered in Hobson [47], for example a square-root model as in Heston [44].

4.6 Conclusion

In this chapter, we have seen how an investor preference based approach for the calculation of the price of volatility risk and the option prices as well as for calibration can be used, considering the computation of the price of volatility risk as a stochastic control problem. It has turned out that, using appropriate methods and boundary conditions, such a calculation is possible at a reasonable time.

For the calculation of the price of volatility risk, the explicit Finite Difference method has turned out to be very efficient, and it is clearly preferable to a simulation solution. On the other hand, for the remaining pricing PDE there is no clear preference of one method to another. This choice depends on the accuracy one would like to have and therefore the amount of grid points, as well as on the question of which points (at the money, out of the money or in the money) are the most relevant. A remarkable feature is that the simplest method which does not make use of sophisticated transformation methods is competitive compared to the other methods.

A calibration has been performed only in a rough way, comparing the average daily implied volatility with the one obtained by our calculations, and only for the case without correlation. The case with a strong negative correlation has not lead to reasonable calibration results. In further empirical research, one could try to calibrate the model with correlation using other stochastic volatility models than we have done, and one could include the dependency of the option prices on the instantaneous volatility, as well as the time to expiry.

Also, as a further empirical analysis, it would be interesting to see whether the calibration performed with our investor preference based approach can give a better fit to the empirical price of volatility risk than the calibrations with the assumption that this price is constant, considering a suitable stochastic volatility model.

Our calibration method has made essential use of the fact that using a model with stochastic volatility as well as power utility functions for the investor preference model, the price of volatility risk, or more generally the dual minimizer of the optimal control problem, can be obtained by the Hobson representation equation. It would be interesting to see whether such an optimal control problem could also be solved in a reasonable time assuming a more general investor preference model, or, as an alternative, if something similar could be made in other models than those with stochastic volatility. In particular, the question arises whether one could apply a similar method to determine the prices of risk in term structure models.

As a further numerical issue, one may try to improve the solution by Monte Carlo simulation by choosing an appropriate adaptive time step, by applying other numerical simulation schemes than Euler or Milstein, for example Andersen [2], or by finding an analytical formula for the conditional density and doing exact simulation. However, to beat the efficiency of the presented Finite Difference solution is ambitious.

Chapter 5

CVaR pricing and hedging in Unit-Linked insurance products

5.1 Introduction

Unit linked insurance products become more and more popular in the insurance industry, because they combine the classical coverage against risks such as death, longevity, and disability in life insurance with the possible chance of large capital earnings that traditionally banks offer.

The pricing of such insurance products needs a combination of classical actuarial principles and principles from financial mathematics. Such combinations have been treated in Møller [65], where the focus mainly is on pricing using a standard deviation principle, but which also contains some hints for a general utility function.

In general, financial valuation principles are based on a replication of a claim, whereas in insurance, a risk-loading is charged, because the claims cannot be hedged. For unit-linked insurance products, one can assume that a full hedge is not possible, because there is a nonhedgeable component. The insurance company will therefore still ask for a risk loading. However, because of the financial component of these products, at least a partial hedge should be possible, which offers the opportunity to reduce the minimal necessary risk loading. The general aim of this chapter is to find a minimal price as well as the corresponding hedging strategy which make a unit-linked insurance claim acceptable for the insurer. Here we define acceptability in terms of coherent risk measures, as in Artzner et al. [3]. We will take CVaR as risk measure. From a practical point of view, this pricing method has the advantage of being more closely related to the cost of capital method than other pricing rules. It is a method which is typically applied in insurance industry.

The topic of risk measure pricing and minimization has recently been treated in several papers. One of the first papers discussing this idea was Carr et al. [15]. In the sequel, general principles of risk measure pricing have been developed in Klöppel and Schweizer [55], Xu [86], Cheridito and Kupper [18], and Jobert and Rogers [50], see also Cherny [19]. The pricing principles used in these papers are similar to the principles discussed in this chapter. An abstract framework for pricing on the basis of coherent risk measures has been firmly established in these papers. However, to calculate prices in specific situations one typically still needs to solve an optimization problem. In this chapter, we address the optimization problem for the case of unit-linked insurance products, and for CVaR as specific risk measure.

A concrete result for the problem of minimization of Worst Conditional Expectation has been developed in Sekine [80]. In this paper, a formula is obtained for the solution of the minimization problem. However, the problem investigated in [80] is not the same one as the one of this chapter. We will again discuss this issue later in the chapter. In Ilhan et al. [49], the authors solve, apart from theoretical considerations, the problem of numerical risk measure pricing in the example of expected shortfall in the sense of Föllmer and Leukert [33]. For the solution, a Hamilton-Jacobi-Bellman method is applied, which leads to a nonlinear partial differential equation with two variables. The authors point out the computational challenge of this procedure.

Risk-minimizing strategies for unit-linked insurance products have already been treated in Møller [64], where risk-minimizing is understood in the sense of local risk minimization, see e.g. Schweizer [78]. Other papers of Møller consider the variance as definition of risk. Minimization of value at risk in unit-linked insurance products as well as corresponding pricing principles have been considered in Melnikov and Skornyakova [62], where the authors mainly focus on the case with only one insured person, and in this way obtain analytic formulas. For the case of many insured persons, they obtain bounds which are derived by considering the financial and insurance risk separately.

As already stated, in our chapter, we consider CVaR as risk measure. We are interested in a risk minimizing strategy when the financial and insurance risk are considered in an integrated manner. Assuming as in Melnikov and Skornyakova [62] that all information about the insurance process is arriving only at the end of the time period, we obtain the minimal price making the claim acceptable as well as the corresponding hedge. It turns out that with this simplification, the problem becomes easy to calculate, and for the specific model we are using, we obtain analytic formulas. And even in situations where this assumption is unrealistic, the method presented here still leads to an upper bound for the CVaR price, including the corresponding hedge.

Actually, the results presented in this chapter can be applied not only for unit-linked insurance products, but also for other situations of CVaR pricing in incomplete markets. The key issue only is the pricing of a payoff which depends on a complete financial market, as well as on another source of uncertainty which is independent of the financial market and which cannot be replicated. One may think about the option of a company to buy a specific commodity at a specific time in a specific currency, where the decision whether or not to buy depends on the foreign exchange rate, but also on other circumstances which are independent of the financial market.

Our approach is based on a result of Rockafellar and Uryasev [73], with which we can connect the problem of CVaR pricing to the earlier results [32] and [33] of Föllmer and Leukert. In these papers the problem of minimization of expected shortfall is connected to the Neyman-Pearson theory (see [85]), and for some specific cases, analytic formulas are developed. We further develop those results for the specific case of unit-linked insurance products. In particular, we show in general how these papers can be connected for obtaining a CVaR price for unit-linked insurance products, as well as in a specific example. Furthermore, we extend a theorem presented in Föllmer and Leukert [33] to the case where the insurance probabilities are discrete. Finally, we apply the results obtained to a specific unit-linked insurance model and explicitly state the formulas and the numerical results.

The structure of the chapter is as follows. In section 2, we formulate the general model, as well as the CVaR pricing principle. In section 3, we present an algorithm for calculating the CVaR price under the additional assumption of continuous distribution. We give an example in which we approximate the discrete insurance probabilities by a normal approximation. In section 4, we prove an extension of the theorem in [33] mentioned before in order to be able to apply the result to discrete probabilities. With this result we are able to obtain analytical formulas for some specific models or to solve the problem numerically. In section 5, we again present an algorithm for the calculation of the CVaR price, without assuming continuous distributions. We give a specific example of a unit linked survival insurance, for which we obtain analytical formulas. Furthermore, we give an explicit numerical example for the CVaR price as well as an analytical formula for the corresponding hedge. Section 6 concludes.

5.2 Problem specification and general statements

5.2.1 Insurance model and problem specification

We are dealing with a probability space $(\Omega, \mathcal{F}, P)$. On this probability space, a vector-valued stochastic process Z_t is defined in order to represent the insurance state process. The states of the financial market are modeled by another vector-valued process X_t on $(\Omega, \mathcal{F}, P)$. The filtration $\mathcal{F}_t$ is given by the natural filtration generated by X_t and Z_t. It is assumed that Z_t and X_t are Markovian, X_t is continuous, and that this market is complete, in the sense that every contingent claim $F(X_T)$ can be replicated by a suitable trading strategy

$$F(X_T) = S_0 + \int_0^T \pi_t dS_t,$$

where S_t is the vector-valued process representing the available financial assets. It is assumed that this process is a vector-valued function of time and the financial state variables, i.e.

$$S_t = \tilde{S}(t, X_t)$$

with a measurable function $\tilde{S}$, such that S_t is a vector-valued continuous semimartingale. A sufficient condition for this would be, by the Itô formula, that X_t is a continuous semimartingale and $\tilde{S}$ is twice differentiable. It is assumed that only the assets S_t can be used for trading. Throughout this chapter, we assume that there exists an equivalent measure Q such that S_t is a local martingale. Furthermore, we assume that X_t and Z_t are independent under P.

The option payoff due to a unit-linked insurance product at the terminal time T is given by a nonnegative product-measurable function $g(X_T, Z_T)$ depending on the financial market as well as on the insurance process.

Remark 5.2.1. By extending the state space, this model also admits payoffs which depend on all states up to time T. Therefore, to restrict to payoffs depending only on states at time T is not really a restriction.

Remark 5.2.2. A genuine restriction is to assume that payoffs can only take place at time T, even if they may depend on earlier times. This has to be assumed because we aim to calculate the CVaR at the terminal time T. It is in general not clear how to define the CVaR if there are different payoff times. In some specific examples, it may make sense to divide the payments at all times by a numéraire (a reasonable choice may be a zero bond with expiry at time T), and take the CVaR at the fixed time T. This situation is also covered by our model.

At this point we recall the CVaR risk measure from Rockafellar [73]:

Definition 5.2.3. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. Then $CVaR_\beta(X)$ at a certain level β is the mean of the distribution function

$$\psi_X(\xi) := \begin{array}{ll} 0 & \text{if} \quad \xi < \xi_\beta \\ (\phi_X(\xi) - \beta)/(1-\beta) & \text{if} \quad \xi \geq \xi_\beta, \end{array}$$

where ϕ_X is the cumulative distribution function of X, and ξ_β the Value-at-Risk at level β, that is

$$\xi_\beta = \min\{\xi \mid \phi_X(\xi) \geq \beta\}.$$

The credit-constrained CVaR pricing rule, which for simplicity we will in the sequel denote by CVaR price, now is the following.

Definition 5.2.4. Let the CVaR level β be given, as well as a self-financing portfolio B_t. Then the *CVaR price* is the minimal capital V_0 such that there exists a self-financing predictable strategy π with respect to the filtration $\mathcal{F}_t$ such that

$$CVaR_\beta\left[(g(X_T, Z_T) - Y_T^\pi)\right] \leq 0 \tag{5.2.1}$$

and such that $Y_t^\pi \geq B_t$ for all t, where the wealth process Y_t^π is defined by

$$Y_t^\pi = V_0 + \int_0^t \pi_u dS_u.$$

The problem now is to find the CVaR price according to Definition 5.2.4. The aim is also to find the corresponding strategy π.

Remark 5.2.5. In the context of the general theory of coherent risk measures, see Artzner et al. [3], equation (5.2.1) means that the risk is acceptable for the insurance company.

Remark 5.2.6. At this stage, it is not clear that such a CVaR price exists. This is the issue of Proposition 5.2.15.

Remark 5.2.7. In contrast to Klöppel and Schweizer [55] or Xu [86], we only consider strategies with wealth processes uniformly (that is independent of the specific process) bounded from below by a self-financing portfolio. Typically, one may think of a zero-bond. Economically, this makes sense, because no insurance company has an unlimited credit line. A lower bound of $-cB_t$, for a constant c and a zero-bond B_t, then means the limit until which the insurance company can use credit.

Mathematically, one needs this uniform bound, because Assumption 5.4 from Klöppel and Schweizer [55] (or Assumption 2.3 from Xu [86]) is not satisfied by the CVaR in (for instance) the Black-Scholes model. Assumption 5.4 states, in our terminology, that $\inf_{g \in \mathcal{C}} \rho(-g) > -\infty$, where ρ is the coherent risk measure and $\mathcal{C}$

the set of superreplicable claims at zero wealth. Essentially the same problem also leads to CVaR prices which are not necessarily market consistent, an issue which will be discussed in the following section.

As a consequence of the limited credit condition, the CVaR price is not translation invariant, in contrast to the risk measure prices in Klöppel and Schweizer [55].

Remark 5.2.8. Mathematically, a uniform lower bound B_t, where B_t is a general self-financing strategy, leads to the same problem as the assumption that the wealth process Y_t^π remains nonnegative. Indeed, adding capital B_0 to the initially available capital and B_T to the terminal payoff, the restriction of a nonnegative wealth process Y_t for this modified problem is the same as the restriction that $Y_t \geq B_t$ for the original problem. In the sequel, we will therefore always assume that $B_t = 0$, i.e. that the wealth cannot be negative.

Remark 5.2.9. Conversely, for a fixed initial capital V_0, one can also ask for the minimal possible CVaR and the corresponding hedging strategy. This question is sensible if the market is competitive and an insurance company is not able to independently price its products.

Remark 5.2.10. The pricing method described here completely differs from the one using an equivalent martingale measure. Actually, there typically does not exist any absolutely continuous probability measure such that the CVaR price of any insurance claim is given as an expectation under this measure. To see this, consider a simple economy which consists only of two insurance states which have both 50% probability, and only one risk-free financial asset with return zero. For an insurance option which pays 1 in the first state and 0 in the second, one can see that the CVaR price must be 1. As a consequence, the only probability measure under which the expected payoff is equal to the CVaR price is $Q = (1, 0)$. On the other hand, when considering an option which pays 0 in the first state and 1 in the second, the only possible probability measure is $Q = (0, 1)$.

Because the price of risk is typically obtained from the pricing measure, it follows that no market price of risk for the insurance variable Z_T can be defined. If we define $Z_T := 1_{\text{state 1}}$, the Sharpe ratio, from the point of view of the insurer, would be 1 for the claim Z_T as well as for $1 - Z_T$.

5.2.2 Market consistent CVaR

The pricing rule from Definition 5.2.4 does not necessarily produce a market consistent price. Indeed, consider a Black-Scholes model and a replicable claim $g(X_T) = 1_{X_T < c}$ with a very small constant c. Take the constant a such that for the martingale measure Q, $Q(X_T < c) = -aQ(X_T > c)$. Then the replicating hedge of the claim $-a1_{X_T > c}$ gives a CVaR strictly smaller than 0, that is $CVaR(1_{X_T<c} + a1_{X_T>c}) < 0$, as can be checked by calculation. Obviously, by definition, $-a1_{X_T>c}$ has the same price as $1_{X_T<c}$. It follows that there is a price $\pi < Q(X_T < c)$ at which the claim is still acceptable.

This is a problem of the CVaR risk measure which does not occur in Klöppel and Schweizer [55], because in this paper, market consistency of the prices is proved under the assumption that the infimum attainable risk over all admissible strategies starting with zero wealth is bounded (cf. again Assumption 5.4 of [55]). This assumption is not satisfied by the CVaR risk measure.

To guarantee market consistency, one has to change the risk measure. One possible way to do this is by adding the measure Q^* to the set of test measures in the dual representation of CVaR. This measure is defined as the one which gives the original probabilities for the insurance process and the risk-neutral ones for the financial process in such a way that X_t and Z_t are independent under Q^*. This gives a coherent risk measure, and the new pricing rule is market consistent and is nothing else than what practitioners applying the CVaR criterion for pricing would probably do. When the price obtained by Definition 5.2.4 would have a negative risk loading, they would set the risk loading to zero, otherwise, they would take the price as obtained by the CVaR criterion. We refer to this new risk measure as market consistent CVaR. Formally, we have the following

Proposition 5.2.11. *The risk measure price with respect to the market consistent CVaR, denoted by V_M, is given by*

$$V_M(g(X_T, Z_T)) = \max\left(E^{Q^*}[g(X_T, Z_T)], V_0(g(X_T, Z_T))\right),$$

where V_0 is the CVaR price.

Proof. It is clear that $V_M \geq \max(E^{Q^*}[g], V_0)$. Let $E^{Q^*}[g(X_T, Z_T)] \leq V_0(g(X_T, Z_T))$. Then there exists a strategy π such that $CVaR(g(X_T, Z_T) - V_T^\pi) \leq 0$. But by assumption $E^{Q^*}[g(X_T, Z_T) - V_T^\pi] = E^{Q^*}[g(X_T, Z_T)] - V_0(g(X_T, Z_T)) \leq 0$. It follows that the risk is also acceptable with respect to the new measure, and therefore $V_M \leq V_0$.

Let now $E^{Q^*}[g(X_T, Z_T)] > V_0(g(X_T, Z_T))$. Then there exists a strategy which makes the risk CVaR-acceptable at a price $V_0 < E^{Q^*}[g]$. By the monotonicity of coherent risk measures, the risk is also CVaR acceptable at the price $E^{Q^*}[g]$. It follows that $E^{Q^*}[g]$ is a price which makes the risk acceptable with respect to the new risk measure, and therefore $V_M \leq E^{Q^*}[g]$. □

In practice, the situations where this new measure leads to different prices than the original one depends on the CVaR level β, on the credit limit, and on the payoff

g. In many situations, and in particular in all of our examples, there is no difference between the price with respect to this new measure and the CVaR price.

5.2.3 Connection to minimization of Expected Shortfall

It follows from Rockafellar and Uryasev [73] that the conditional value at risk of a random variable X at a level β is given by

$$CVar(X) = \min_a \left(a + \frac{1}{1-\beta} E\left[(X-a)^+\right] \right). \tag{5.2.2}$$

Using this, the problem of CVaR pricing from Definition 5.2.4 can be reformulated as follows. Find the minimal initial capital V_0 such that there exist an allowed strategy π and a parameter a with

$$f(a, \pi; V_0) \leq -a(1-\beta), \tag{5.2.3}$$

where $f(a, \pi; V_0)$ is given by

$$f(a, \pi; V_0) := E\left[\left((g(X_T, Z_T) - a)^+ - Y_T^\pi \right)^+ \right]. \tag{5.2.4}$$

Remark 5.2.12. From (5.2.2), or from (5.2.3) and (5.2.4), it is clear that a must be nonpositive. It follows that $g(X_T, Z_T) - a \geq 0$ is always satisfied if the insurance claim is nonnegative.

Remark 5.2.13. By the nonnegativity of Y_T^π, it follows that

$$\left((g(X_T, Z_T) - a)^+ - Y_T^\pi \right)^+ = \left(g(X_T, Z_T) - a - Y_T^\pi \right)^+ .$$

The reason why we write the expectation as in (5.2.4) is that it helps to guarantee nonnegativity of Y_T^π.

Remark 5.2.14. By Rockafellar and Uryasev [73], the minimum on the right hand side of (5.2.2) is always attained. It follows that the minimum of V_0 under condition (5.2.1) is attained if the minimum $V_0(a)$ under condition (5.2.3) is attained for any fixed a.

For fixed a, we can define $V_0(a)$ as the minimal initial capital such that there exists a strategy π which satisfies (5.2.3).

Proposition 5.2.15. *The following holds:*

1. *For each a, the minimum $V_0(a)$ under condition (5.2.3) is attained.*

2. *The function $V_0(a)$ is convex in a.*

3. *If a^* minimizes $V_0(a)$ and π^* is the strategy which minimizes $V_0(a^*)$ for the given a^*, then π^* is the strategy which makes the claim at initial capital $V_0(a^*)$ acceptable in the sense of criterion (5.2.1).*

Proof. For the first statement, we follow the arguments of Föllmer and Leukert [32], which are based on the Neyman-Pearson lemma. For any a and any given strategy π, we can define an $\mathcal{F}_T$-measurable random variable $\phi \in [0,1]$ by

$$\big((g(X_T, Z_T) - a)^+ - Y_T^\pi\big)^+ = (1-\phi)(g(X_T, Z_T) - a)^+.$$

The variable ϕ has been called the success ratio in [32]. We are interested in the success ratio ϕ which minimizes the price of a superhedge, i.e. we want to minimize

$$\sup_{Q \in \mathcal{Q}} E^Q[\phi(g(X_T, Z_T) - a)^+],$$

where $\mathcal{Q}$ is the set of all equivalent martingale measures, under the condition that equation (5.2.3) is satisfied, i.e.

$$\hat{E}[\phi] \geq \frac{1 + a(1-\beta)}{E[(g(X_T, Z_T) - a)^+]},$$

where $\hat{E}$ denotes the expectation under the measure $\hat{P}$, defined by

$$\frac{d\hat{P}}{dP} = \frac{(g(X_T, Z_T) - a)^+}{E[(g(X_T, Z_T) - a)^+]}.$$

The existence of an optimal ϕ now follows by the same argument as in Föllmer and Leukert [32] which is based on the Neyman-Pearson lemma. The corresponding optimal strategy then is given as the superhedge of the claim $\phi(g(X_T, Z_T) - a)^+$.

For the second statement, let V_1 and V_2 be the minimal required capital for a_1 and a_2, respectively, where a_1 and a_2 are arbitrary real numbers, and let π_1 and π_2 be the corresponding strategies. Then (5.2.3) is satisfied for (a_1, π_1, V_1) as well as for (a_2, π_2, V_2), and for an arbitrary $t \in [0,1]$ we have

$$tf(a_1, \pi_1; V_1) + (1-t)f(a_2, \pi_2; V_2) \leq -(ta_1 + (1-t)a_2)(1-\beta).$$

By the convexity of the function $x \mapsto x^+$, it follows that the left hand side is larger than or equal to $f(ta_1 + (1-t)a_2, t\pi_1 + (1-t)\pi_2; tV_1 + (1-t)V_2)$, hence

$$f(ta_1 + (1-t)a_2, t\pi_1 + (1-t)\pi_2; tV_1 + (1-t)V_2) \leq -(ta_1 + (1-t)a_2)(1-\beta).$$

It follows that at capital $tV_1 + (1-t)V_2$, there exists a strategy such that this equation is satisfied, and the minimal capital must therefore be smaller or equal. The third statement immediately follows from (5.2.2). □

It follows that we can first minimize the required capital $V_0(a)$ for a fixed a and subsequently minimize this expression with respect to a. The first minimization problem is a problem of the type discussed in Föllmer and Leukert [33], namely the minimization of the capital required under an expected shortfall constraint. We can therefore apply the observations contained in [33].

It will sometimes be easier to consider a related problem, namely to minimize $f(a,\pi;V_0)$ with respect to π at a given initial capital V_0. It is clear that

$$f_{min}(a,V_0) := \min_{\pi} f(a,\pi;V_0) \tag{5.2.5}$$

is a nonincreasing function in V_0 for given a. If we obtain $f_{min}(a,V_0)$ for all a and V_0, we can choose the minimal V_0 such that

$$f_{min}(a,V_0) \leq -a(1-\beta).$$

An advantageous situation occurs if f_{min} is continuous in V_0. In this case, we can replace this inequality by the corresponding equality.

If we aim to minimize the CVaR at a given initial capital V_0, we can again apply equation (5.2.5). By Rockafellar and Uryasev [73], the function

$$a \mapsto a + \frac{1}{1-\beta} f_{min}(a,V_0)$$

is convex in a, and we can again minimize over all values of a.

It becomes clear that the essential minimization problem is (5.2.5), which is a problem of minimizing an expected shortfall in the sense of Föllmer and Leukert [33], and from which everything else follows. In the sequel, we will therefore focus on this problem. As in Föllmer and Leukert, [32] and [33], we reformulate the problem of minimizing the expected shortfall as a problem of maximizing a state-dependent utility function. We write

$$E[(g(X_T,Z_T) - Y_T^\pi)^+] = E[g(X_T,Z_T)] - E[g(X_T,Z_T) \wedge Y_T^\pi].$$

Therefore we can, instead of solving the minimization problem (5.2.5), maximize

$$E[(g(X_T,Z_T) - a)^+ \wedge Y_T^\pi] \tag{5.2.6}$$

under the condition that

$$B_0 E^Q[Y_T^\pi] \leq V_0$$

for all equivalent martingale measures Q, where B_0 is the value of the zero bond with expiry time T, which is here taken as numéraire. In the sequel, we will focus on (5.2.6) as objective function.

5.2.4 Insurance information at the end of the period

As already stated in the introduction, one general assumption of this chapter is that nonfinancial information is only available at the terminal time T. This assumption has not been used until now, but we will now use it for the rest of the chapter.

If insurance information arrives only at the end of the period, the idea is that, similarly as in Föllmer and Leukert [32], we can integrate out the insurance random variable. For any strategy we follow up to time T^-, the objective function at time T^- is given by

$$E[(g(X_T, Z_T) - a)^+ \wedge Y_T^\pi | X_{T^-}, Y_{T^-}^\pi].$$

By the Markov property and the predictability of X_t and Y_t^π, this expression is equal to $\alpha(X_T, Y_T^\pi)$, where

$$\alpha(x, y) := \int \big((g(x, z) - a)^+ \wedge y\big)\, dP(z), \tag{5.2.7}$$

and where $dP(z)$ is the distribution function of Z. By assumption, Z is independent of $\mathcal{F}_{T^-}$, and therefore in particular independent of the strategy. The function α is concave in the second argument, and the optimization problem can now be reformulated as

$$\max_\pi E[\alpha(X_T, Y_T^\pi)] \tag{5.2.8}$$

subject to

$$B_0 E^Q[Y_T^\pi] \le V_0.$$

Because this problem is independent of $\sigma(Z)$, we are in a complete model, and the martingale measure Q is unique. Because the number of insured persons has a maximum, we can, in life insurance, mostly assume that

$$u(x) := \sup_z g(x, z) < \infty$$

and that

$$V_{sup} := B_0 E^Q[u(X_T)] < \infty.$$

In this case, a superhedge is possible at a finite initial capital, and we can apply to a large extent the theory developed in Föllmer and Leukert, [32] or [33]. We will explain this procedure in the subsequent section.

Remark 5.2.16. Even if we have reduced the problem to a problem of a complete market, the results of Sekine [80] are not applicable here. First, the risk measures are not the same. In [80], the author minimizes the expression, with our terminology,

$$\sup_{A \in \mathcal{F}: P(A) \ge 1-\beta} E\left[\left(g - V_0 - \int_0^T \pi dS\right)^+ \,|\, A\right]$$

which is not equal to CVaR and is in fact not a coherent risk measure due to the $(\cdot)^+$ function in the expectation.

Moreover, if we would integrate out the random variable Z_T in the framework of [80], this must be done for all test probability measures $\tilde{P}$ which give the dual representation of the CVaR risk measure. To be specific, for the sets A from above, the measures $\tilde{P}$ have the densities $\frac{d\tilde{P}}{dP} = \frac{1_A}{P(A)}$. The expected payoff then is (with x the state variable, y the terminal wealth)

$$\tilde{g}(x, y) = -\int g(x, z) d\tilde{P}(z) + y.$$

This is a payoff in a complete market, but it depends on the test probability measure. As a consequence, it does not seem obvious how to translate the idea of [80] to our setting.

5.3 Continuous probability distributions

5.3.1 Assumptions

In this section, we study the case of continuous insurance probabilities, which in some way leads to a simplified problem. The general problem is then studied in the subsequent sections.

Throughout this section, we make the following assumptions:

Assumption 5.3.1. The insurance variable Z_T has a continuous distribution, and the law of the financial variable X_T is absolutely continuous with respect to the Lebesgue measure.

Assumption 5.3.2. For almost every x, the payoff $g(x, Z_T)$ has a continuous distribution. By "almost every x", we refer to the law of X_T.

Assumption 5.3.3. There exists a function $H : \mathbb{R} \to \mathbb{R}_+$ such that $H(X_T)$ is integrable and for almost every x, $g(x, Z_T)$ is contained in $[0, H(x)]$.

Assumption 5.3.4. The Radon-Nikodym density of the financial market admits small values, i.e. for any $\epsilon > 0$ the set $\{\frac{dQ}{dP} < \epsilon\}$ has strictly positive probability.

Assumption 5.3.5. There exists a measurable function $q(x)$ such that the Radon-Nikodym density of the financial market is equal to $q(X_T)$. Furthermore, all level sets of $q(X_T)$ have zero measure, i.e. $P[q(X_T) = c] = 0$ for all constants $c \geq 0$.

Remark 5.3.6. All these assumptions are satisfied in the Black-Scholes market.

For being able to fully explore the continuity, we need some technical lemmas.

Lemma 5.3.7. *Let X and Z be two independent random variables and $g(x,z)$ a measurable function. If for almost every x, $g(x,Z)$ has a continuous distribution, then the distribution of $g(X,Z)$, i.e. $F(y) := P[g(X,Z) \leq y]$, is continuous too.*

Proof. Define $f(x;y) := P[g(x,Z) \leq y]$. Then $f(X,y)$ is a version of the conditional probability $P[g(X,Z) \leq y \mid \sigma(X)]$. For any $y \in \mathbb{R}$, let y_n be a sequence converging to y. Then $f(X;y_n) \to f(X,y)$ almost surely by the continuity of the distribution of $g(x,Z)$. But for any $n \in \mathbb{N}$, $|f(X;y_n)| \leq 1$, since $f(X;y_n)$ is a probability. By the dominated convergence theorem, it follows that

$$P[g(X,Z) \leq y_n] = E[f(X;y_n)] \to E[f(X;y)] = P[g(X,Z) \leq y].$$

This proves the lemma. □

Lemma 5.3.8. *For any $x \in \mathbb{R}$, let $f(x,\cdot) : \mathbb{R}_+ \to \mathbb{R}$ be a nonincreasing function, $d\nu(x)$ a measure which is absolutely continuous with respect to the Lebesgue measure, and $\int f(x,0^+)d\nu(x) < \infty$. Let $g(x)$ be a function with zero $d\nu$-measure on all level sets, i.e. for any $c \in \mathbb{R}$, $\nu(\{x \mid g(x) = c\}) = 0$. Then the function*

$$\gamma \mapsto \int f(x,\gamma g(x))d\nu(x)$$

is continuous.

Proof. Because f is nonincreasing in y, there are at most countably many points of discontinuity. For any $\gamma > 0$, let γ_n be a sequence converging to γ. Then, for $d\nu(x)$-almost all x, the sequence $(f(x,\gamma_n g(x)))_{n\geq 1}$ converges to $f(x,\gamma g(x))$, because convergence can only fail at countably many values of $g(x)$, and by the assumption that it has zero $d\nu$-measure on level sets, the points x at which convergence fails also build a $d\nu(x)$-null set. By the fact that $f(x,0^+)$ is integrable, one can apply the dominated convergence theorem which yields the result. □

5.3.2 Calculation of CVaR price

Theorem 5.3.9. *Let the insurance and financial processes Z and X as well as the insurance claim $g(X_T,Z_T)$ satisfy the model assumptions stated in 5.2.1 and 5.2.4. Let furthermore Assumptions 5.3.1-5.3.5 be valid. Then the CVaR price as formulated in Definition 5.2.4 is given by*

$$V = E^Q[I_a(X_T,\gamma q(X_T))], \tag{5.3.9}$$

where I_a is the inverse function given by

$$I_a(x,y) := \inf\{z \geq 0 \mid P[g(x,Z_T) > z + a] < y\}, \tag{5.3.10}$$

a is given by

$$a = -\frac{E\left[(g(X_T, Z_T) - I_0(X_T, \gamma q(X_T)))^+\right]}{1 - \beta - P[q(X_T) > \frac{1}{\gamma}]}, \tag{5.3.11}$$

and γ is determined by the equation

$$\gamma = 1 - \beta + \epsilon_\gamma \tag{5.3.12}$$

with

$$\epsilon_\gamma = \gamma E[q(X_T) 1_{q(X_T) > \frac{1}{\gamma}}] - P[q(X_T) > \frac{1}{\gamma}],$$

or, otherwise formulated,

$$\gamma = \frac{1 - \beta - P[q(X_T) > \frac{1}{\gamma}]}{Q[q(X_T) < \frac{1}{\gamma}]}. \tag{5.3.13}$$

Remark 5.3.10. From (5.3.12) it follows that γ only depends on the financial process X, on T, and on the value β. In particular it depends neither on the insurance process Z nor on the payoff $g(x, z)$.

Remark 5.3.11. For many reasonable parameter values, we have $\epsilon \approx 0$, and equation (5.3.12) gives even an explicit equation for γ, which does not need recursion to solve. The reason is that the CVaR level β is typically near to 1, so that γ is small. It follows that $q(X_T) > \frac{1}{\gamma}$ occurs only for a few events, with respect to probability Q as well as to probability P.

Remark 5.3.12. By (5.3.13), the parameter γ does not depend on the constraint (5.3.11). It follows that minimizing the initial wealth such that the corresponding $CVaR$ is smaller than or equal to any constant would lead to the same parameter γ. This also implies that minimization of $CVaR$ with any given initial wealth V_0 leads to the same value of γ.

The proof of Theorem 5.3.9 is based on a theorem stated in Föllmer and Leukert [33], which we present here in an abbreviated version. It gives a solution for the optimization problem

$$u(z) = \sup_Z \{E[U(Z, \cdot)] | 0 \leq Z \leq H \quad \text{and} \quad E^*[Z] \leq z\}, \tag{5.3.14}$$

where $U(Z, \cdot)$ is a state-dependent utility function which is nondecreasing and concave in the second variable, as well as strictly concave and differentiable on $]0, H(\omega)[$, and E^* is the expectation under the risk-neutral measure, which is unique in the complete market. Here we use the same notations as in [33].

Theorem 5.3.13 (Theorem 7.1 of [33])**.** *For each $z \leq E^*[H]$ there is a unique solution $\tilde{Z}$ of the equation $u(z) = E[U(\tilde{Z}, \cdot)]$. It takes the form*

$$\tilde{Z}(\omega) = I(y(z)\rho^*(\omega), \omega) \wedge H(\omega),$$

where $y(z)$ is the solution of

$$E^*[I(y(z)\rho^*(\omega), \omega) \wedge H(\omega)] = z.$$

The function ρ^* plays the role of the density of the risk neutral measure with respect to the original one, and the function I is the inverse of U', i.e.

$$I(y, \omega) = \inf\{z \in [0, H(\omega)] | U'(z, \omega) < y\}.$$

Remark 5.3.14. The theorem also holds if the concavity is not strict, as it will be the case in our application. Only the uniqueness may fail, but this does not affect the statements of Theorem 5.3.9. We will omit the proof of Theorem 5.3.13 in the case of the relaxed assumption, because this situation will be treated in the more general Theorem 5.4.2, which does not even need differentiability of U.

For proving Theorem 5.3.9, we first prove another lemma:

Lemma 5.3.15. *Let the function α be given by*

$$\alpha(y) := E[Y \wedge y],$$

where Y is a random variable with continuous distribution. Then

$$\frac{\partial}{\partial y}\alpha(y) = P[Y > y]. \tag{5.3.15}$$

In particular, α is differentiable and concave, and strictly concave as long as the density of Y exists and is strictly positive.

Proof. Assume that $\Delta y > 0$; the other cases can be proved by a similar argument. Then

$$\begin{aligned}\tfrac{1}{\Delta y}\left(\alpha(y + \Delta y) - \alpha(y)\right) \quad &= \quad \tfrac{1}{\Delta y}\textstyle\int_{Y \in [y, y+\Delta y]} (Y - y)\, dP(Z_T) \\ &\quad + \textstyle\int_{Y > y+\Delta y} dP(Z_T).\end{aligned}$$

The second expression on the right hand side is equal to $P[Y > y + \Delta y]$, and the first one can be estimated by

$$\frac{1}{\Delta y} \mid \int_{Y \in [y, y+\Delta y]} (Y - y)\, dP(Z_T) \mid \leq \int_{Y \in [y, y+\Delta y]} dP(Z_T).$$

As $\Delta y \to 0$, the latter expression converges to 0 by the continuity of the distribution of Y and the former one to $P[Y > y]$.

If the density of Y exists and is strictly positive, the function $y \mapsto P[Y > y]$ is strictly decreasing, and $\alpha(y)$ is strictly concave. □

Proof of Theorem 5.3.9. Let V_0 be the CVaR price. Then, by the considerations in section 5.2.3, there exist an $a \in \mathbb{R}$ and a strategy π^* such that $f_{min} \leq -a(1-\beta)$, where as in (5.2.5),

$$f_{min} = E[\Big(g(X_T, Z_T) - a - Y_T^{\pi^*}\Big)^+].$$

By (5.2.7) and (5.2.8), this can be written as

$$f_{min} = E[g(X_T, Z_T) - a] - \max_{\pi} E[\alpha(X_T, Y_T^{\pi})].$$

By Lemma 5.3.15, the function $\alpha(x, y)$ is differentiable and concave in y with the derivative given by $\frac{\partial \alpha}{\partial y}(x, y) = P[g(x, Z_T) > y + a]$. Furthermore, the second term involves an optimization problem in a complete market setting. Theorem 5.3.13 then states that, for a fixed a, there exists a constant γ such that

$$f_{min}(a, V_0) = E[g(X_T, Z_T) - a] - E[\alpha\left(X_T, I_a(X_T, \gamma q(X_T))\right)].$$

By the convexity and the finiteness of f_{min} in a, it is clear that it is continuous in a. We will now show that it is also continuous in V_0. First, f_{min} is continuous in γ by Lemma 5.3.8, since $h(x, y) := \alpha(x, I_a(x, y))$ is nonincreasing in y. Furthermore, by Lemma 5.3.8, $V_0(\gamma) := E^Q[I_a(X_T, \gamma q(X_T))]$ is continuous as well, and strictly decreasing for $\gamma > 0$. Indeed, $I_a(x, y)$ is strictly decreasing in y for $y < 1$ and $I_a(x, y) = 0$ for $y \geq 1$. Because of Assumption 5.3.4, the set of all X_T with $q(X_T) < \frac{1}{\gamma}$ has positive Q-probability, and therefore $V_0(\gamma)$ is strictly decreasing by the monotonicity of the expectation. The inverse $\gamma(V_0)$ is therefore continuous, and hence also $f_{min}(a, V_0)$ in V_0.

We now claim that

$$E[(g(X_T, Z_T) - a - I_a(X_T, \gamma q(X_T)))^+] = -a(1-\beta). \tag{5.3.16}$$

To prove this, note that if (5.3.16) would be a strict inequality, by continuity there would be a $V < V_0$ satisfying $f_{min}(a, V) \leq -a(1-\beta)$. In other words, there would exist an allowed strategy at initial wealth V, and therefore V_0 could not be the CVaR price.

From equation (5.3.10), one can conclude that

$$I_a(x, y) = \begin{array}{lll} I_0(x, y) - a & \text{if} & y \leq 1, \\ 0 & \text{if} & y > 1. \end{array}$$

Therefore equation (5.3.16) is equivalent to

$$-a(1-\beta) = E\left[(g(X_T, Z_T) - I_0(X_T, \gamma q(X_T)))^+\right] - aP[q(X_T) > \frac{1}{\gamma}].$$

This equation can be solved for a and yields (5.3.11).

By Assumptions 5.3.1 and 5.3.2 and Lemma 5.3.7, the random variable $R := g(X_T, Z_T) - I_a(X_T, \gamma q(X_T))$ has a continuous distribution. This means that the Value-at-Risk a^* at level β of R is given by

$$P[R > a^*] = P[R \geq a^*] = 1 - \beta.$$

We will show that $f_{min}(a^*, V_0) = -a^*(1-\beta)$. We know that this equality is satisfied for a instead of a^*, and by Rockafellar and Uryasev [72], a^* gives a minimum and therefore $f_{min}(a^*, V_0) \leq -a^*(1-\beta)$. On the other hand, let the inequality be strict. By the continuity of f_{min} in V, there exists a $V < V_0$ such that the inequality is still satisfied for V. But this again implies that there is an allowed strategy at the price $V < V_0$, and V_0 could not be the CVaR price. We can therefore take $a = a^*$. By Assumptions 5.3.1 and 5.3.2, we have

$$P[g(x, Z_T) > a + I_a(x, y)] = y \wedge 1.$$

It follows that

$$1 - \beta = P[R > a] = P[g(X_T, Z_T) - I_a(X_T, \gamma q(X_T)) > a] = E[\gamma q(X_T) \wedge 1].$$

Equations (5.3.12) and (5.3.13) now follow immediately. □

5.3.3 Algorithm

Theorem 5.3.9 now suggests the following algorithm for calculating the CVaR price:

1. Calculate γ from (5.3.13).
2. Determine $I_0(x, y)$ by (5.3.10).
3. Determine a by equation (5.3.11).
4. Determine the CVaR price by equation (5.3.9).

Remark 5.3.16. Even if we originally have an optimization problem, this algorithm is a straightforward calculation and does not require any optimization algorithm anymore.

5.3.4 Application to a unit-linked insurance model

We now apply Theorem 5.3.9 to the case of a unit-linked survival insurance, where the stock price S_T is paid at time T if the person is still alive, and nothing is paid if the insured has died before this time. The process S_t is assumed to follow a geometric Brownian motion, i.e.

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{5.3.17}$$

and the amount of survivors is assumed to follow a binomial distribution. It is further assumed that the insurance outcome is independent of the Brownian motion process. For simplicity, we assume the risk-free interest rate to be zero. If n is the amount of insured persons and p the probability of surviving, then the amount of survivors N_T is

$$N_T \sim BIN(n,p). \tag{5.3.18}$$

The total insurance payoff is

$$g(S_T, N_T) = S_T N_T. \tag{5.3.19}$$

We furthermore assume that the information about the survivors is first revealed at time T. For large values of n, it seems reasonable to approximate the amount of survivors by a truncated normal distribution, i.e.

$$N_T \sim \mathcal{N}_{trunc}(np, np(1-p)). \tag{5.3.20}$$

The density of the truncated normal distribution is defined by

$$f_{trunc}(z) := \begin{array}{lll} 0 & \text{if} & z < 0 \\ cf(z) & \text{if} & 0 \leq z \leq n \\ 0 & \text{if} & z > n, \end{array} \tag{5.3.21}$$

where $f(z)$ is the density of the normal distribution with mean np and variance $np(1-p)$ and c is a normalization constant. With equation (5.3.15), one has

$$\frac{\partial}{\partial y}\alpha(x,y) = P[N_T x > y + a] = 1 - c\left(\Phi\left(\frac{\frac{y+a}{x} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{-np}{\sqrt{np(1-p)}}\right)\right) \wedge 1$$

with $c^{-1} = \Phi\left(\sqrt{\frac{n(1-p)}{p}}\right) - \Phi\left(-\sqrt{\frac{np}{1-p}}\right)$.

As can be checked for the Black-Scholes model (5.3.17), the Radon-Nikodym density is given by $q(S_T) = S_0^{\frac{\mu}{\sigma^2}} e^{\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 T - \frac{\mu}{2}T} S_T^{-\frac{\mu}{\sigma^2}}$.

Let us now apply Theorem 5.3.9 to this unit-linked insurance model. By (5.3.12), γ is given by

$$\gamma = 1 - \beta + \epsilon, \tag{5.3.22}$$

where

$$\epsilon = \gamma\Phi\left(\frac{\ln\left(\frac{c_0}{S_0}\right) + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{\ln\left(\frac{c_0}{S_0}\right) - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}\right)$$

Table 5.1: Minimal capital for different parameters, $\beta = 95\%$

n	p	γ	a	V_0	*Load*
1000	0.5	0.05	−6.62	532.61	6.5%
1000	0.1	0.05	−3.97	119.56	19.6%
50	0.5	0.05	−1.48	32.29	29.2%
50	0.1	0.05	−0.89	9.38	87.6%

with

$$c_0 := \gamma^{\frac{\sigma^2}{\mu}} S_0 e^{\frac{1}{2}(\mu-\sigma^2)T}.$$

By (5.3.10), we have

$$I_a(x,y) = \begin{cases} nx - a & \text{if} \quad y = 0, \\ npx - a + \sqrt{np(1-p)}x\Phi^{-1}\left(\frac{1-y}{c} + \Phi\left(-\sqrt{\frac{np}{1-p}}\right)\right) & \text{if} \quad 0 < y < 1, \\ 0 & \text{if} \quad y \geq 1. \end{cases} \tag{5.3.23}$$

By (5.3.11), we must calculate by numerical integration

$$a = -\frac{E[f(S_T)]}{1 - \beta - \Phi\left(\frac{\ln\left(\frac{c_0}{S_0}\right) - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}\right)}, \tag{5.3.24}$$

where $f(x)$ is determined by integration of (5.3.11),

$$\begin{aligned} f(x) \quad = \quad & \sqrt{np(1-p)}x \times \\ & \left[\frac{c}{\sqrt{2\pi}}\left(e^{-\frac{1}{2}\left(\Phi^{-1}\left(\frac{1-\gamma q(x)}{c}+\Phi_0\right)\right)^2} - e^{-\frac{n(1-p)}{2p}}\right) - \gamma q(x)\Phi^{-1}\left(\frac{1-\gamma q(x)}{c} + \Phi_0\right)\right] \end{aligned}$$

if $\gamma q(x) < 1$ and npx otherwise, where $\Phi_0 = \Phi(-\sqrt{\frac{np}{1-p}})$. Finally, we can now again perform a numerical integration in order to solve (5.3.9).

5.3.5 Numerical results

We solved the problem numerically using the parameters $\mu = 0.07$, $\sigma = 0.2$, $T = 1$, $S_0 = 1$, $\beta = 0.95$. The resulting parameter γ from formula (5.3.22) is $\gamma = 0.05$ with $\epsilon = 0$, and one can check that indeed $\epsilon = 5.0 \times 10^{-20} \approx 0$ for this choice of parameters. The results for the CVaR price are given in Table 5.1. The last column, *Load*, is the risk loading as percentage of the pure premium. The optimal hedge then is the complete delta hedge of the payoff $I_a(S_T, \gamma S_0^{\frac{\mu}{\sigma^2}} e^{\frac{T}{2}(-\mu+(\frac{\mu}{\sigma^2})^2)} S_T^{-\frac{\mu}{\sigma^2}})$, with a and the corresponding γ given in Table 5.1.

5.4 Optimal solution for nonsmooth state dependent utility function

In typical situations of life insurance, the distribution function for the amount of survivors is discontinuous. Furthermore, among the assumptions of section 5.3.1, in particular Assumption 5.3.2 is too restrictive. Applying (5.2.7) to discrete distributions leads to a function $\alpha(x, y)$ which is piecewise linear in y and therefore not differentiable nor strictly concave. We can therefore not apply Theorem 5.3.13.

In this section, we essentially prove an extension of Theorem 5.3.13 to the case of discrete probabilities. However, for this we use a formulation which resembles Proposition 5.14 in Föllmer and Leukert [32], because this is more closely related to our application. The result then shows how to find the optima. The extension with respect to [32] consists of dropping the requirement that the state dependent utility function is strictly concave and differentiable.

Here we first state a condition which will often be used throughout this section. We denote this condition by FHFC (full hedge with finite capital), because, economically speaking, this condition means that one can make a full hedge using only a finite amount of capital. In life insurance, this condition is mostly satisfied, because the assumption that everyone survives describes the worst case, or the assumption that everyone dies in the case of a death insurance.

Definition 5.4.1. A function $\alpha(x, y)$ is said to satisfy the *FHFC condition with respect to the probability measure* μ on the Borel set $\mathcal{B}(A)$, $A \subset \mathbb{R}^n$, if there exists a measurable and μ-integrable function $h(x) > 0$ such that $\sup_y \alpha(x, y) = \alpha(x, h(x))$.

Theorem 5.4.2. *Let $A \subset \mathbb{R}^n$ an interval, and let ν and μ be two finite equivalent measures on $\mathcal{B}(A)$. Let $\alpha : D := A \times [0, \infty) \to \mathbb{R}$ be a function which is concave, nondecreasing in the second argument, and satisfies the FHFC condition with respect to μ, and let $\alpha(x, h(x))$ be ν-integrable. Define $\alpha(x, y) := -\infty$ for all $y < 0$, so that the concavity holds for all real numbers.*

Let $v : A \to [0, \infty)$ be a function on C, where C is the set of all Borel-measurable functions

$$f : A \to \mathbb{R}$$

with the property that

$$||f|| := \sup_{x \in A} \left| \frac{f(x)}{h(x) + 1} \right| < \infty, \tag{5.4.25}$$

where $h(x)$ is the function of the FHFC condition.

Then the following statements are equivalent:

1. *There exists a function $\beta : D \to \mathbb{R}$ such that for each $(x, y) \in D$, $\beta(x, y)$ is a point in the superdifferential of α with respect to the second argument, and such that*
$$\beta(x, v(x)) = \gamma \frac{d\mu}{d\nu}(x) \tag{5.4.26}$$
for a constant $\gamma > 0$.

2. *The function $v(x)$ optimizes*
$$\int_A \alpha(x, f(x)) d\nu(x)$$
over all functions $f \in C$ with
$$\int_A f(x) d\mu(x) \le \int_A v(x) d\mu(x). \tag{5.4.27}$$

Remark 5.4.3. Economically speaking, this is a functional analytical version of the statement that in the optimum, the price is proportional to the marginal utility, where μ is the pricing functional of f, and $\nu(\alpha(\cdot, f))$ is its utility.

Proof. We first prove the easy direction from (1) to (2). Let $f(x)$ be any nonnegative function on C, where C is the set defined in Theorem 5.4.2. We define the function
$$f_\lambda(x) := (1 - \lambda)v(x) + \lambda f(x).$$
It is clear that for all $1 \ge \lambda \ge 0$, $f_\lambda \in C$ and f_λ is nonnegative. By the concavity of α in the second argument, we have, for any choice $\beta(x, y)$ of the superdifferential, that
$$\alpha(x, f_\lambda(x)) - \alpha(x, f_0(x)) \le \beta(x, f_0(x))(f(x) - v(x))\lambda.$$
Therefore, because α satisfies FHFC and $\alpha(x, h(x))$ is integrable, it follows by $\alpha(x, f_\lambda(x)) \le \alpha(x, h(x))$ that $\alpha(x, f_\lambda(x))$ is integrable for all $\lambda \ge 0$ and that
$$\int_A \alpha(x, f(x)) d\nu(x) - \int_A \alpha(x, v(x)) d\nu(x) \le \int_A \beta(x, v(x))(f(x) - v(x)) d\nu(x). \tag{5.4.28}$$
Now let the superdifferential satisfy property (5.4.26). Then it follows for the right hand side of equation (5.4.28) that
$$\int_A \beta(x, v(x))(f(x) - v(x)) d\nu(x) = \gamma \left(\int_A f(x) d\mu(x) - \int_A v(x) d\mu(x) \right) \le 0,$$
where the last inequality follows if f satisfies property (5.4.27), and the integrability is again guaranteed by FHFC with respect to μ. It follows from (5.4.28) that
$$\int_A \alpha(x, f(x)) d\nu(x) \le \int_A \alpha(x, v(x)) d\nu(x),$$

and therefore $v(x)$ is optimal for all nonnegative functions $f \in C$. If $f(x) < 0$ on a set with ν-positive measure, $\alpha(x, f(x)) = -\infty$ on a set with ν-positive measure, and the integral is $-\infty$, which cannot be optimal.

Now let us turn to the other direction. Here we need functional analytical arguments from infinite-dimensional convex analysis. Note that C, with the norm from (5.4.25), is a convex normed vector space. Furthermore, the function F on C defined by

$$F(f) := \int_A \alpha(x, f(x))d\nu(x)$$

is a concave function, which easily follows from the concavity of α. The function $f \mapsto \int_A f(x)d\mu(x)$ is a continuous linear functional on C. Furthermore, if $v(x)$ is not identically zero, the Slater condition is satisfied, and there exists a point f such that F is continuous in f, for example $f(x) = 1 + h(x)$. For applying the Kuhn-Tucker theorem, it remains to show, by Theorem 9.6.1 of Attouch et al. [4], that $F(f)$ is closed. We will show that the set $\{F(f) < \tilde{\alpha}\}$ is open for all $\tilde{\alpha} \in \mathbb{R}$. Indeed, let first f be nonnegative. Then, by definition of C, for $g \in C$ with $||g - f|| \leq t$,

$$g(x) \leq f(x) + t(1 + h(x)).$$

Furthermore, as $t \downarrow 0$, $\alpha(x, f(x) + t(1 + h(x))) \downarrow \alpha(x, f(x))$ almost surely, by the fact that α is nondecreasing and right-continuous in the second variable for $y \geq 0$. By FHFC, $\alpha(x, f(x) + t(1 + h(x))) \leq \alpha(x, h(x))$, and $\alpha(x, h(x))$ is ν-integrable. By the dominated convergence theorem, it follows that $F(f + t(1+h)) \downarrow F(f)$. For any $\epsilon > 0$, we can therefore find a $\delta > 0$ such that $F(g) \leq F(f + t(1 + h)) < F(f) + \epsilon$ for all $||g - f|| \leq t < \delta$. Because $F(f) < \tilde{\alpha}$, we find an $\epsilon > 0$ such that $F(f) + \epsilon < \tilde{\alpha}$, and therefore $F(g) < \tilde{\alpha}$ for all g with $||f - g|| < \delta$. If f is not nonnegative, there exists a set $A' \subset A$, $\nu(A') > 0$, such that $f(x) \leq -\epsilon < 0$ on A'. Because $h(x)$ is finite, we can furthermore find a further subset with nonzero $d\nu$-measure, again denoted by A', such that $h(x) \leq K < \infty$ for all $x \in A'$. Now choose g such that $||f - g|| < \frac{\epsilon}{2K+2}$. Then, on the set A',

$$g(x) \leq f(x) + \frac{\epsilon}{2K + 2}(1 + h(x)) \leq -\epsilon + \frac{\epsilon}{2} < 0.$$

It follows that $\alpha(x, g(x)) = -\infty$ on a set with positive $d\nu$-measure, and therefore $F(g) = -\infty < \tilde{\alpha}$ for all $\tilde{\alpha}$. As a consequence, $\{F(f) < \tilde{\alpha}\}$ is open, from which it follows that F is closed.

By the Kuhn-Tucker theorem in infinite dimensions (Theorem 9.6.1 of Attouch et al. [4]), there exists a continuous linear functional $\phi \in \delta F(v)$ in the superdifferential of F and a constant $\gamma > 0$ such that

$$\phi = \gamma\mu.$$

By the fact that μ is absolutely continuous with respect to ν, it follows that

$$\phi(f) = \int_A \gamma \frac{d\mu}{d\nu}(x) f(x) d\nu(x),$$

which means that ϕ is even contained in $L^1(A, \nu)$.

Now we define, for a function $f \in C$, the new function

$$g(t) := F(v + tf).$$

Then $\phi(f)$ must be an element of the superdifferential of $\delta g(0)$, because we have

$$g(t) - g(0) = F(v + tf) - F(v) \leq \phi(tf) = t\phi(f)$$

by the fact that ϕ is contained in the superdifferential of F. Now choose f such that g is continuous in 0 (that is $v + tf \geq 0$ for $|t|$ small enough). Then

$$g(t) - g(0) = \int_A [\alpha(x, v(x) + tf(x)) - \alpha(x, v(x))] \, d\nu(x) \leq t\phi(f) \ \forall t \in B_\epsilon(0)$$

must be satisfied. But for $t \downarrow 0$, we have

$$\lim_{t \downarrow 0} \frac{g(t) - g(0)}{t} = \lim_{t \downarrow 0} \int_A \frac{1}{t} [\alpha(x, v(x) + tf(x)) - \alpha(x, v(x))] \, d\nu(x)$$

The integrand converges $d\nu$-a.s. to

$$\beta_-(x, v(x)) f^+(x) - \beta_+(x, v(x)) f^-(x),$$

where $\beta_-(x, y)$ is the right limit of the difference quotient of $\alpha(x, y)$ in y, and $\beta_+(x, y)$ the left limit, and f^+ and f^- are the nonnegative and nonpositive parts of f, respectively. Similarly, for $t \uparrow 0$, the integrand converges to

$$\beta_+(x, v(x)) f^+(x) - \beta_-(x, v(x)) f^-(x).$$

By the fact that $g(t)$ is concave and that there exists an $\epsilon > 0$ with $g(-\epsilon) > -\infty$ the dominated convergence theorem yields

$$\begin{aligned} \phi(f) \ \in \ & [\textstyle\int_A (\beta_-(x, v(x)) f^+(x) d\nu(x) - \beta_+(x, v(x)) f^-(x) d\nu(x)), \\ & \textstyle\int_A (\beta_+(x, v(x)) f^+(x) d\nu(x) - \beta_-(x, v(x)) f^-(x) d\nu(x))]. \end{aligned} \tag{5.4.29}$$

Now let $\hat{\phi}(x) = \gamma \frac{d\mu}{d\nu}(x)$ be the function with $\phi(f) = \int_A \hat{\phi}(x) f(x) d\nu(x)$. Assume that on a set $A' \subset A$ with $v(x) > 0$ on A' and $\nu(A') > 0$, we have $\hat{\phi}(x) > \beta_+(x, v(x))$. Then there exists a subset of A', again denoted by A', on which $v(x) \geq \epsilon > 0$, with $\nu(A') > 0$. The function $1_{A'}$ is obviously contained in C and nonnegative, and for this function, $g_{1_{A'}}(t)$ is continuous at 0. It follows from (5.4.29) that

$$\phi(1_{A'}) \leq \int_{A'} \beta_+(x, v(x)) d\nu(x).$$

But by definition,

$$\begin{aligned}\phi(1_{A'}) &= \textstyle\int_{A'} \hat{\phi}(x)d\nu(x) = \int_{A'} \beta_+(x, v(x))d\nu(x) \\ &\quad + \textstyle\int_{A'} \left(\hat{\phi}(x) - \beta_+(x, v(x))\right) d\nu(x) > \int_{A'} \beta_+(x, v(x))d\nu(x) \geq \phi(1_{A'}),\end{aligned}$$

a contradiction. In the same way, the assumption $\hat{\phi}(x) < \beta_-(x, v(x))$ leads to a contradiction on $\{v(x) > 0\}$.

On $\{v(x) = 0\}$, if $f(x) \geq 0$ but $g(t)$ is not necessarily continuous at 0, we can still take the right limit $g(t) \downarrow 0$, and the monotone convergence theorem yields

$$\phi(f) \geq \int_A \beta_-(x, v(x)) f(x) d\nu(x).$$

Furthermore, we have $\beta_+(x, 0) = \infty$ by definition of α. Note that $\beta_-(x, 0) < \infty$, because otherwise, defining $f(x) := 1_{v(x)=0}(x)$, we would have $\phi(f) = \infty$, a contradiction to the continuity of ϕ. Now assume that $\hat{\phi}(x) < \beta_-(x, v(x))$ on a subset A' of $\{v(x) = 0\}$, with $\nu(A') > 0$. Again,

$$\begin{aligned}\textstyle\int_{A'} \hat{\phi}(x)d\nu(x) &= \textstyle\int_{A'} \beta_-(x, v(x))d\nu(x) - \int_{A'} \left(\beta_-(x, v(x)) - \hat{\phi}(x)\right) d\nu(x) \\ &< \textstyle\int_{A'} \beta_-(x, v(x))d\nu(x) \leq \phi(1_{A'}),\end{aligned}$$

which is a contradiction. It follows that in all cases,

$$\hat{\phi}(x) = \gamma \frac{d\mu}{d\nu}(x) \in [\beta_-(x, v(x)), \beta_+(x, v(x))].$$

For any $x \in A$, defining $y = v(x)$, we can therefore find a point $\beta(x, y)$ in this interval such that (5.4.26) holds. But by definition, this interval precisely coincides with the superdifferential of $\alpha(x, y)$ at the point y. For a y for which no x exists with $v(x) = y$, we can choose an arbitrary point $\beta(x, y)$ in the superdifferential of α. It follows that the point $\beta(x, y)$ defined in this way satisfies property (5.4.26), and the theorem now is completely proved. □

Remark 5.4.4. For a continuous and strictly concave function and with $\mu = \nu$, this theorem is essentially Proposition 5.14 of Föllmer and Leukert [32].

Remark 5.4.5. If $\alpha(x, y)$ is strictly concave, Theorem 5.4.2 gives an algebraic equation from which we can find the optimal function $v(x)$.

Indeed, the optimal function exists, which can be proved in the same way as in Föllmer and Leukert [32]. Theorem 5.4.2 is therefore just an answer to the question of how an optimum may be found. Similarly to [32], we also state the existence theorem.

Theorem 5.4.6. *Let μ and ν be two finite equivalent measures, and let $\alpha(x,y)$ be given as in Theorem 5.4.2. Let $V_0 < \mu(h)$ be larger than 0. Then there exists a measurable function $v(x) \in C$, C defined as in Theorem 5.4.2, such that*

$$\int_A \alpha(x, v(x))d\nu(x) = \sup_f \int_A \alpha(x, f(x))d\nu(x),$$

where the supremum is taken over all measurable functions f with

$$\mu(f) \leq V_0.$$

Proof. Because $\alpha(x,y) = -\infty$ for $y < 0$, we can restrict the discussion to nonnegative functions f. Furthermore, we can restrict ourselves to functions contained in C, because $h(x) \in C$, and because any nonnegative measurable function f satisfies

$$\int_A \alpha(x, f(x))d\nu(x) = \int_A \alpha(x, f(x) \wedge h(x))d\nu(x),$$

so that f even can be chosen bounded by $h(x)$. Furthermore, we can choose f such that $\mu(f) = V_0$, because if $\mu(f) < V_0 < \mu(h)$, $f_t(x) := f(x) + t(h(x) - f(x))$ is for $t \in [0,1]$ still a nonnegative function bounded by h, and by the fact that $\alpha(x,y)$ is nondecreasing in y,

$$\int_A f_t(x)d\nu(x)$$

is a nondecreasing function in t, and

$$g(t) := \int_A f_t(x)d\mu(x)$$

is a continuous (linear) function with $g(0) = \mu(f)$ and $g(1) = \mu(h)$. By standard real analysis, there exists a $0 < t < 1$ with $\mu(f_t) = V_0$. If we now define the set

$$C' := \{0 \leq f(x) \leq h(x) : \mu(f) = V_0\},$$

then C' is a convex set which is weakly compact in $\mathcal{L}^1$, and we are precisely in the same situation as in Föllmer and Leukert [32]. The existence now follows by the same arguments. □

Corollary 5.4.7. *Let $(\Omega, \mathcal{F}_T, \mathcal{F}_t, P)$ be a filtered probability space and X_t a continuous semimartingale with values in a convex set $A \subset \mathbb{R}^n$. Assume that there exists a unique equivalent local martingale measure Q, and that $\frac{dQ}{dP}$ is $\sigma(X_T)$-measurable. Let $\alpha(x,y)$ satisfy the properties of Theorem 5.4.2, with μ and ν the laws of X_T under Q and P, respectively. Then the hedge which optimizes $E[\alpha(X_T, V_T)]$ at initial capital V_0 is given by the hedge of the claim $v(X_T)$, with v from Theorems 5.4.2 and 5.4.6.*

Proof. The proof follows similar arguments as in Föllmer and Leukert [32]. Let π be any admissible strategy, and let its value process be

$$V_t = V_0 + \int_0^t \pi_s dX_s.$$

At time T, we define the X_T-measurable random variable $f(X_T) := E[V_T|X_T]$. By the concavity of α in the second argument we have that

$$E[\alpha(X_T, V_T)] \leq E[\alpha(X_T, E[V_T|X_T])] = E[\alpha(X_T, f(X_T))].$$

But by the fact that $\frac{dQ}{dP}$ is $\sigma(X_T)$-measurable, we have

$$E^Q[f(X_T)] = E[\frac{dQ}{dP}E[V_T|X_T]] = E[E[\frac{dQ}{dP}V_T|X_T]] = E^Q[V_T] = V_0.$$

If v is optimal in the sense of Theorems 5.4.2 or 5.4.6, it follows that

$$\begin{aligned} E[\alpha(X_T, V_T)] &\leq E[\alpha(X_T, f(X_T))] = \int_A \alpha(x, f(x))d\nu(x) \\ &\leq \int_A \alpha(x, v(x))d\nu(x) = E[\alpha(X_T, v(X_T))], \end{aligned}$$

and therefore the replication of the claim $v(X_T)$ is optimal. □

5.5 CVaR pricing with discrete insurance probabilities

5.5.1 Problem formulation and assumptions

The idea of this section is to develop an algorithm for obtaining the CVaR price analogous to section 5.3, but without Assumptions 5.3.1 and 5.3.2. The first idea would be to apply Theorem 5.4.2 in the same way as Theorem 5.3.13 has been applied for proving Theorem 5.3.9. However, there is one further problem. The derivation of equations (5.3.12) and (5.3.13) relies on the assumption that $P[g(x, Z_T) - I_a(x, y) = a] = 0$, which is typically not satisfied when the distribution of $g(x, Z_T)$ is not continuous. Actually, there may be cases where this probability is even quite large as will be shown in the example later in this section.

However, the formulas (5.3.12) and (5.3.13) also hold in many cases where $P[g(x, Z_T) - I_a(x, y) = a] > 0$, so that the continuity of the distribution helps to prove the theorem, but it is not the key assumption. Here we will derive another argument, which relies on the classical first-order condition of a minimization problem, that is $V_0'(\gamma) = 0$ at the minimum, where the minimal price V_0 is calculated as a function of γ, and satisfying the constraint (5.3.16).

To develop general conditions under which Theorem 5.3.9 extends to the case of possibly discontinuous distributions is a technically delicate issue and would exceed the scope of this chapter. In particular, one would have to carefully analyze the boundary regions of sets $\{x : \gamma q(x) \in I\}$, where I is an interval in $[0, 1]$, and q again such that $q(X_T)$ is equal to the Radon-Nikodym density. We will prove a simplified version of this theorem under some additional assumptions, which cover at least the Black-Scholes model with a finite insurance state space.

Assumption 5.5.1. The insurance state space is finite, i.e. the set of all insurance outcomes is $\{Z_1, ..., Z_n\}$. We assume that all these outcomes have a positive probability.

Assumption 5.5.2. For the claim $g(x, Z)$, there is a uniform ordering, i.e. one can order $Z_1, Z_2, ..., Z_n$ in a way that

$$g(x, Z_1) \leq g(x, Z_2) \leq ... \leq g(x, Z_n)$$

uniformly for all x.

Assumption 5.5.3. The state space variable X_T has a density function which is continuous, under the original measure P as well as under the martingale measure Q.

Assumption 5.5.4. There exists a differentiable function $q(x)$ such that the Radon-Nikodym density of the financial market is given by $\frac{dQ}{dP} = q(X_T)$.

Assumption 5.5.5. The state space variable X_T is one-dimensional, and the density $q(x)$ is a strictly decreasing function, which surjective on $]0, \infty[$.

Assumption 5.5.6. The Radon-Nikodym density of the financial market admits small values, i.e. for any $\epsilon > 0$ the set $\frac{dQ}{dP} < \epsilon$ has strictly positive probability.

Remark 5.5.7. The assumption that the financial state variable is one-dimensional does not seem to be essential. However, it considerably facilitates the proof. Without this assumption, one precisely has the technical problems mentioned above.

Remark 5.5.8. In the Black-Scholes model, the Radon-Nikodym density is given by $q(X_T)$, where $q(x) = cx^{-\frac{\mu}{\sigma^2}}$ with a suitable constant c, and therefore Assumption 5.5.5 is satisfied.

Lemma 5.5.9. *Let Assumptions 5.5.1-5.5.2 be satisfied. Then the inverse function $I(x, y)$, defined in (5.3.10), is given by*

$$I(x, y) = \sum_{k=1}^{n} g(x, Z_k) 1_{y \in]P_{k+1}, P_k]} - a 1_{y<1}, \tag{5.5.30}$$

where the values $(P_k)_{1\leq k\leq n}$ *are given by*

$$P_k = P[\{Z_k, Z_{k+1}, .., Z_n\}]. \tag{5.5.31}$$

Proof. Let $y \in]P_{k+1}, P_k]$. Then, for $\xi < g(x, Z_k) - a$, it follows by Assumption 5.5.2 that $\xi < g(x, Z_j) - a$ for all $j \geq k$, and therefore $P[g(x, Z) - a > \xi] \geq P_k$. On the other hand, if $\xi > g(x, Z_k) - a$, then $\xi < g(x, Z_j) - a$ is only possible for $j \geq k+1$, and therefore $P[g(x, Z) - a > \xi] < y$ for all $y > P_{k+1}$. Consequently, the infimum of all these ξ's is $\xi = g(x, Z_k) - a$. Equation (5.5.30) follows. □

Lemma 5.5.10. *Let the insurance and financial processes* Z *and* X *as well as the insurance claim* $g(X_T, Z_T)$ *satisfy the model assumptions stated in sections 5.2.1 and 5.2.4, and assume that Assumptions 5.5.1-5.5.5 are satisfied. Then, for fixed* a, $\gamma \mapsto V_0(\gamma) = E^Q[I_a(X_T, \gamma q(X_T))]$ *is a nonincreasing differentiable function. Furthermore, the function* $\gamma \mapsto E[(g(X_T, Z_T) - I_0(X_T, \gamma q(X_T)))^+]$ *is differentiable in* γ*. The differentials are given by*

$$\begin{aligned}\frac{d}{d\gamma}E^Q[I_a(X_T, \gamma q(X_T))] = &\sum_{k=2}^{n} f(c_k)q(c_k)\left(g(c_k, Z_{k-1}) - g(c_k, Z_k)\right)\frac{dc_k}{d\gamma}\\ &- (g(c_1, Z_1) - a)f(c_1)q(c_1)\frac{dc_1}{d\gamma}\end{aligned} \tag{5.5.32}$$

and

$$\begin{aligned}&\frac{d}{d\gamma}E[(g(X_T, Z_T) - I_0(X_T, \gamma q(X_T)))^+] =\\ &\qquad\sum_{k=2}^{n} P_k\left(g(c_k, Z_k) - g(c_k, Z_{k-1})\right)f(c_k)\frac{dc_k}{d\gamma} + g(c_1, Z_1)f(c_1)\frac{dc_1}{d\gamma},\end{aligned} \tag{5.5.33}$$

where

$$c_k := q^{-1}\left(\frac{P_k}{\gamma}\right). \tag{5.5.34}$$

Proof. By Lemma 5.5.9, we know the formula of the integrand of the expectation operator. Furthermore, by Assumption 5.5.5, we know that $\gamma q(X_T) \in]P_{k+1}, P_k]$ if and only if $X_T \in [c_k, c_{k+1}[$ with c_k given by (5.5.34). By Assumption 5.5.3, it follows that the integrals

$$E^Q[g(X_T, Z_k)1_{X_T\in[c_k,c_{k+1}[}] = \int_{c_k}^{c_{k+1}} f(x)q(x)g(x, Z_k)dx$$

are differentiable functions, and by the fundamental theorem of differential calculus and the chain rule one has

$$\frac{d}{d\gamma}\int_{c_k}^{c_{k+1}} f(x)q(x)g(x,Z_k)dx = f(c_{k+1})q(c_{k+1})g(c_{k+1},Z_k)\frac{dc_{k+1}}{d\gamma} - f(c_k)q(c_k)g(c_k,Z_k)\frac{dc_k}{d\gamma}. \tag{5.5.35}$$

A summation and reordering of the terms yields the result (5.5.32), considering that q is differentiable and invertible and therefore $\frac{dc_k}{d\gamma} = \frac{1}{q'\left(q^{-1}\left(\frac{P_k}{\gamma}\right)\right)}\left(-\frac{P_k}{\gamma^2}\right)$. That the function is nonincreasing follows from the facts that $f(c_k) \geq 0$ for any $1 \leq k \leq n$, that $\frac{dc_k}{d\gamma} > 0$ by Assumption 5.5.5, and that $q(X_T)$ is equal to the density function and therefore, because of the equivalence of the two measures, $q(c_k) > 0$. By Assumption 5.5.2, all terms in the sum of (5.5.32) are nonpositive. The proof of equation (5.5.33) essentially follows the same arguments. □

5.5.2 Calculation of the CVaR price for general insurance state models

Theorem 5.5.11. *Let the insurance and financial processes Z and X as well as the insurance claim $g(X_T, Z_T)$ satisfy the model assumptions stated in sections 5.2.1 and 5.2.4, and assume that Assumptions 5.5.3, 5.5.4, and 5.5.6 hold. Moreover, let* $\sup_{Z_T} g(x, Z_T)$ *be finite for each x. Then the CVaR price V_0 and the parameter a are again given by (5.3.9)-(5.3.11). The parameter γ is determined by minimizing V_0 with respect to γ. If Assumptions (5.5.1)-(5.5.6) all are satisfied, formulas (5.3.12) and (5.3.13) are valid as well.*

Remark 5.5.12. As for the case of continuous distributions, we often have that $P[q(X_T) > \frac{1}{\gamma}] \approx 0$ and $Q[q(X_T) \leq \frac{1}{\gamma}] \approx 1$, so that formula (5.3.12) holds with $\epsilon \approx 0$. In this case, γ can be obtained explicitly, otherwise a zero-finding algorithm is needed.

Remark 5.5.13. Again, equations (5.3.12) and (5.3.13) neither depend on the insurance process Z_t nor on the payoff $g(X_T, Z_T)$.

Here we first prove another lemma:

Lemma 5.5.14. *The function $E[(g(X_T, Z_T) - a - I_a(X_T, \gamma q(X_T)))^+]$ can be written as*

$$\begin{aligned} E[(g(X_T, Z_T) - a - I_a(X_T, \gamma q(X_T)))^+] &= E[(g(X_T, Z_T) - I_0(X_T, \gamma q(X_T)))^+] \\ &\quad - aP[q(X_T) \geq \tfrac{1}{\gamma}]. \end{aligned} \tag{5.5.36}$$

In particular, the function is continuous in a for fixed γ.

Proof. We still have $I_a(x,y) = (I_0(x,y) - a)1_{y\leq 1}$ from the proof of Theorem 5.3.9. Taking the expectation and conditioning gives

$$E[(g(X_T,Z_T) - a - I_0(X_T,\gamma q(X_T)) + a)^+ 1_{\gamma q(X_T)\leq 1}] + E[(g(X_T,Z_T) - a)1_{\gamma q(X_T)>1}]$$

Recognizing that $I_0(x,y) = 0$ if $y > 1$, the result follows. □

Proof of Theorem 5.5.11. Let V_0 be the CVaR price. Then, by the arguments of section 5.2.3, there exist an a and a strategy π^* such that

$$f_{min}(a,V_0) = f(a,\pi^*,V_0) = E[g(X_T,Z_T) - a] - E[\alpha(X_T,Y_T^{\pi^*})] \leq -a(1-\beta),$$

where $\alpha(x,y)$ is nondecreasing and concave. Furthermore, the FHFC condition is satisfied. From Theorem 5.4.2 and Corollary 5.4.7, it follows that there exists a $\gamma > 0$ with

$$f_{min}(a,V_0) = E[g(X_T,Z_T) - a] - E[\alpha(X_T, I_a(X_T,\gamma q(X_T)))],$$

and $V_0(\gamma) = E^Q[I_a(X_T,\gamma q(X_T))]$.

We now show equality (5.3.11). The inequality

$$E[(g(X_T,Z_T) - a - I_a(X_T,\gamma q(X_T)))^+] \leq -a(1-\beta) \qquad (5.5.37)$$

follows from the acceptability condition. If this would be a strict inequality, then $a < 0$, because on the left hand side, the expression cannot be negative. By increasing a, it follows from Assumption 5.5.6 that

$$V_0(\gamma) = E^Q[I_0(X_T,\gamma q(X_T))] - aQ(X_T \leq \frac{1}{\gamma})$$

is strictly decreasing in a. On the other hand, by Lemma 5.5.14, the left hand side of (5.5.37) is continuous in a, and therefore the inequality would still be satisfied for a strictly larger a. But this means that V_0 cannot be the CVaR price. The inequality must therefore be an equality. Again by Lemma 5.5.14, it is clear that we can solve (5.5.37) for a. If we substitute (5.3.11) into (5.3.9), we obtain for each γ a price $V_0(\gamma)$ which satisfies the constraint (5.2.3) and is therefore acceptable. Therefore, the CVaR price is the one which minimizes $V_0(\gamma)$.

Let now all Assumptions 5.5.1-5.5.6 be satisfied. By Lemmas 5.5.10 and 5.5.14, it follows that a is differentiable in γ with

$$\frac{da}{d\gamma} = -\frac{\frac{d}{d\gamma}E[(g(X_T,Z_T) - I_0(X_T,\gamma q(X_T)))^+]}{1-\beta-P[q(X_T) > \frac{1}{\gamma}]} + a\frac{\frac{d}{d\gamma}P[q(X_T) > \frac{1}{\gamma}]}{1-\beta-P[q(X_T) > \frac{1}{\gamma}]}$$

and

$$V_0'(\gamma) = \frac{d}{d\gamma} E^Q[I_a(X_T, \gamma q(X_T))] - Q[q(X_T) \leq \frac{1}{\gamma}] \frac{da}{d\gamma}.$$

Substituting (5.5.32) and (5.5.33) into these equations and recognizing that $q(c_k) = \frac{P_k}{\gamma}$ leads to the expression

$$\frac{dV_0}{d\gamma} = (\text{nonpositive terms}) \left(\frac{1}{\gamma} - \frac{Q[q(X_T) \leq \frac{1}{\gamma}]}{1 - \beta - P[q(X_T) > \frac{1}{\gamma}]} \right),$$

and γ given by equations (5.3.12) and (5.3.13) must be a minimum for $V_0(\gamma)$. □

5.5.3 Application to the unit-linked insurance model

We again take the same model for the financial market as in section 5.3 and the same unit-linked survival insurance, but now with a discrete distribution of the amount of the survivors N_T which is not specified at this stage. It follows that Assumptions 5.5.1-5.5.6 are all satisfied. We can therefore apply Theorem 5.5.11 and follow the same algorithm as presented in section 5.3.3.

We again have that $q(x) = X_0^{\frac{\mu}{\sigma^2}} e^{\frac{1}{2}(-\mu + (\frac{\mu}{\sigma})^2)T} x^{-\frac{\mu}{\sigma^2}}$, and therefore, together with the assumption of a geometric Brownian motion, we have by (5.3.13) that

$$\gamma = \frac{1 - \beta - \Phi\left(\frac{\ln\left(\frac{c_0}{X_0}\right) - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \right)}{\Phi\left(\frac{\ln\left(\frac{X_0}{c_0}\right) - \frac{\sigma^2}{2}T}{\sigma\sqrt{T}} \right)}, \tag{5.5.38}$$

where Φ is the cumulative normal distribution function. The payoff to be hedged is

$$v(X_T) := \sum_{k=0}^{n-1} (kX_T - a) 1_{c_k < X_T < c_{k+1}} + (nX_T - a) 1_{X_T > c_n} \tag{5.5.39}$$

with

$$c_k := q^{-1}\left(\frac{P[N_T \geq k]}{\gamma} \right) = \left(\frac{\gamma}{\sum_{j=k}^{n} p_j} \right)^{\frac{\sigma^2}{\mu}} X_0 e^{\frac{T}{2}(\mu - \sigma^2)}. \tag{5.5.40}$$

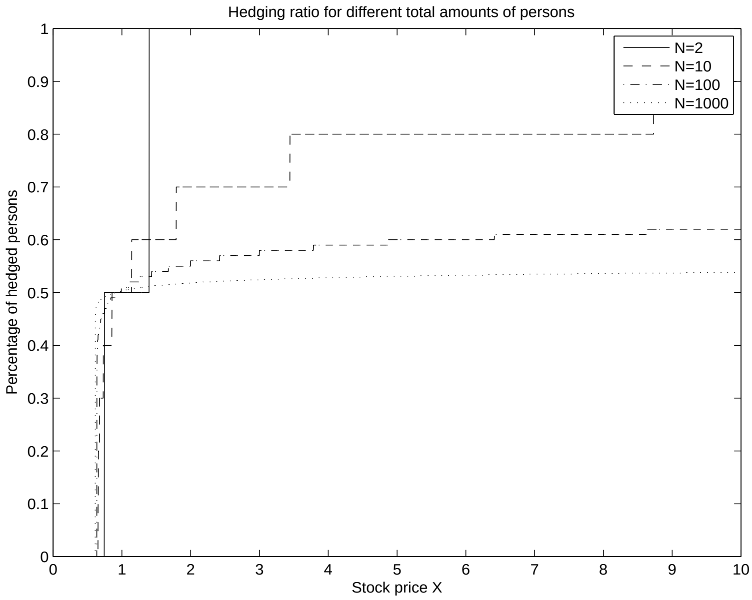
The parameter a is given by

$$a = \frac{a_0}{1 - \beta - \Phi\left(\frac{\ln\left(\frac{c_0}{X_0}\right) - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \right)}, \tag{5.5.41}$$

where a_0 is given by

$$a_0 = X_0 e^{\mu T} \sum_{k=1}^{n} \left(\sum_{l=k}^{n} p_l \right) \Phi\left(\frac{\ln\left(\frac{c_k}{X_0}\right) - (\mu + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \right). \tag{5.5.42}$$

Figure 5.1: Optimal payoff for different amounts of insured persons, $a = 0$



Finally, we obtain the CVaR price by

$$V_0(a) = \sum_{k=1}^{n} X_0 \Phi\left(\frac{\ln\left(\frac{X_0}{c_k}\right) + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right) - a\Phi\left(\frac{\ln\left(\frac{X_0}{c_0}\right) - \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right). \tag{5.5.43}$$

In this context, we use the notation $c_{n+1} := \infty$.

Remark 5.5.15. Even if under the original probability measure the processes N_t and X_t are independent, the optimal strategy is not simply the delta hedge corresponding to the payoff npX_T; instead, a higher survival rate is hedged for larger values of X_T. This also means that under the worst-case martingale measure as defined in Föllmer and Leukert [32] the two events are no longer independent.

Remark 5.5.16. As in the case of continuous distributions, with the approximation $\epsilon \approx 0$, by substituting (5.5.38) into (5.5.40) and then into (5.5.41) and (5.5.43), we obtain an analytical formula for the CVaR price.

As an illustration, Figure 5.1 shows the hedge ratio of the optimal payoff to be hedged, i.e. $\frac{v(x)}{nx}$, as a function of the terminal stock price for different amounts of insured persons. $v(x)$ is the function given in (5.5.39), where for simplicity we take

Table 5.2: Minimal capital for different parameters, $\beta = 0.95$

n	p	γ	a	V_0	*Load*
1000	0.5	0.05	-6.62	532.60	6.5%
1000	0.1	0.05	-4.19	120.0	20.0%
50	0.5	0.05	-1.45	32.24	29.0%
50	0.1	0.05	-1.07	9.76	95.2%

$a = 0$. From this, one can see that the optimal payoff is a sum of knock-in options. We take for the insurance probabilities a binomial distribution with n persons and $p = 0.5$.

For only a few persons (in practice, one can think about a special insurance for a few persons with very large payoffs), zero payoff is hedged if the stock price is below a limit c_1, whereas a full hedge is implemented if the stock price is large enough. The reason for this is that in the case of high stock prices, the risk that there are more survivors than expected plays a much larger role than in the case of low stock prices. If the amount of insured persons increases, the optimal hedge ratio converges more and more to the one which is usually used by actuaries in practice, namely the hedge of the expected amount of survivors.

Numerical results

Even if CVaR is a translation invariant risk measure, it is not the same as to calculate only the minimal CVaR at capital zero and take this as the minimal capital required. The reason is that for larger initial capitals, more trading strategies are allowed.

We repeat the numerical example from section 5.3, with the same parameters. With these parameters we obtain

$$\Phi\left(\frac{\ln\left(\frac{c_0}{X_0}\right) - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}\right) = 1.2 \times 10^{-18} \approx 0$$

and

$$\Phi\left(\frac{\ln\left(\frac{X_0}{c_0}\right) - \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right) = 1 - 2.6 \times 10^{-17} \approx 1,$$

which justifies the approximation $\epsilon \approx 0$. The results for the discrete probability case are given in Table 5.2. Comparing these results with the ones of section 5.3, one can see that the differences to the normal approximation are rather small, even for the case of only 50 persons.

For looking at the sensitivity of the risk loading with respect to the CVaR level β, we have also performed the calculations also for $\beta = 0.99$. This gives a risk loading for 50 persons of 37% when $p = 0.5$, and 127% for $p = 0.1$.

5.6 Conclusion

In this chapter, we have shown how the CVaR price can be obtained in order to make an insurance payoff acceptable for the seller, under the assumption that nonfinancial information is only obtained at the time of maturity. Using this assumption, the problem simplifies to a complete market problem, and the CVaR pricing problem including the corresponding hedging strategy become almost explicit formulas.

The approach we have chosen for this problem is to apply the relationship between expected shortfall and CVaR given in Rockafellar and Uryasev [73], and to connect the CVaR pricing problem to the one of minimization of expected shortfall treated in Föllmer and Leukert [33]. For the case where the insurance claim has a continuous distribution function, the state dependent utility function satisfies the requirements given there, and one can directly apply the results of Föllmer and Leukert. We have shown that for this situation, and under some additional technical assumptions which are satisfied in the Black-Scholes model, we approximately have an analytical formula for the CVaR price. Without approximation, we at least have a straightforward algorithm which does not require any optimization procedure.

However, in life insurance, the probabilities are typically discrete, and even if they are continuous, the insurance payoff may have some points with nonzero probability. As a consequence, one cannot directly apply the results from before. Therefore, a considerable part of the chapter was devoted to extending the result of Föllmer and Leukert to the case where the state dependent utility function is not necessarily differentiable nor strictly concave, in order to be able to apply the result for discrete probabilities as well. The result we obtained holds quite generally and applies to any state dependent utility maximization problem in complete markets.

With this additional result, we could formulate, under some additional assumptions, an algorithm similar to the one for the case of continuous distribution functions. However, to find necessary conditions in order to have the algorithm which does not require any numerical optimization is a technically delicate problem, which we left open for further research. Instead, we proved a result under some more restrictive assumptions which are satisfied by the Black-Scholes model. Again using an approximation, we obtained an almost analytical formula for the CVaR price.

We applied the theoretical results to the case of a unit-linked survival insurance, where we applied a truncated normal approximation for being able to apply the results of section 5.3, as well as a binomial model where we applied the results of section 5.5. We obtained analytical formulas for the hedge, as well as for the CVaR price, again using the approximation which is very accurate for reasonable parameters. Otherwise, if this approximation is not good enough, one has to solve numerically an equation in one variable, which is also a feasible problem.

In practice, the assumption that insurance information arrives only at the end of the period may be unrealistic. However, the price obtained by this assumption can be thought of as an upper bound for the actual CVaR price, because for the hedge, it is always possible to ignore additional information. The question then arises how good this upper bound is. The extension from insurance information at the end to a general information structure will be covered in the next chapter. Similarly to the considerations of Föllmer and Leukert, the idea is to introduce more and more information steps, and to prove the convergence when information arrives continuously. However, our first numerical examples show that there is only little gain when considering more than only information at the end. On the other hand, there is no analytical solution any more when one considers more information steps, and the numerics becomes more difficult.

Here we have only considered insurance payoffs which occur at a specific terminal time. However, in practice, one often has payment processes, such as pension rents. In such a situation, one has to reconsider what is meant by the requirement that the wealth process is bounded from below, because the insurance payoffs may reduce the wealth at a time before the terminal one. This also has consequences concerning the assumption that information about the insurance process only arrives at the end, because if state dependent payments have to been done before the terminal time, one obviously knows more about the insurance process. The study of such payment processes, as well as a reasonable redefinition of the corresponding CVaR price, would be another issue of further research.

A natural next step in further research would also be to consider the problem of satisfying the CVaR criterion and maximizing the profits in the sense of Basak and Shapiro [6]. This is a typical problem a financial institution has to deal with.

The application has essentially been done with a specific model for the financial and the insurance process. The general considerations, however, could also be applied to other situations. In particular, it would be interesting to apply them to a situation where there exists no analytical solution of the model and where the solution has to be found numerically, such as in financial models with nonconstant coefficients or with stochastic volatility.

Finally, one could think about more general risk measures than the CVaR. In Föllmer and Schied [34], it has been shown that the CVaR is a building block for all law-invariant risk measures. Even if it is not obvious how to combine the theories in order to calculate a risk measure price for a general law invariant risk measure, it would be interesting to see whether an extension in this direction is possible.

Chapter 6

Continuous information limit

6.1 Introduction

In Chapter 5, we have seen how the CVaR price can be calculated under the assumption that nonfinancial information is only available at the terminal time. The question however is if one can generalize the method to the case when the financial as well as the insurance process are continuous-time processes.

In principle, this is indeed possible, and the idea is, similarly to Föllmer and Leukert [32], to discretize the time at which information is available, and to apply for each time step the results of Chapter 5. This is the issue of this chapter, as well as to prove that the limit of the CVaR prices obtained using this discretization is indeed the CVaR price for the general (continuous-time) case.

However, there are several difficulties in this procedure. By the tower property of conditional expectations, one can condition on everything which is known at a certain time when new insurance information comes up, similarly to Föllmer and Leukert [32], and essentially follow the same method as in section 5.2.4. However, this only applies to the expected shortfall and not to the CVaR. Furthermore, unless one is in the terminal time step, the derivative of the function α from (5.2.7) is no longer of the simplified form as stated in Lemma 5.3.15, with the consequence that our main results, Theorems 5.3.9 and 5.5.11, can no longer be applied. It cannot even be expected that α still has an analytic form even for the standard Black-Scholes model. Furthermore, it is not obvious how to find the parameter γ of these theorems, and it is also not clear that this parameter is constant for all time steps. As a consequence one has to determine it numerically at each time step.

This chapter presents a procedure which can in principle be applied for obtaining the CVaR price. We give some numerical examples for the model that has already been treated in Chapter 5. The problem is numerically solved by a discrete Legendre transformation method and an iterative solution for the parameter γ at each time step. This leads to a feasible solution as long as the information only arrives at a few time points, but it requires a large computational effort for finer information time grids.

6.2 Information about amount of survivors during hedging process

Let the model again be the one of section 5.2, but without the assumption of section 5.2.4 that information is only available at the terminal time T. As a first step, we assume that there is a time t_0 with $0 < t_0 < T$ where the information about the insurance process Z_{t_0} arrives. This information can therefore be used for the further hedge.

Let us already be at the time t_0 such that we know the variable X_{t_0}, the capital V_{t_0}, and the insurance outcome Z_{t_0}. Then our problem is precisely the one of section 5.2.4, and our aim is to optimize the expectation

$$E[\alpha(X_T, Y_T^\pi)|\mathcal{F}_{t_0}]$$

under the restriction that the initial value is $Y_{t_0}^\pi = V_{t_0}$. $\alpha(x, y)$ is the measurable function defined in (5.2.7), which is concave in y, with exception that the expectation now may depend on the outcome Z_{t_0}, because the distribution function $dP(Z_T)$ in the integral (5.2.7) now is conditional on Z_{t_0}. We assume again, as in Assumptions 5.3.5 and 5.5.5, that the unique Radon-Nikodym density for the financial process $\frac{dQ}{dP}$ is a function of the terminal state variable X_T. By the Markov property of X_t, the whole density process then is a function of X_t. We may therefore apply Corollary 5.4.7 to obtain a measurable function $v_{V_{t_0}}$ such that $v_{V_{t_0}}(X_T)$ is the optimal claim to hedge given the information at time t_0. The optimal value at time t_0 is therefore

$$u(X_{t_0}, V_{t_0}, z) = E[\alpha(X_T, v_{V_{t_0}}(X_T))|Z_{t_0} = z], \tag{6.2.1}$$

where z denotes the insurance outcome at time t_0.

Proposition 6.2.1. *Let the function $\alpha(x, y)$ be nondecreasing and concave in the second argument. Then for all z, the function u defined by equation (6.2.1) is nondecreasing and concave in the second argument.*

Proof. For simplifying notation, we set $X_0 := X_{t_0}$ and $V_0 := V_{t_0}$. Let V_0 and $\tilde{V}_0$ be two initial capitals. It is trivial that the function is nondecreasing, because if

$\tilde{V}_0 > V_0$, we have

$$\begin{aligned} u(X_0, V_0, z) &= \sup_\pi E[\alpha(X_T, Y_T^\pi)|X_0, Y_0 \leq V_0, Z_{t_0} = z] \\ &\leq \sup_\pi E[\alpha(X_T, Y_T^\pi)|X_0, Y_0 \leq \tilde{V}_0, Z_{t_0} = z] = u(X_0, \tilde{V}_0, z). \end{aligned}$$

By the concavity of α,

$$\begin{aligned} &tu(X_0, V_0, z) + (1-t)u(X_0, \tilde{V}_0, z) \\ &\qquad = E[t\alpha(X_T, v_{V_0}(X_T)) + (1-t)\alpha(X_T, v_{\tilde{V}_0}(X_T))|\mathcal{F}_{t_0}] \\ &\qquad\qquad \leq E[\alpha(X_T, tv_{V_0}(X_T) + (1-t)v_{\tilde{V}_0}(X_T))|\mathcal{F}_{t_0}]. \end{aligned} \tag{6.2.2}$$

But by definition,

$$B_{t_0} E^Q[v_{V_0}(X_T)|\mathcal{F}_{t_0}] \leq V_0,$$

where B_t is the price process of the zero bond with expiry at time T. The same holds for $\tilde{V}_0$, so that under Q the claim $tv_{V_0}(X_T) + (1-t)v_{\tilde{V}_0}(X_T)$ has the discounted expectation of less than or equal to $tV_0 + (1-t)\tilde{V}_0$. This claim is therefore (super-) replicable at initial capital $tV_0 + (1-t)\tilde{V}_0$, and the optimal value available at this capital must satisfy

$$E[\alpha(X_T, tv_{V_0}(X_T) + (1-t)v_{\tilde{V}_0}(X_T))|\mathcal{F}_{t_0}] \leq E[\alpha(X_T, v_{tV_0+(1-t)\tilde{V}_0}(X_T))|\mathcal{F}_{t_0}],$$

where the function $v_{tV_0+(1-t)\tilde{V}_0}(X_T)$ denotes the optimal claim available at capital $tV_0 + (1-t)\tilde{V}_0$. Together with equation (6.2.2), this implies the concavity of the function u. □

At time t_0^-, we do not know yet the outcome Z_{t_0} at time t_0, and our value function is therefore

$$\alpha_0(X_{t_0}, V_{t_0}) = \int u(X_{t_0}, V_{t_0}, z) dP_{Z_0}(z), \tag{6.2.3}$$

where dP_{Z_0} denotes the distribution function of Z_{t_0}. This is a convex combination of concave functions and therefore still concave. As a consequence, we can repeat our procedure for arbitrarily many points $t_n < t_{n-1} < ..t_1 < t_0 < T$.

As all traded assets, B_t is assumed to be a function of X_t, so that a stochastic interest rate can easily be included.

6.3 Limiting case: continuous-time information

The aim of this section is to prove the following theorem:

Theorem 6.3.1. *Let the model assumptions of section 5.2.1 be satisfied, but let now Z_t, the insurance process, be a continuous-time process. Let the claim $g(X_T, Z_T)$ of section 5.2.1 be integrable. Let Π_k be a sequence of partitions of $[0,T]$, and let $u_k(x,y)$ be the optimal solution using the method above, with information about mortality at the points $0 \leq t_1^{(k)} < .. < t_k^{(k)} \leq T$, i.e. at the points of the partition, and let $u(x,y)$ be the optimal value function in continuous-time. Then, if the mesh of Π_k converges to 0 as $k \to \infty$ and if for all k, Π_{k+1} is a refinement of Π_k, the value functions $u_k(x,y)$ converge monotonically increasing to $u(x,y)$.*

For the proof of Theorem 6.3.1, we first need several lemmas.

Lemma 6.3.2. *If Π_{k+1} is a refinement of Π_k, then $u_{k+1}(x,y) \geq u_k(x,y)$.*

Proof. The strategy leading to the value $u_k(x,y)$ is also allowed if we have information at the time points Π_{k+1} (by only considering the information at time points Π_k). Because $u_{k+1}(x,y)$ is the optimal value for the information times Π_{k+1}, it must be larger than or equal to $u_k(x,y)$. □

Proposition 6.3.3. *If $g(X_T, Z_T)$ is integrable, the sequence $u_k(x,y)$ converges pointwise to a limit.*

Proof. By Lemma 6.3.2, for each fixed point (x,y), $u_k(x,y)$ is a monotonically increasing sequence of real numbers. Furthermore

$$u_k(x,y) \leq E[(g(X_T, Z_T) - a)^+ | X_0 = x, V_0 = y],$$

and therefore, for all (x,y), the sequence is bounded from above. It follows that the sequence converges pointwise. □

Proposition 6.3.4. *Let the claim $g(X_T, Z_T)$ to be hedged be integrable, and let π^* be the optimal strategy at initial capital y and initial state $X_0 = x$. Assume that this strategy is càglàd, and that the traded assets S in the financial market are continuous. Then, for every sequence of partitions Π_k with mesh converging to zero, the value functions $\tilde{u}_k(x,y)$ of the constant policy strategy*

$$\tilde{\pi}_k(t) := \begin{matrix} \pi_k(t) & \text{if} & t < \hat{T}^{(k)}, \\ 0 & \text{if} & t \geq \hat{T}^{(k)}, \end{matrix} \tag{6.3.4}$$

with

$$\pi_k := \sum_{j=1}^{k} \pi^*(T_{j-1}^{(k)}) 1_{(T_{j-1}^{(k)}, T_j^{(k)}]} \tag{6.3.5}$$

converge pointwise to the optimal value function $u(x, y)$. Here, for any $k \geq 1$, $(T_j^{(k)})_{j\geq 1}$ is the sequence of stopping times inductively defined by $T_0^{(k)} = 0$ and

$$T_j^{(k)} := \begin{array}{ll} t_j & \text{if} \quad \tilde{Q}_{j-1,k}(t) > 0 \quad \forall \quad t \leq t_j \\ t_j & \text{if} \quad \exists t \leq T_{j-1}^{(k)} : Q_{j-1,k}(t) \leq 0 \\ \inf\{t > 0 : \tilde{Q}_{j-1,k}(t) \leq 0\} & \text{otherwise,} \end{array} \tag{6.3.6}$$

where Q_{jk} and $\tilde{Q}_{jk}$ are defined again inductively by

$$\begin{aligned} Q_{jk}(t) &:= V_0 + \textstyle\sum_{l=1}^{j} \pi^*(T_{l-1}^{(k)}) \left(S_{t\wedge T_l^{(k)}} - S_{t\wedge T_{l-1}^{(k)}}\right), \\ \tilde{Q}_{jk}(t) &:= Q_{jk}(t) + \pi^*(T_j^{(k)}) \left(S_t - S_{t\wedge T_j^{(k)}}\right). \end{aligned}$$

Furthermore $\hat{T}^{(k)}$ is the stopping time

$$\hat{T}^{(k)} := \inf\left\{t > 0 : V_0 + \int_0^t \pi_k dS \leq 0\right\}. \tag{6.3.7}$$

Proof. Let V_T^* be the optimal capital of strategy π^*, i.e.

$$V_T^* = V_0 + \int_0^T \pi_t^* dS_t.$$

Obviously,

$$\int_0^T \tilde{\pi}_k dS = \int_0^T \pi_k dS + \int_0^T (\tilde{\pi}_k - \pi_k) dS.$$

By standard theory, see e.g. Protter [69], the random variables V_k, defined by

$$V_k := V_0 + \int_0^T \pi_k dS = V_0 + \sum_{j=1}^{k} \pi_{T_{j-1}}^* (S_{T_j} - S_{T_{j-1}}),$$

converge in probability to the random variable V_T^*. For showing that the second term converges to zero in probability, by the continuity of the integral operator (see [69]), it suffices to show that the integrand converges in *ucp* to zero, because it is a sequence of simple predictable processes.

Let $A := \{V_T^* > 0\}$. Then, for each fixed $\omega \in A$, $V_t^*(\omega) > 0$ for all $t \leq T$, since $V_t^* = 0$ on an $\mathcal{F}_t$-measurable subset of A for some $t < T$ and $V_T^* > 0$ on A would be an arbitrage opportunity. By the continuity of S_t, the continuity of V_t^* follows, and hence also the uniform continuity of V_t^* on $[0, T]$. It follows that A can be written as

$$A = \cup_{n\geq 1} A_n := \cup_{n\geq 1} \left\{\omega : \inf_t V_t^*(\omega) \geq \frac{1}{n}\right\}.$$

We have

$$P[\{|\tilde{\pi}_k - \pi_k| > \epsilon\} \cap A] \leq P[\cup_{n\geq N} A_n] + P[\{|\tilde{\pi}_k - \pi_k| > \epsilon\} \cap A_N]$$

and

$$P[\{|\tilde{\pi}_k - \pi_k| > \epsilon\} \cap A_N] \leq P[\{V_0 + \inf_t \int_0^t \pi_k(r) dS_r \leq 0\} \cap A_N].$$

But for fixed N, the right hand side converges to zero, because of the *ucp* convergence of the integral to V_t^* and the fact that $V_t^* \geq \frac{1}{N}$ uniformly on A_N. Therefore, for each $\delta > 0$, we can choose N so large that $P[A \setminus A_N] < \frac{\delta}{2}$, and then choose k so large that $P[\{|\tilde{\pi}_k - \pi_k| > \epsilon\} \cap A_N]$ becomes small. Therefore, we have *ucp* convergence of $\tilde{\pi}_k - \pi_k$ to 0 on A, and the integral has the same property.

On A^c, i.e. on $\{V_T^* = 0\}$, either $\tilde{\pi}_k = \pi_k$ or $V_0 + \int_0^T \tilde{\pi}_k dS = 0$, by definition of $\tilde{\pi}_k$. In both cases, the stochastic integral converges to 0. It follows that $\tilde{\pi}_k$ is an approximating sequence of strategies as well. By definition, these strategies give a nonnegative wealth process.

It follows that the sequence $(g(X_T, Z_T) - a)^+ \wedge \tilde{V}_k)$ converges in probability to $(g(X_T, Z_T) - a)^+ \wedge V_T^*)$, where $\tilde{V}_k := V_0 + \int_0^T \tilde{\pi}_k dS$. But since $(g(X_T, Z_T) - a)^+$ is integrable, the sequence is dominated by an integrable random variable. It follows that the sequence is uniformly integrable. It therefore converges in L^1 to $(g(X_T, Z_T) - a)^+ \wedge V_T^*)$, and hence

$$\tilde{u}_k(x, y) \to u(x, y).$$

This concludes the proof. □

Proposition 6.3.5. *If there does not exist an optimal strategy which is càglàd, there exists a subsequence of partitions Π_{k_j} such that Proposition 6.3.4 still holds for this subsequence.*

Proof. There exists a sequence of càglàd strategies $(\pi_l)_{l \geq 1}$ such that

$$\hat{u}_l(x, y) := E[(g(X_T, Z_T) - a)^+ \wedge (V_0 + \int_0^T \pi_l dS_t) | X_0 = x, V_0 = y] \to u(x, y)$$

pointwise in (x, y). For every $\hat{u}_l$, we can apply the same arguments as in Proposition 6.3.4, and we obtain, for the partitions Π_k, a constant policy approximation $\tilde{u}_{kl}(x, y) \to \hat{u}_l(x, y)$. Let now $\epsilon > 0$. Then, for any fixed (x, y) and any $j \geq 1$ we find an l_j with $|\hat{u}_{l_j}(x, y) - u(x, y)| < 2^{-j}$. Similarly, for any j, we find a k_j with $|\tilde{u}_{k_j l_j}(x, y) - \hat{u}_{l_j}(x, y)| < 2^{-j}$. The sequence $w_j(x, y) := \tilde{u}_{k_j l_j}(x, y)$ then converges pointwise to $u(x, y)$, and is of the form (6.3.4) for the partitions Π_{k_j}. □

We now apply Proposition 6.3.5 to the case where the optimal strategy is predictable but not necessarily left-continuous. The following lemma seems to be quite standard, but we did not find a suitable reference. We therefore give a proof.

Lemma 6.3.6. *Let S be a semimartingale and π a predictable S-integrable process. Then there exists a sequence π_n of càglàd processes such that*

$$\int_0^T \pi_n dS \to \int_0^T \pi dS \tag{6.3.8}$$

almost surely.

Proof. If S is square integrable and special and if π is bounded, it follows by Theorems 2 and 3 of Chapter 4 in Protter [69] that there exists a sequence of càglàd strategies π_n such that the integral $\pi_n \cdot S$ converges to $\pi \cdot S$ in the space of square-integrable semimartingales. If π is not bounded, we can approximate it by the sequence π_m of bounded processes defined by $\pi_m := \pi 1_{|\pi| \leq m}$. Then the integral $\pi_m \cdot S$ converges to $\pi \cdot S$ in the space of square-integrable semimartingales, by Theorem 14 of the same chapter. It follows that there exists a subsequence of bounded càglàd processes such that the integral converges in this space. For a fixed $T > 0$, we have the convergence

$$\int_0^T \pi_{n_k} dS \to \int_0^T \pi dS \tag{6.3.9}$$

in probability, with π_{n_k} given by the subsequence from above. Hence, we have a subsequence of π_{n_k}, again denoted by π_{n_k}, such that the convergence is almost surely.

By definition in Protter [69], a predictable process π is $S-$ integrable if there exists a sequence $(T_n)_{n\geq 1}$ of stopping times such that the processes $S^{T_n^-}$ are square-integrable special semimartingales and π is integrable with respect to $S^{T_n^-}$. Since the stochastic integral is defined in such a way that

$$\int_0^T \pi dS^{T_n^-} \to \int_0^T \pi dS$$

almost surely, the statement of the lemma follows from the above considerations. □

Corollary 6.3.7. *If the optimal strategy π^* from Proposition 6.3.4 is predictable and S-integrable but not necessarily càglàd, Proposition 6.3.4 still holds, provided the claim $g(X_T, Z_T)$ to be hedged is integrable.*

Proof. By Lemma 6.3.6, there exist càglàd strategies π_n satisfying (6.3.8). For the sequence $((g(X_T, Z_T) - a)^+ \wedge Y_T^{\pi_n})_{n\geq 1}$, we have the convergence

$$(g(X_T, Z_T) - a)^+ \wedge Y_T^{\pi_n} \to (g(X_T, Z_T) - a)^+ \wedge Y_T^{\pi_*}$$

almost surely, and the sequence is dominated by $(g(X_T, Z_T) - a)^+$. It follows that the value of the supremum over all càglàd strategies is the same as the value of the optimal predictable strategy, and the result follows from Proposition 6.3.5. □

Now we turn to the proof of Theorem 6.3.1.

Proof of Theorem 6.3.1. By Lemma 6.3.2 and Proposition 6.3.3, we have already seen that the sequence $u_k(x,y)$ converges monotonically to a limit $\bar{u}(x,y)$. By the optimality of $u(x,y)$, it follows that $\bar{u}(x,y) \leq u(x,y)$. Let u_k now be fixed and Π_k its partition. Using this partition, we can define a constant policy $\tilde{\pi}_k$ as in equation (6.3.4). This strategy only uses the information on Z_t at times $t_j^{(k)}$ and is therefore also allowed in the model which only uses information at times in Π_k. By the optimality of $u_k(x,y)$ for this information, it follows that

$$u_k(x,y) \geq \tilde{u}_k(x,y).$$

By Proposition 6.3.5, there exists a subsequence Π_{k_j} of partitions such that $\tilde{u}_{k_j}(x,y) \to u(x,y)$, and therefore

$$u(x,y) \geq \bar{u}(x,y) \geq u_{k_j}(x,y) \geq \tilde{u}_{k_j}(x,y) \to u(x,y).$$

By the monotonicity in k of the sequence $u_k(x,y)$, the result follows. □

6.4 Limit for the CVaR

The aim of this section is to apply the continuous time limit results to the CVaR criterion. To this end, we reconsider the results of section 5.2.3 in order to see that the functions $f_{min}^{(n)}$ of section 5.2.3 are in the present context given by

$$f_{min}^{(n)}(a,V_0) = E\left[(g(X_T,Z_T)-a)^+\right] - u_n(X_0,V_0;a),$$

and accordingly for f_{min} with u instead of u_n. Let V_{max} be the minimal initial capital such that a superhedge of $(g(X_T,Z_T)-a)^+$ in the continuous information model is possible.

Proposition 6.4.1. *The following holds:*

1. *The functions $f_{min}^{(n)}$ and f_{min} are strictly decreasing and convex in V_0, and $E\left[(g(X_T,Z_T)-a)^+\right]$ at $V_0 = 0$. Furthermore, at V_{max}, $f_{min} = 0$.*
2. *It follows that the functions $f_{min}^{(n)}$ and f_{min} are continuous and invertible in the interval $]0, V_{max}[$. Furthermore, $(f_{min}^{(n)})^{-1} \to f_{min}^{-1}$ $(n \to \infty)$ pointwise.*
3. *We have $V_0^{(n)}(a) = (f_{min}^{(n)})^{-1}(a, -a(1-\beta))$ and $V_0(a) = f_{min}^{-1}(a, -a(1-\beta))$, where $V_0^{(n)}(a)$ is the minimal capital such that there exists a strategy satisfying (5.2.3) for information at time points of partition Π_n and $V_0(a)$ the one for the continuous-time information.*

4. *If a^* minimizes $V_0(a)$, then $V_0(a^*)$ is the price of the claim for continuous-time information due to the CVaR criterion.*

5. *If a_n^* minimizes $V_0^{(n)}(a)$, then $V_0(a_n^*)$ converges monotonically decreasing to $V_0(a^*)$ $(n \to \infty)$.*

Proof. The crucial issue is statement 1. The other ones are standard or follow from our definitions.

It is clear that $f_{min}^{(n)}$ and f_{min} are nonincreasing and convex, because u_n and u are nondecreasing and concave. Furthermore, $u_n(x, 0; a) = 0$ and $u(x, 0; a) = 0$, because with zero wealth one cannot go into any hedging position different from zero without the possibility that the wealth becomes negative. It follows that $f_{min}^{(n)}(a, 0) = E\left[(g(X_T, Z_T) - a)^+\right]$ and $f_{min}(a, 0) = E\left[(g(X_T, Z_T) - a)^+\right]$. At V_{max}, we have a superhedge, and therefore $f_{min}(a, V_{max}) = 0$.

The fact that the functions are strictly decreasing follows from the fact that for all initial capitals V with $0 < V < V_{max}$, the functions must be strictly between 0 and $E\left[(g(X_T, Z_T) - a)^+\right]$. Furthermore, $f_{min}^{(n)}(a, V_{max}) \geq f_{min}(a, V_{max}) = 0$. If $f_{min}^{(n)}(a, V_{max}) > 0$, then $f_{min}^{(n)}(a, V_{max} + \Delta V) < f_{min}^{(n)}(a, V_{max})$ for $\Delta V > 0$, since one can follow the strategy which gives $f_{min}^{(n)}(a, V_{max})$ with the capital V_{max} and put ΔV in a numéraire. This gives a terminal wealth which is almost surely strictly larger than the one obtained with the capital V_{max}, and therefore $f_{min}^{(n)}(a, V_{max} + \Delta V) < f_{min}^{(n)}(a, V_{max})$. The functions $f_{min}^{(n)}$ are therefore strictly decreasing at V_{max}. Together with the fact that the functions are nonincreasing and convex, it follows that $f_{min}(a, V)$ and $f_{min}^{(n)}(a, V)$ are strictly decreasing on $]0, V_{max}[$.

The second statement follows from the first one and the standard result that if we have $f_n \to f$ for decreasing and invertible functions f_n and f, then $f_n^{-1} \to f^{-1}$. The other statements follow from what is stated before. □

6.5 Application to the unit-linked survival insurance

Due to the fact that we cannot expect the value function (6.2.1) at the intermediate time t_0 to have an analytical form, we switch to a numerical method in order to solve (5.2.8). Equation (5.4.26) gives a possibility to find the optimal payoff function. Indeed, since we know that the Radon-Nikodym density is proportional to $X_T^{-\frac{\mu}{\sigma^2}}$, it follows that

$$v(x) = \frac{\partial}{\partial y} \alpha^*(x, \gamma x^{-\frac{\mu}{\sigma^2}}), \tag{6.5.10}$$

Table 6.1: Minimal capital with information at the end, discrete Legendre method

n	p	a	V_0	*Load*
1000	0.5	-6.30	532.80	6.5%
1000	0.1	-4.25	120.00	20.0%
50	0.5	-1.48	32.25	29.0%
50	0.1	-1.09	9.78	95.6%

Table 6.2: Minimal capital with one information step between

n	p	a	V_0	*Load*
1000	0.5	-6.30	532.80	6.5%
1000	0.1	-4.25	120.00	20.0%
50	0.5	-1.48	32.25	29.0%
50	0.1	-1.09	9.78	95.6%

where α^* is the Legendre transform of α of Theorem 5.4.2 and γ is a constant which again has to be determined iteratively by the budget constraint. Our strategy now is to solve (6.5.10) by a discrete Legendre transform.

We will first test our method by again calculating the situation when information is available only at the end of the period, and then add one further information step between zero and the terminal time. It follows from Table 6.1 that the results obtained numerically are very similar to the ones obtained analytically in Table 5.2. This encourages us to apply this numerical method for the case with information between zero and the terminal time. It follows that adding a further information step makes almost no difference to the optimal price when information arrives only at the end of the time period, at least in our example. This is not obvious, and we would like to know the reason for this. First, the reduction of the risk loading is only about 10% with respect to the capital that would be needed in order to have a nonpositive CVaR when performing a naive hedging strategy. This naive strategy consists of investing precisely an amount of np in the stock and the remaining part of the capital in the risk-free asset.

Second, numerical calculations have shown the following. We first apply the hedging strategy which is optimal in the case where information arrives only at the end of the period. In the middle of the period, we look for the optimal strategy given the financial and insurance outcome at this time. It follows that the difference of the optimal payoffs for different amounts of survivors is only large for events which have a small probability.

We have done this calculation in particular in the case of only two insured persons and a survival probability of 10%. However, this effect seems to hold for all binomial distributions. It is rather likely that this changes when considering distributions which are not infinitely divisible, and in particular insurance processes whose conditional distributions for the future depend on their current state.

A more detailed treatment of the question when and why precisely further information gives almost no improvement is a topic of future research. For the purpose of this chapter, we state that the method of Chapter 5 appears to produce a reasonable approximation of the CVaR based price and can be obtained quite easily.

6.6 Conclusion

In this chapter, an algorithm has been given for obtaining the CVaR price in the case where financial as well as nonfinancial information arrives continuously. The method for doing this is to discretize the time at which information arrives, and to apply the results of Chapter 5. A limit theorem has been proved which shows the convergence of this method. A numerical example has been given, where the CVaR price has been calculated by a discrete Legendre transformation, an iterative procedure for finding γ as well as a numerical optimization procedure for determining a.

One of the main drawbacks of this method is that, in contrast to the one presented in Chapter 5, one still has an optimization problem to solve, even if this problem is only one-dimensional. Furthermore, it turns out that the solution method needs a lot of computational effort if one wants to calculate the CVaR prices using many information time steps. Therefore, finding more efficient numerical algorithms for this problem remains a topic of future research.

Another topic of future research is to determine the order of convergence. In section 6.3, we have only proved that the method converges, but not how fast it converges. For numerical purposes, however, one would like to have a convergence order. A first estimate of such an order may be done in a similar way as in Krylov [59], where a convergence order has been proved for constant policy approximations in a similar framework. Because the problem treated in [59] is different to the one here, it is not obvious how to apply the results of [59] to our situation, and we leave this for future research.

As already mentioned in our example it turned out that there is only little gain when considering more insurance information steps than only the one at the terminal time. Obtaining a bound for this difference is therefore another topic of future research.

Chapter 7

Conclusions and further research

This thesis has treated some theoretical as well as practical aspects of the pricing and hedging problem in incomplete markets. From a theoretical point of view, in Chapter 2 we have proposed to define the risk aversion of a utility function as a measure. This definition extends the classical one and has the advantage that it is also applicable for nonsmooth utility functions. Essentially, Chapter 2 completes the theory of Pratt [68], because it translates the main results on risk aversion of [68] to the new, generalized framework which is also suitable for nonsmooth utility functions. In particular, we have shown that the utility function can be reconstructed from the risk aversion, that an investor being more risk averse is equivalent to the utility function being more concave, and both are equivalent to the certainty equivalent of the investor being larger.

In Chapter 3, we have answered the question whether or not the indifference pricing martingale measure, or specifically the minimax martingale measure according to Bellini and Frittelli [7], is equivalent to the objective one in a measure theoretical sense. Complementary to the existing literature, we have given a proof for satiated utility functions, i.e. functions which have a maximal value. The question has been connected to relative risk aversion, and using Chapter 2, the results can also be applied for nonsmooth utility functions. Equivalence has been proved in a similar way as Delbaen and Schachermayer [25] did for the variance optimal measure, but with considerable extensions. In particular, we have given a conditional version of the Hölder inequality for Luxemburg norms.

From the practical point of view, there are many problems in incomplete markets which differ from each other rather strongly and therefore also need a different treatment. It cannot be expected that there exists a unifying framework for all these problems. Even the Hamilton-Jacobi-Bellman approach which is standard for optimization problems is not only numerically difficult to apply in some situations, but also inapplicable if the preferences are not of the type of expected utility, but for example of the type of risk measure pricing. To find an efficient numerical algorithm which is applicable for all situations in incomplete markets simultaneously has thus not been the issue of this thesis. On the other hand, treating very specific problems, as it has been done in this thesis, also gives the opportunity of finding numerical algorithms which are much more efficient than a standard algorithm developed for a unifying framework.

In this thesis, we have treated numerically two different specific situations in which incompleteness of the market plays a role. In Chapter 4, an indifference price for a stochastic volatility model has been calculated by a nonlinear PDE method using an explicit Finite Difference scheme. It has been shown that this indifference price can be calculated rather precisely in a reasonable time. On the other hand, a Monte Carlo algorithm has turned out to be inefficient for this specific problem.

The problem of CVaR pricing of unit-linked insurance payoffs has been treated in Chapter 5. With some reasonable assumptions, an analytical formula for an upper bound of the CVaR price has been obtained, which can very quickly be calculated numerically, and without any optimization procedure. Furthermore, the corresponding hedge is also given by an analytic formula, and the strategy then is given by the corresponding Delta hedge. The reason why the results are only upper bounds is that we have made the assumption that information about the insurance process is only available at the end of the time period. If this would be true in practice, the bounds would be exact solutions. We have applied the results to a very specific situation. However, we have formulated the main theorems in a general way.

The method to achieve this goal has been a combination of the results of Rockafellar and Uryasev [73] on the CVaR and the ones on expected shortfall minimization of Föllmer and Leukert, [32] and [33]. Furthermore, we have extended one of the main results of [33] to the case of state-dependent utility functions which are not necessarily differentiable nor strictly concave, a result which can in principle be used for any state dependent utility optimization problem in complete markets.

In order to obtain better upper bounds, the information time step can be decreased, and the procedure of Chapter 5 can be applied for all information time steps. In Chapter 6, we have shown how to accomplish this, applying an idea of Föllmer and Leukert [32]. Furthermore, we have proved a convergence result, where we have shown that this discretization converges to the optimal solution for the case where information arrives continuously.

For further research, there are still many open problems, from the theoretical as well as from the practical point of view. Theoretically, the question of equivalence is not completely solved; there are still open questions. First, there does not exist a unifying framework which captures the satiated as well as the unsatiated utility functions. Moreover, even if there exist simple counterexamples to the equivalence result in the case of satiated utility functions and financial price processes allowing jumps, there still is the question if one can find conditions for the general financial market driven by semimartingales in order to have the equivalence result. Finally, this equivalence result only applies to the case where investor preferences are given by expected utility. An interesting question would be what the properties of a pricing measure are when other preference structures are applied, for example according to the dual theory of choice under risk, developed by Yaari [87].

With respect to the numerical aspect of incomplete markets, one may think about a generalization of the algorithms to other situations. In particular, concerning the indifference pricing in a stochastic volatility model, the underlying theory of Hobson [47] does apply very specifically to this situation. For obtaining a suitable partial differential equation for other financial models, for example with jumps or with a stochastic interest rate, or a more general class of utility functions, a new theory, or at least an extension of the theory of [47], would have to be developed.

With respect to the CVaR pricing, one question is whether the assumption could be dropped that the replicable financial process and the non-replicable one from the insurance are independent. This would allow an application of the theory to many other risk measure pricing problems than the one for unit-linked insurance payoffs, for example a model with stochastic volatility with the volatility given by a jump process. Because independence has been a key assumption in this thesis in order to turn the incomplete market problem into one of a complete market, a relaxation of this assumption needs some careful analysis and may lead to a completely different theory. Additionally, the assumptions in the main theorems, in particular the one-dimensionality and the fact that the insurance state space is finite, are likely to be not necessary but only to facilitate the proofs. To find the necessary assumptions for the main theorems would be an interesting topic of future research.

The topic of risk measure pricing has only been treated with CVaR as specific risk measure. In further research, one could thus also consider other (coherent or convex) risk measures, or try to find a general algorithm for the risk measure pricing problem. From Föllmer and Schied [34], it follows that CVaR is a building block for all law-invariant risk measures, and it may be interesting to see whether this fact could be applied in order to obtain a more general risk measure price.

In order to obtain better upper bounds for the CVaR price than in Chapter 5, Chapter 6 gives a theoretical convergence proof in the situation where information arrives continuously. However, this proof does not include any order of convergence. From a numerical point of view, it would be interesting to find such an order of convergence. Furthermore, even if the problem would be theoretically solved, the numerical calculation of better bounds is time-consuming even for a simple model and becomes almost infeasible for a large amount of information time steps. Therefore, it would also be interesting to develop a more efficient numerical treatment of the CVaR pricing problem with continuous insurance information.

Bibliography

[1] Y. Achdou and N. Tchou. Variational Analysis for the Black and Scholes equation with stochastic volatility. *Mathematical Modelling and Numerical Analysis*, 36(3):373–396, 2002.

[2] L. B. G. Andersen. Efficient simulation of the Heston stochastic volatility model. Working paper, Bank of America Securities, New York, 2007.

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

[4] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV spaces*. SIAM and MPS, Philadelphia, 2005.

[5] J. P. Aubin. *Mathematical Methods of Game and Economic Theory*. North-Holland, Amsterdam, 2nd edition, 1982.

[6] S. Basak and A. Shapiro. Value-at-risk-based risk management: Optimal policies and asset prices. *The Review of Financial Studies*, 14(2):371–405, 2001.

[7] F. Bellini and M. Frittelli. On the existence of minimax martingale measures. *Mathematical Finance*, 12(1):1–21, 2002.

[8] S. K. Berberian. *Fundamentals of Real Analysis*. Springer, New York, 1999.

[9] S. Biagini and M. Frittelli. A unified framework for utility maximization problems: An Orlicz space approach. *The Annals of Applied Probability*, 18(3):929–966, 2008.

[10] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[11] S. Bloom and R. Kerman. Weighted Orlicz space integral inequalities for the Hardy-Littlewood maximal operator. *Studia Mathematica*, 110(2):149–167, 1994.

[12] T. Bollerslev, F. Engle, and J. M. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96(1):116–131, 1988.

[13] H. P. Boswijk. Volatility mean reversion and the market price of volatility risk. In P. H. Franses and M. McAleer, editors, *Proceedings of the International Conference on Modelling and Forecasting Financial Volatility*, Perth, 2001. The University of Western Australia.

[14] B. Bouchard, N. Touzi, and A. Zeghal. Dual formulation of the utility maximization problem: The case of nonsmooth utility. *Annals of Applied Probability*, 14(2):678–717, 2004.

[15] P. Carr, H. Geman, and D. B. Madan. Pricing and hedging in incomplete markets. *J. of Financial Economics*, 62(1):131–167, 2001.

[16] P. Carr and D. B. Madan. Option valuation using the fast Fourier transform. *J. of Computational Finance*, 2(4):61–73, 1999.

[17] A. Chen, A. Pelsser, and M. Vellekoop. Approximate solutions for indifference pricing under general utility functions. Working paper, University of Amsterdam and University of Twente, 2008.

[18] P. Cheridito and M. Kupper. Time-consistency of indifference prices and monetary utility functions. Preprint, Princeton University, Princeton, 2006.

[19] A. S. Cherny. Pricing with coherent risk. *Theory of Probability and Its Applications*, 52(3):506–540, 2008.

[20] N. Clarke and K. Parrot. Multigrid for American option pricing with stochastic volatility. *Applied Mathematical Finance, Taylor and Francis Journals*, 6(3):177–195, 1999.

[21] J. H. Cochrane and J. Saá-Requejo. Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy*, 108(1):79–119, 2000.

[22] R. Cont and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, Boca Raton, Florida, 2004.

[23] F. Delbaen. Coherent risk measures on general probability spaces. In K. Sandmann and P.J. Schönbucher, editors, *Essays in Honour of Dieter Sondermann*, Heidelberg, 2002. Springer Verlag.

[24] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300:463–520, 1994.

[25] F. Delbaen and W. Schachermayer. The variance-optimal martingale measure for continuous processes. *Bernoulli*, 2(1):81–105, 1996.

[26] F. Delbaen and W. Schachermayer. The fundamental theorem of asset pricing for unbounded processes. *Mathematische Annalen*, 312(2):215–250, 1998.

[27] D. Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, Princeton, 3rd edition, 2001.

[28] E. Eberlein and F. Özkan. Time consistency of Lévy models. *Quantitative Finance*, 3(1):40–50, 2003.

[29] E. Eberlein and F. Özkan. The Lévy Libor Model. *Finance and Stochastics*, 9(3):327–348, 2005.

[30] E. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(4):34–105, 1965.

[31] G. Fiorentini, A. Leon, and G. Rubio. Estimation and empirical performance of Heston's stochastic volatility model: the case of a thinly traded market. *Journal of Empirical Finance*, 9(2):225–255, 2002.

[32] H. Föllmer and P. Leukert. Quantile hedging. *Finance and Stochastics*, 3(3):251–273, 1999.

[33] H. Föllmer and P. Leukert. Efficient hedging: Cost versus shortfall risk. *Finance and Stochastics*, 4(2):117–146, 2000.

[34] H. Föllmer and A. Schied. *Stochastic Finance.* Walter de Gruyter, Berlin, 2nd edition, 2004.

[35] J.-P. Fouque, G. Papanicolaou, and K. R. Sircar. *Derivatives in Financial Markets with Stochastic Volatility.* Cambridge University Press, Cambridge, 2000.

[36] M. Frittelli. The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10(1):39–52, 2000.

[37] P. Glasserman. *Monte Carlo Methods in Financial Engineering.* Springer Verlag, New York, 2003.

[38] M. R. Grasselli and T. R. Hurd. A Monte Carlo Method for exponential hedging of contingent claims. Technical report, McMaster University, 2006.

[39] A. Gundel and S. Weber. Utility maximization under a shortfall risk constraint. Preprint, Humboldt-Universität zu Berlin and Cornell University, 2005.

[40] J. Haezendonck and M. Goovaerts. A new premium calculation principle based on Orlicz norms. *Insurance: Mathematics and Economics*, 1(1):41–53, 1982.

[41] V. Henderson. Valuation of claims on nontraded assets using utility maximization. *Mathematical Finance*, 12(4):351–373, 2002.

[42] V. Henderson. Analytical comparisons of option prices in stochastic volatility models. *Mathematical Finance*, 15(1):49–59, 2005.

[43] V. Henderson, D. Hobson, and S. Howison. A comparison of option prices under different pricing measures in a stochastic volatility model with correlation. *Review of Derivatives Research*, 8:5–25, 2005.

[44] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[45] N. Hilber, A.-M. Matache, and C. Schwab. Sparse wavelet methods for option pricing under stochastic volatility. *J. of Computational Finance*, 8(4):1–42, 2005.

[46] J.-B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms*. Springer, New York, 2nd edition, 1996.

[47] D. Hobson. Stochastic volatility models, correlation and the q-optimal measure. *Mathematical Finance*, 14(4):537–556, 2004.

[48] S. Hodges and A. Neuberger. Optimal replication of contingent claims under transaction costs. *Review of Futures Markets*, 8(2):222–239, 1989.

[49] A. Ilhan, M. Jonsson, and R. Sircar. Optimal static-dynamic hedges for exotic options under convex risk measures. Working Paper, Goldman Sachs International, University of Michigan and Princeton University, 2008.

[50] A. Jobert and L. C. G. Rogers. Valuations and dynamic convex risk measures. *Mathematical Finance*, 18(1):1–22, 2008.

[51] D. Kabanov and M. Sirbu. Sensitivity analysis of utility-based prices and risk-tolerance wealth processes. *The Annals of Applied Probability*, 16(4):2140–2194, 2006.

[52] Y. Kabanov and Chr. Stricker. On the optimal portfolio for the exponential utility maximization: Remarks to the six-author paper. *Mathematical Finance*, 12(2):125–134, 2002.

[53] I. Karatzas, J. Lehoczky, S. Shreve, and G. Xu. Martingale and duality methods for utility maximization in an incomplete market. *SIAM Journal on Control and Optimization*, 29(3):702–730, 1991.

[54] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2nd edition, 1998.

[55] S. Klöppel and M. Schweizer. Dynamic indifference valuation via convex risk measures. *Mathematical Finance*, 17(4):599–627, 2007.

[56] D. Kramkov and W. Schachermayer. The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *The Annals of Applied Probability*, 9(3):904–950, 1999.

[57] D. O. Kramkov. Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probability Theory and Related Fields*, 105:459–479, 1996.

[58] D. M. Kreps. Arbitrage and equilibrium in economies with infinitely many commodities. *Journal of Mathematical Economics*, 8(1):15–35, 1981.

[59] N. V. Krylov. Approximating value functions for controlled degenerate diffusion processes by using piece-wise constant policies. *Electronic Journal of Probability*, 4:1–19, 1999.

[60] J. Lim. A numerical algorithm for indifference pricing in incomplete markets. Technical report, University of Texas at Austin, 2006.

[61] B. Melenberg and B. J. M. Werker. The implicit price of volatility risk: An empirical analysis. Working paper, Tilburg University, Tilburg, 2001.

[62] A. Melnikov and V. Skornyakova. Pricing of equity-linked life insurance contracts with flexible guarantees. Technical Report 1/04, Department of Mathematics and Statistics, Concordia University, Montreal, Canada, 2004.

[63] R. C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

[64] T. Møller. Risk-minimizing hedging strategies for insurance payment processes. *Finance and Stochastics*, 5(4):419–446, 2001.

[65] T. Møller. On valuation and risk management at the interface of insurance and finance. *British Actuarial Journal*, 8(4):787–827, 2002.

[66] M. Monoyios. Efficient option pricing with transaction costs. *J. of Computational Finance*, 7(1):107–128, 2003.

[67] M. Monoyios. Characterization of optimal dual measures via distortion. Technical report, Department of Mathematics, Imperial College, London, 2005.

[68] J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1):122–136, 1964.

[69] Ph. Protter. *Stochastic Integration and Differential Equations.* Springer Verlag, Heidelberg, 2nd edition, 2005.

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

[71] R. T. Rockafellar. Maximal monotone relations and the second derivatives of nonsmooth functions. *Ann. Inst. Henri Poincaré*, 2(3):167–184, 1985.

[72] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000.

[73] R. T. Rockafeller and S. Uryasev. Conditional Value at Risk for general loss distributions. *J. of Banking and Finance*, 26(7):1443–1471, 2002.

[74] M. Romano and N. Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–410, 1997.

[75] M. Rubinstein. Non-parametric tests of alternative option pricing models using all reported trades and quotes. *The Journal of Finance*, 40(2):455–480, 1985.

[76] W. Schachermayer. Optimal investment in incomplete markets when wealth may become negative. *The Annals of Applied Probability*, 11(2):694–734, 2001.

[77] U. Schmock, S. E. Shreve, and U. Wystup. Valuation of exotic options under shortselling constraints. *Finance and Stochastics*, 6(2):143–172, 2002.

[78] M. Schweizer. Option hedging for semimartingales. *Stochastic Processes and their Applications*, 37(2):339–363, 1991.

[79] U. Segal and A. Spivak. First-order risk aversion and non-differentiability. *Economic Theory*, 9(1):179–183, 1997.

[80] J. Sekine. Dynamic minimization of worst conditional expectation of shortfall. *Mathematical Finance*, 14(4):605–618, 2004.

[81] R. Seydel. *Tools for Computational Finance.* Springer Verlag, Berlin, 2nd edition, 2004.

[82] A. N. Shiryaev. *Probability.* Springer, New York, 2nd edition, 1996.

[83] D. Williams. *Probability with Martingales.* Cambridge University Press, Cambridge, 1991.

[84] P. Wilmott, J. Dewynne, and S. Howison. *Option Pricing: Mathematical Models and Computation.* Oxford Financial Press, Oxford, 1995.

[85] H. Witting. *Mathematische Statistik 1.* Teubner Verlag, Stuttgart, 1985.

[86] M. Xu. Risk measure pricing and hedging in incomplete markets. *Annals of Finance*, 2(1):51–71, 2006.

[87] M. E. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.

Summary

This thesis addresses questions on pricing and hedging in incomplete financial markets. In such markets, not all contingent claims can be hedged, and the pricing measure is not uniquely determined from arbitrage considerations. Both indifference pricing and risk measure pricing are included in the scope of this thesis.

The thesis contains a theoretical as well as a practical part, where specific financial models as well as numerical issues are discussed. In the theoretical part, the question is treated whether the indifference pricing measure is equivalent to the objective measure. Complementary to the existing literature, the thesis answers this question for utility functions which are nondecreasing but not strictly increasing, i.e. functions which have a satiation point. It is shown how this question is connected to the relative risk aversion of the utility function.

In Chapter 2, a general definition of the relative risk aversion is provided, which is applicable for all nondecreasing and concave utility functions. In particular, this notion also applies to nonsmooth and not necessarily strictly increasing utility functions. Using this definition, it is shown that the main statements on risk aversion continue to hold for nonsmooth utility functions.

In Chapter 3, the results of Chapter 2 are applied in order to prove that the indifference pricing measure is equivalent to the physical measure for satiated utility functions in the case where the filtration is continuous, provided that some assumptions about the relative risk aversion hold, and that there exists an equivalent local martingale measure which has sufficient integrability properties. This generalizes an earlier result of Delbaen and Schachermayer, where the equivalence of the variance-optimal measure has been proved. Other examples which were not known before are q-optimal measures and one-sided risk functions.

In the practical part, two different specific models are treated. The first one is indifference pricing in a stochastic volatility model, where only the underlying stock and a risk-free asset can be used for trading. This is the topic of Chapter 4. Here we consider the class of power utility functions. A recent theoretical result of Hobson is used which states that the price of volatility risk corresponding to the indifference

pricing measure can be obtained by a nonlinear partial differential equation. A method based on Monte Carlo simulation and an explicit Finite Difference method are compared with respect to their numerical efficiency. It is shown that the Finite Difference method produces quite accurate results in reasonable time, whereas the Monte Carlo method is too slow.

In Chapter 5, the financial market is assumed to be complete with respect to the underlying assets, but there additionally exists an insurance risk which cannot be hedged. Instead of indifference pricing, risk measure pricing is used to calculate a price, with CVaR as specific risk measure. Under the assumption that nonfinancial information is only available at the end of the time period, a straightforward algorithm is obtained for obtaining the CVaR price as well as the corresponding hedge, which does not involve any optimization algorithm. For some specific models, analytical formulas are provided both for the CVaR price and for the corresponding hedge.

In Chapter 6, the CVaR pricing method is further developed to general continuous-time non-financial processes, with a discretization of the time at which information is available. At each information time step, the results of Chapter 5 are applied. It is proved that this discretization converges to the true continuous-time value, and for a specific model, the CVaR price is numerically calculated by a discrete Legendre transformation. The main difficulty of this method is that, in particular for many information steps, the numerical effort for the computation becomes very large. On the other hand, for only few information time steps, the method can be rather efficient.

Samenvatting

Dit proefschrift gaat over waardering en risico-afdekking in onvolledige financiële markten. In zulke markten kan niet ieder kontrakt worden afgedekt door te handelen in bestaande financiële produkten, en de kansmaat die voor waardering wordt gebruikt valt niet te bepalen op basis van alleen overwegingen ontleend aan afwezigheid van arbitragemogelijkheden.

Het proefschrift bestaat uit een theoretisch deel en een praktisch deel, waarin specifieke financiële modellen en kwesties van numerieke aard worden besproken. In het theoretische deel wordt de vraag aan de orde gesteld wanneer de kansmaat voor waardering op basis van het indifferentieprincipe equivalent is met de objectieve kansmaat. In aanvulling op de bestaande literatuur wordt hierop een antwoord gegeven voor nutsfuncties die niet dalen maar die ook niet strikt stijgend zijn, in andere woorden, voor nutsfuncties met een verzadigingspunt. Getoond wordt hoe de besproken vraag verband houdt met de relatieve risico-aversie die door de gegeven nutsfunctie wordt gedefinieerd.

Hoofdstuk 2 bevat een algemene definitie van relatieve risico-aversie die van toepassing is op alle niet-dalende en concave nutsfuncties. Het hierdoor gegeven begrip is in het bijzonder toepasbaar op niet-gladde nutsfuncties en op niet strikt stijgende nutsfuncties. Aangetoond wordt dat de belangrijkste relaties die gelden voor risico-aversie nog steeds opgaan als van de algemenere definitie gebruik wordt gemaakt.

In hoofdstuk 3 worden de resultaten van hoofdstuk 2 gebruikt in het bewijs van een theoretisch resultaat dat zegt dat de kansmaat voor waardering op basis van het indifferentieprincipe equivalent is met de natuurlijke kansmaat in het geval van nutsfuncties met een verzadigingspunt als de gegeven filtratie continu is, mits verder enkele aannames over de relatieve risico-aversie vervuld zijn en de markt een lokale martingaalmaat toelaat die equivalent is met de natuurlijke kansmaat en die voldoet aan zekere integreerbaarheidseigenschappen. Dit is een generalisatie van een resultaat van Delbaen en Schachermayer betreffende equivalentie van de variantie-optimale maat. Eveneens volgen nieuwe uitspraken over q-optimale kansmaten en

eenzijdige risicofuncties.

In het praktische gedeelte worden twee verschillende specifieke modellen behandeld. De eerste toepassing betreft waardering op basis van het indifferentieprincipe in een model met stochastische volatiliteit, waarbij wordt verondersteld dat alleen gehandeld kan worden in het onderliggende aandeel en in een risicovrije belegging. Dit is het onderwerp van hoofdstuk 4. We gaan hierbij uit van de klasse van nutsfuncties die door machten worden beschreven. Gebruik wordt gemaakt van een recent theoretisch resultaat van Hobson, volgens welke de prijs van volatiliteitsrisico die volgt uit het indifferentieprincipe kan worden bepaald door het oplossen van een niet-lineaire partiële differentiaalvergelijking. In het hoofdstuk wordt een Monte Carlo-methode en een expliciete eindige differentiemethode met elkaar vergeleken aangaande numerieke efficiëntie. Aangetoond wordt dat de eindige differentiemethode leidt tot vrij nauwkeurige resultaten bij een redelijke rekentijd, terwijl de Monte Carlo-methode te traag is.

In hoofdstuk 5 wordt uitgegaan van een situatie waarin de financiële markt op zich volledig is in termen van beschikbare financële produkten, maar waarin ook sprake is van een verzekeringsrisico dat niet kan worden afgedekt. Als waarderingsprincipe wordt nu niet het indifferentieprincipe toegepast, maar een principe gebaseerd op het gebruik van een risicomaat. In het bijzonder wordt de risicomaat CVaR gebruikt. Onder de aanname dat niet-financiële informatie alleen beschikbaar is aan het eind van de kontraktperiode wordt een eenvoudig implementeerbaar algoritme verkregen waarmee de prijs op basis van CVaR en de bijbehorende afdekkingsstrategie kunnen worden berekend zonder gebruik te maken van een optimalisatieroutine. Voor enkele specifieke modellen worden analytische uitdrukkingen gegeven zowel voor de CVaR-prijs als voor de afdekkingsstrategie.

In hoofdstuk 6 wordt de waarderingsmethode op basis van CVaR verder ontwikkeld voor toepassingen waarbij het verzekeringsproces plaatsvindt in continue tijd, terwijl informatie hierover in de tijd is gediscretiseerd. De methode van hoofdstuk 5 wordt toegepast op elk tijdstip waarop informatie beschikbaar komt. Een convergentieresultaat wordt aangetoond voor de limiet waarin informatie vaker en vaker beschikbaar is. Voor een specifiek model wordt de CVaR numeriek berekend met behulp van een discrete Legendre-transformatie. De gepresenteerde methode wordt zeer rekenintensief bij toename van het aantal informatiestappen, maar voor situaties waarin informatie beschikbaar komt op slechts enkele tijdstippen is de methode tamelijk efficiënt.