A dissertation submitted to
the Faculty of Business Administration and Economics of Bielefeld University
in partial fulfillment of the requirements for the doctoral degree Dr. rer. pol.

Essays on Optimal Stopping and Stochastic Control in Finance

LAZAR OBRADOVIĆ

Center for Mathematical Economics
Faculty of Business Administration and Economics
Bielefeld University

January 2018

Essays on Optimal Stopping and Stochastic Control in Finance

Dissertation zur Erlangung des Doktorgrades
der Wirtschaftswissenschaften (Dr. rer. pol.)
an der Fakultät für Wirtschaftswissenschaften
der Universität Bielefeld
vorgelegt von
Lazar Obradović

Erstgutachter:
Prof. Dr. Frank Riedel

Zweitgutachter:
Jun. Prof. Dr. Jan-Henrik Steg

Januar 2018

Gedruckt auf alterungsbeständigem Papier nach DIN-ISO 9706

Acknowledgements

I feel very fortunate to have had Frank Riedel as my supervisor. All his insight and guidance throughout the years served as invaluable lessons on being a researcher, an academic and a leader. I am very grateful for the whole process with him and, in particular, for all the opportunities he created for me.

I want to especially thank my second supervisor Jan-Henrik Stegg for his endless understanding, kindness and encouragement.

I gratefully acknowledge funding and support from DAAD, CRC 1283 and the *Robust finance* research group. Special thanks to ZiF.

I owe so much to my dear friend, the greatest office mate, and my *Wilde Kuh* partner Oliver Claas. He is extremely well versed at selfless actions, ingenious problem solving and perfect advice. The friendship with Oli is the best byproduct of my PhD studies.

Finalizing this thesis would have been absolutely impossible without my friend Jennifer Rasell and her generous editing. There are no enough fun summery things for me to come up with to thank her. Ever.

Floor W10 of Bielefeld University is legendary (albeit in very narrow circles) - I thank all of those who occupied it during my PhD studies. Thanks to Chris (a dear friend, semi-genius and Sarah's husband), Nikoleta (my understanding compatriot and initial office mate), Tim (for all the chess), Freddy, Katha, Pascal, Jörg, Martin and Johannes (for all the coffees, beers and conversations), and Jodi, Patrick and Torben (the awesome next generation). Thanks also to the great Giorgio Ferrari and to the one and only Walter Trockel for being Walter Trockel.

Over the course of time it took to write this thesis three different deans were at the helm of my employer and alma mater *Faculty of Science and Mathematics* of *University of Montenegro*: Predrag Stanišić, Žana Kovijanić and Predrag Miranović. I am indebted to all of them for their support and understanding. I am also grateful to Oleg Obradović for all the encouragement, especially when things were tough.

I am extremely fortunate to have so many good people working with me. Aleksandar is my spiritual animal and my career guru. Nikola is an epic lifesaver who I will never be able to thank adequately. I am grateful to my Münster buddies Jelena and Milica, to my great office buddy Dragana, to Goran, Anton, Darko, Dušica and to the rest of PMF for all kinds of help. Special thanks to Vladimir Božović.

My parents are Duda and Deki (they are one), my sisters are Jecka the Genije and Dagili the Pet Saver. I love them and they love me, and that helps a lot when you are writing a thesis for six years. It helps with other things, too. And it really doesn't hurt to have the best grandparents ever, Ogi and Miko, in your corner at all times, no matter what. I thank families Djurašković and Krgović for their love and support.

I hereby acknowledge that Kum Bojan, Kumara Igor and Kuma Jovana are well groomed and generally awesome. Thanks to Milena (for all the sausages and, sure, twenty years of friendship), dr. Marko Mengesh Radišić, Jelena, Adi, Dado, Miraš, Saška and Girica. I really appreciate your sustained curiosity about the progress of my PhD and am certain that every single one of you will read this thesis, cover to cover, now that it has finally seen the light of day.

Lea is the best. *The best.*

January 2018
Podgorica

Contents

Chapter 1

General Introduction

In one way or another, most of the literature on mathematics of finance pertains to making optimal decisions under uncertainty. The uncertainty in finance, although inevitable, is inconvenient: it is intangible, hard to quantify, and it significantly complicates decision making. It is worth noting that usually the focus of research in mathematical finance is not on the uncertainty itself, but rather on optimal decisions. However, what is optimal in a given situation depends heavily on the way the uncertainty is modelled. The framework one works in, or merely the way the question is posed, can naturally lead to different models of uncertainty, which in turn influence conclusions about what the optimal behavior in a certain scenario is.

This thesis is a result of different investigations[1] into the relationship between the way uncertainty is taken into account by mathematical models in theoretical finance and the optimal behavior prescribed by the solutions of those models. Our focus is on the dynamical setting in both continuous and discrete time. We present three self contained essays that, although different in topics, explore interconnected ideas related to optimization under multiple priors, risk measures, optimal stopping and stochastic control in finance. We

[1] The seminal early paper in modern mathematical finance by Samuelson (1965) begins modestly with: "This is a compact report on desultory researches stretching over more than a decade". Given the long and winding road of this thesis' creation, it could have easily begun in the exact same spirit.

contribute to well known models of pricing financial derivatives, best choice problems, coherent risk measures and optimal portfolios, among others.

Each of the chapters that follow contains one essay. Before we present them, we briefly review the literature of relevant fields of research: we hope to demonstrate the relationship between our contributions on the one hand and the established ideas and contemporary research questions on the other.

1.1 Uncertainty as a Single Probability Measure

Traditionally, uncertainty in finance is modelled as a single probability measure, also called a *prior*. This classical worldview is mathematically very convenient: one assumes that all the uncertainty relevant to the situation being modeled is perfectly captured by the one probability measure.

There are several implicit assumptions that are made when using this approach and they depend on the interpretation of the situation being modelled, the model itself, and even on the meaning and interpretation of probability.

From an objective point of view, it is implicitly assumed that the probability of each relevant event is measurable and known: the "true probability measure" exists, and it completely describes all the randomness of the world. A classical justification for this view is that the agents using the model have sufficient data to make correct estimations of the probability measure. It is further assumed that this probability measure is fixed even in dynamical models: calculations regarding classical updating of probabilities due to the passage of time and arrival of new information is performed using the one probability measure.

From a subjective point of view, the probabilities are merely beliefs about the likelihood of certain events. This is the approach of subjective expected utility following Savage (1954) and Von Neumann and Morgenstern (1953). After accepting what seem to be reasonable axioms, it follows that each agent is capable of assigning a probability to each event in a manner that is

mathematically consistent. Furthermore, in this context it is assumed that each agent also has a concave and increasing utility function that they are perfectly aware of.

Using a single prior often implies further implicit assumptions. Notably, if a model with a single prior is dynamical it involves stochastic processes, the distribution of which is assumed to be known, and it is furthermore assumed that the values of the parameters of the model (for example values of the drift and diffusion coefficients in geometric Brownian motion) are somehow known.

As can be seen, this approach is quite idealistic and not very robust. Indeed, assuming that uncertainty is completely measurable and, in a certain sense, completely known has been criticized throughout the literature. In the influential work by Knight (1921) this kind of uncertainty is referred to as *risk* and it is contrasted with *ambiguity* which is considered to be intangible and unmeasurable[2]. Ambiguity is clearly not captured by considering only one probability measure to describe uncertainty in a model. In fact, there are serious reasons to question all of the above assumptions and maxmin expected utility theory and the theory of risk measures address most of them from different perspectives with a similar mathematical foundation. Before we discuss both approaches in sections 1.3 and 1.4 we briefly review the classical applications of stochastic control and optimal stopping in the mathematics of finance.

1.2 Stochastic Control and Optimal Stopping in Mathematical Finance

From its conception the theory of probability was used to analyze possible financial gains in the face of uncertainty. For example, a very early problem

[2]Indeed, the terms *ambiguity* and *Knightian uncertainty* are often used interchangeably.

in probability known as the *problem of points*[3] from the seventeenth century is, in the language of contemporary finance, a problem of pricing a binary option on a binomial tree. The pricing of financial derivatives remains one of the most important and actively researched topics in contemporary finance. In particular, the pricing of options of the American type can be reduced to an optimal stopping problem; solving the problems of that type also prescribes if and when early exercise is optimal.

The theory of optimal stopping investigates problems of choosing the optimal time, called a *stopping time*, to take a certain action (selling an asset, making an irreversible investment), while the theory of stochastic control investigates problems where actions that influence the values in the model are needed at each point in time (choosing the optimal portfolio, or of consumption) in order to maximize some gain function. As such, the theory of optimal stopping can be considered a special (constrained) case of the theory of optimal stochastic control.

Optimal stopping theory was first developed in discrete time by Wald (1945) and Snell (1952); in that sense early problems of optimal stopping in discrete time can be counted among precursors to the modern theory of option pricing. Among the best known optimal stopping problems in discrete time are best choice problems. They date back to at least the 1950s[4] and are still actively researched, often in explicitly financial contexts as in Bruss, Ferguson, et al. (2002). The best known among them are the *secretary problem* and the *full information best choice problem*[5]. Both of these models deal with detecting the maximum of a certain random sequence; in the next chapter of this thesis we will explore a robust version of the latter.

An important financial application of the theory of optimal stopping in discrete time is pricing of American options on a binomial tree model of a financial market (see Cox et al. (1979)). In this simple model prices are

[3] See, for example, Hald (2003).

[4] Early history of best choice problems is vague; an amusing and detailed review can be found in Ferguson (1989).

[5] See, for example, Gilbert and Mosteller (1966).

modelled as an asymmetric random walk. It allows for explicit calculations of optimal exercises times and values of virtually any option. It is shown in Cox et al. (1979) that the price process, under the appropriate limiting procedure, converges to geometric Brownian motion, one of the standard models for the asset value movement in continuous time.

Use of geometric Brownian motion in finance to model the movement of the price of an asset in continuous time was first introduced in Samuelson (1965). Applications to the theory of option pricing were first presented in seminal papers by Black and Scholes (1973) (for European call and put options) and Merton (1973) (for perpetual American options). In Black and Scholes (1973) geometric Brownian motion was used to model the movement of prices, and the authors proceeded to calculate the value of a European option by a hedging argument. The pricing of American options in the same setting required the theory of optimal stopping in continuous time. In Merton (1973) the price of the perpetual American put is calculated using what is essentially a free boundary approach to optimal stopping problems. This approach can be used for pricing other perpetual American options[6]. In the last chapter of this thesis we give a modest contribution to that theory by improving on the well known results on the *perpetual American Straddle*.

One of the classical applications of stochastic control theory in finance pertains to the choice of the optimal portfolio. A pioneering paper in this area is Merton (1969) where it was described, under the assumption that asset prices follow a geometric Brownian motion, how one should invest in order to maximize the utility of terminal wealth; this problem and its extensions is known as the *Merton portfolio problem*. The elegant solution led to a simple, proportional optimal portfolio now known as the *Merton portfolio*. The extensions and generalizations have since been numerous, an extensive review can be found in Rogers (2013). In the second chapter of this thesis, we will address what can be viewed as a formulation of Merton's problem

[6] A detailed introduction to the theory with multiple examples can be found in Peskir and Shiryaev (2006).

in the context of coherent risk measures. This was already done in, among others, Gambrah and Pirvu (2014); in addition to presenting our own results we will improve on their work, too.

We note that the theory reviewed in this section is based on the classical (single prior) theory of optimal stopping and stochastic control. In the next two sections we discuss two different, but related approaches that involve *multiple priors*.

1.3 Maxmin Expected Utility Theory

There are serious drawbacks to using a single probability measure in models of financial mathematics. Indeed, one can never be certain that the measure in use is the "correct one", or even that it exists. Even if one assumes it does – estimating the measure and its parameters is prone to errors and can have significant effects in models that are not sufficiently robust. From a more subjective point of view, it was demonstrated that even in simple experiments economic agents do not act as if they would assign a subjective probability measure to possible outcomes. That kind of behavior violates certain axioms of subjective expected utility theory, but the observed behavior of the agents in the experiment is far from irrational. An early and influential paper in this field of research is Ellsberg (1961), but the literature on violations of the subjective probability paradigm is vast.

The theory of *maxmin expected utility* is one of the approaches that addresses the drawbacks of using a single prior. It was axiomatically founded in Gilboa and Schmeidler (1989), by extending the framework of subjective expected utility of Savage (1954) to include *ambiguity aversion*. It can be shown that in the setting of maxmin expected utility theory the agent optimizes with respect to a set of priors: the optimal decisions can be interpreted as a result of a certain maxmin procedure in which the agent does not choose the optimal decisions by calculating the probabilities and expectations with respect to a single measure, but rather with respect to a set of measures.

Every measure in the set leads to a different optimal decision and the agent acts cautiously – by choosing the least favorable optimal decision, i.e. as if the measure that gives the worst payoff is the "correct one".

The decision making framework of maxmin expected utility is fruitful and goes well beyond finance[7]. It was originally formulated in a static, single period setting, but has been extended dynamically in Epstein and Schneider (2003). Extending the theory to multiple periods, and further to continuous time, is not trivial. In particular, the set of priors used by the agent needs to satisfy certain regularity conditions that avoid dynamical inconsistencies. This regularity condition is known as *time consistency*[8] and it, informally speaking, ensures that at any point the agent can "change their mind" about what they perceive to be the "worst measure from now on", without changing their mind about what they perceived the "worst measure" was up to that point in time. In the next chapter of this thesis we give a contribution in this direction by describing an explicit procedure for constructing a time consistent set of priors for discrete time dynamical models by pasting together single period sets of priors. Our approach can be considered a generalization of similar procedures that appear in Riedel (2009), Chudjakow and Riedel (2013) and references therein.

Optimal Stopping under Multiple Priors

The theory of optimal stopping in discrete time was extended to maxmin expected utility setting in Riedel (2009). In this stetting instead of just maximizing the payoff over stopping times the agent also minimizes over the set of measures. Extending the ideas of the classical optimal stopping theory required developing the multiple prior version of martingales and proving the multiple prior versions of results on iterated conditional expectations, Doob decomposition and optional sampling theorem. Existence of the solutions of

[7]For a recent review of ambiguity aversion literature and maxmin expected utility theory's place within it see Gilboa and Marinacci (2016).

[8]For different formal definitions of time consistency see lemma 8 in Riedel (2009).

optimal stopping problems under multiple priors is proven, under suitable conditions, in both finite and infinite time. The solutions are characterized via a multiple priors version of the Snell envelope and the smallest multiple prior supermartingale dominating the payoff process. With these results established, it is possible to prove a maxmin duality result: maximizing over the stopping times and then minimizing over the set of measures or the other way around leads to the same solution.

The theory of optimal stopping under multiple priors is quite general and allows for wide applications. In Riedel (2009) the problems of pricing various options in a multiple priors version of a binomial pricing model are considered. The author uses ideas from statistics to introduce a set of priors called *exponential neighborhood* and all of the examples solved in that work are part of it. The *secretary problem*, probably the best known optimal stopping problem in discrete time, is solved in Chudjakow and Riedel (2013). In the second chapter of this thesis we will perform a similar, but significantly more involved analysis for the *full information best choice problem*.

1.4 Risk Measures

The theory of monetary risk measures explores ways of offsetting the risk of a financial position with a cash reserve. The classical example is value-at-risk (VaR), a quantile of possible loss of the position. This risk measure has been and still is used in practice by financial institutions and regulatory agencies. Theoretical and practical shortcomings of VaR are well documented and have even been counted among the causes of the 2007-08 financial crisis[9].

The theory of coherent risk measures begins with Artzner et al. (1999). It is mathematically similar to maxmin expected utility: under elegant axiomatic requirements (monotonicity, cash invariance, homogeneity and subadditivity) on the monetary risk measures it can be proven that the risk of a financial position can be represented as the largest expected loss calculated

[9]See for example Sollis (2009).

with respect to a set of probability measures[10]. This result allows the theory of risk measures to be interpreted as an alternative way to address the shortcomings of considering a single prior in financial analysis: by imposing reasonable requirements on a financial position the risk is estimated robustly in monetary terms. On a technical level, risk measures allow for models that are, in a certain sense, between the single prior setting and maxmin utility expectation approach. Examples include applications of stochastic control to optimal investment problems with risk constraints as in Emmer et al. (2001) (single period) and Gambrah and Pirvu (2014) (continuous time). In these models, the dynamic and the probability of the model are assumed to be known, just like in the traditional case, but risk measure requirements are absorbed into the optimization criteria. One could argue that this use of risk measures increases the robustness of the models around both the assumed measure and the assumed values of the parameters of the model. We explore these ideas in the third chapter of this thesis by introducing a new risk measure and examining optimal portfolios related to it.

1.5 Thesis Outline and Contributions

The contributions of this thesis range from purely technical (as in chapter 4), to solving a well known problem in a new framework (as in chapter 2), to the introduction of what can be considered a new concept (as in chapter 3).

In chapter 2 we solve the classical full information best choice (FIBC) problem (from Gilbert and Mosteller (1966)) under multiple priors. In order to do so several contributions to the theory of optimal stopping under multiple priors itself are presented. First, an explicit procedure for the construction of the set of priors is introduced, allowing one to consider more general sets of priors than have been previously considered in the literature.

[10]The results of coherent risk measures have been extended to convex risk measures where convexity of the risk measure is required in addition to monotonicity and cash invariance. For a detailed introduction to the theory of risk measures see Föllmer and Schied (2011).

We also generalize results on monotone problems from Riedel (2009) in multiple ways: by considering monotonicity in a wider sense and by allowing for a more general set of priors. Finally we show how any non-adapted optimal stopping problem under multiple priors problem can be reduced to an equivalent adapted problem; it is in fact a more general version of the procedure that appears in Chudjakow and Riedel (2013).

The FIBC problem is about detecting the highest valued realization in a sequence of finitely many independent and identically distributed random variables. We offer a contemporary interpretation of the problem in the context of a venture capitalist's optimal investment. We formally formulate and solve the problem under multiple priors: by a suitable recursive procedure we fully characterize the (minimal) optimal stopping time and the minimizing measure. We contrast the results to the similar investigation of the secretary problem in Chudjakow and Riedel (2013); in that work, the agent could stop earlier or later depending on the shape of the set of multiple priors while in our investigations the agent always stops earlier.

The analysis we perform involves two sets of priors. The first is the already known exponential neighborhood defined in Riedel (2009). The second is a set of priors that we introduce: *locally constant uncertainty neighborhood*. It is a set of priors that is defined using ideas from risk measures by describing the ambiguity about "small probabilities" of any set to be within a certain suitably chosen interval. This is achieved by considering the set of Radon-Nikodym derivatives of the measures "close to" the original measure. It is worth noting that the set is essentially different from the other sets of priors used in problems of optimal stopping under multiple priors in that it cannot be parametrized by even countably many real parameters. Identifying the minimizing measure under the locally constant uncertainty neighborhood is not trivial and requires solving several problems that are essentially problems of (deterministic) optimal control (see lemma 2.4.1).

In the third chapter we turn to exploring the locally constant uncertainty neighborhood set of priors from the perspective of risk measure. Naturally,

it defines a coherent measure that we call *locally constant model uncertainty* (LCMU). The measure can, due to its relation to optimization under multiple priors, be considered a measure of the uncertainty of the model (the assumed probability measure) that is in a certain sense locally constant.

The chapter essentially consists of two parts. In the first part we explore the technical properties of the LCMU risk measure and establish a connection with the well known average value-at-risk: we prove that LCMU can be represented as a convex combination of expected loss and average value-at-risk which further establishes its relevance.

The second part of the chapter deals with optimal investment problems with respect to LCMU. We consider a frictionless market with multiple securities, the dynamics of which are time dependent. We solve three problems of optimal investment that include minimizing the risk (prescribed by LCMU) and maximizing expected profits under the risk constraint. The results lead to Merton portfolios (Merton (1969)); this rather surprising fact has already been proven for value-at-risk and average value-at-risk in Gambrah and Pirvu (2014). Even more surprising is that, although the risk measures LCMU and AVaR both lead to Merton portfolios, the optimal portfolios can be radically different under the two measures. This leads to an interesting discussion about the nature of modelling optimal portfolio problems and use of risk measures in such models; theorem 3.3.1 and the discussion that follows addresses this.

The final chapter on the perpetual American straddle gives a modest contribution to the literature of perpetual American options in the Black-Scholes-Merton model by presenting a compact way to characterize the price of that particular portfolio. Technically, we solve an optimal stopping problem using the standard approach that involves the Hamilton-Jacobi-Bellman equation combined with the smooth pasting conditions and a verification theorem.

While the focus of the essays varies, each explores the way uncertainty affects optimal behaviour from different angles. The second chapter exam-

ines a well known problem under a more general version of uncertainty that makes the problem more robust from both subjective and objective perspectives. We compare our results with the classical version of the problem, thus providing a deeper insight into the effects of uncertainty (and the way it is modelled!) on the optimal behavior. In the third chapter we introduce a risk measure that, although closely related to concepts already known and well studied in the literature on risk measures and maxmin expected utility, can be considered a new way to describe the uncertainty about the model. Again, the effects of this way of modelling uncertainty on optimal behavior is explicitly contrasted to similar relevant models with surprising conclusions. Finally, the technical results of the last chapter show how even in well known models there is still room for simplifying the characterizations of optimal behavior under risk, i.e. a classical model of uncertainty. Hopefully, the three essays convincingly show that the way the uncertainty is modelled has profound effects on the optimal decisions in the context of finance.

References

Artzner, Philippe et al. (1999). "Coherent measures of risk". In: *Mathematical finance* 9.3, pp. 203–228.

Black, Fischer and Myron Scholes (1973). "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3, pp. 637–654.

Bruss, F Thomas, Thomas S Ferguson, et al. (2002). "High-risk and competitive investment models". In: *The Annals of Applied Probability* 12.4, pp. 1202–1226.

Chudjakow, Tatjana and Frank Riedel (2013). "The best choice problem under ambiguity". In: *Economic Theory* 54.1, pp. 77–97.

Cox, John C, Stephen A Ross, and Mark Rubinstein (1979). "Option pricing: A simplified approach". In: *Journal of financial Economics* 7.3, pp. 229–263.

Ellsberg, Daniel (1961). "Risk, ambiguity, and the Savage axioms". In: *The quarterly journal of economics*, pp. 643–669.

Emmer, Susanne, Claudia Klüppelberg, and Ralf Korn (2001). "Optimal portfolios with bounded capital at risk". In: *Mathematical Finance* 11.4, pp. 365–384.

Epstein, Larry G and Martin Schneider (2003). "Recursive multiple-priors". In: *Journal of Economic Theory* 113.1, pp. 1–31.

Ferguson, Thomas S. (1989). "Who Solved the Secretary Problem?" In: *Statistical Science* 4.3, pp. 282–289.

Föllmer, Hans and Alexander Schied (2011). *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

Gambrah, Priscilla Serwaa Nkyira and Traian Adrian Pirvu (2014). "Risk measures and portfolio optimization". In: *Journal of Risk and Financial Management* 7.3, pp. 113–129.

Gilbert, John P. and Frederick Mosteller (1966). "Recognizing the Maximum of a Sequence". In: *Journal of the American Statistical Association* 61.313, pp. 35–73.

Gilboa, Itzhak and Massimo Marinacci (2016). "Ambiguity and the Bayesian paradigm". In: *Readings in Formal Epistemology*. Springer, pp. 385–439.

Gilboa, Itzhak and David Schmeidler (1989). "Maxmin expected utility with non-unique prior". In: *Journal of mathematical economics* 18.2, pp. 141–153.

Hald, Anders (2003). *A history of probability and statistics and their applications before 1750*. Vol. 501. John Wiley & Sons.

Knight, Frank H (1921). "Risk, Uncertainty and Profit". In:

Merton, Robert C (1969). "Lifetime portfolio selection under uncertainty: The continuous-time case". In: *The review of Economics and Statistics*, pp. 247–257.

Merton, Robert C (1973). "Theory of rational option pricing". In: *The Bell Journal of economics and management science*, pp. 141–183.

Peskir, Goran and Albert Shiryaev (2006). *Optimal stopping and free-boundary problems*. Springer.

Riedel, Frank (2009). "Optimal stopping with multiple priors". In: *Econometrica* 77.3, pp. 857–908.

Rogers, Leonard CG (2013). *Optimal investment*. Springer.

Samuelson, Paul A (1965). "Rational theory of warrant pricing". In: *IMR; Industrial Management Review (pre-1986)* 6.2, p. 13.

Savage, Leonard J (1954). *The foundations of statistics*. John Wiley, New York.

Snell, James Laurie (1952). "Applications of martingale system theorems". In: *Transactions of the American Mathematical Society* 73.2, pp. 293–312.

Sollis, Robert (2009). "Value at risk: a critical overview". In: *Journal of Financial Regulation and Compliance* 17.4, pp. 398–414.

Von Neumann, John and Oskar Morgenstern (1953). "Theory of games and economic behavior". In:

Wald, Abraham (1945). "Sequential tests of statistical hypotheses". In: *The Annals of Mathematical Statistics* 16.2, pp. 117–186.

Chapter 2

Robust Maximum Detection: Full Information Best Choice Problem under Multiple Priors

Abstract

We consider a robust version of the *full information best choice* problem (Gilbert and Mosteller (1966)): there is ambiguity (represented by a set of priors) about the measure driving the observed process. We solve the problem under a very general class of multiple priors in the setting of Riedel (2009). As in the classical case, it is optimal to stop if the current observation is a running maximum that exceeds certain thresholds. We characterize the decreasing sequence of thresholds, as well as the (history dependent) minimizing measure. We introduce *locally constant ambiguity neighborhood* (LCAn) which has connections to coherent risk measures. Sensitivity analysis is performed using LCAn and *exponential neighborhood* from Riedel (2009).

2.1 Introduction

How is one to make the best choice among sequentially presented options when no recall is possible? Many scenarios in economics can be reduced to this question. In the well known *secretary problem* the employer is trying to pick the highest ranked among sequentially presented candidates for a position (Ferguson (1989)). In *job search* models, the unemployed agent is choosing among job offers trying to maximize life wealth (Lippman and McCall (1976)). In *house selling problems* the realtor is maximizing the profit in a series of take-it-or-leave-it bids (Porteus (2002)). These admittedly stylized problems are not trivial, and as such represent a useful first step towards more complex models and applications of the theory of optimal stopping (Ferguson (2006), Peskir and Shiryaev (2006)).

We consider the following best choice problem: a venture capitalist (the agent) is looking to invest and her budget allows her to invest in only one of the several sequentially presented start-up companies. She assumes the start-ups are similar and evaluates them by calculating a certain score. Due to the similarity of the start-ups and her familiarity with the matter, she treats the scores as realizations of independent and identically distributed random variables, the distribution of which is known to her. She believes that, given the high competition and failure rate among start-ups, only the company with the highest overall evaluation is the one that will be profitable. There is no recall: the decision not to invest in a start-up cannot be reversed. Hence, she is interested in maximizing the probability of choosing the start-up company with the highest valuation[1] .

This is one of the ways to formulate *full information best choice* (FIBC) problem, one of the best known optimal stopping problems in discrete time (Gilbert and Mosteller (1966), Bojdecki (1978), Samuels (1982), Ferguson (1989)). Formally, the agent is interested in detecting the maximum of a finite sequence of i.i.d. random variables (X_t), i.e. identifying the stopping time τ that maximizes the probability $P(X_\tau = \max(X_1, X_2, \ldots, X_T))$. The

[1]This formulation is based on a related problem in Bruss, Ferguson, et al. (2002).

solution is elegant: it is optimal to stop only if the current observation is also the current maximum that exceeds some threshold value. The thresholds are decreasing, can be calculated in advance and depend only on the number of remaining observations (Gilbert and Mosteller (1966), Bojdecki (1978)).

The "full information" in the name of the FIBC problem refers to the fact that the agent knows the distribution of start-ups' scores. The reasons to question this strong assumption are numerous. There is no objective way to be certain that the distribution the agent uses is the correct one. Considering a set of measures "around the assumed probability" would make the solution more robust. Even if one adopts the subjective probability approach, a single prior is not a reasonable assumption as shown by the Ellsberg paradox (Ellsberg (1961)). Indeed, even rational agents allow for Knightian uncertainty, or ambiguity, and behave in a way that is ambiguity-averse. A well established model of ambiguity aversion is *maxmin expected utility* theory, formulated by Gilboa and Schmeidler (1989). It assumes that the agent considers a set of priors and behaves pessimistically in a certain sense: when choosing the optimal action the agent first considers optimal actions over all of the priors and then chooses the one which has the lowest expected payoff[2].

In this paper we formulate and solve the FIBC problem under multiple priors in the setting of Riedel (2009). We show that the optimal stopping time is of the same form as in the classical case: it is completely characterized by a decreasing sequence of thresholds. We also characterize the measure under which the single prior problem is equivalent to the multiple priors one; it is highly history dependent.

The theory of optimal stopping under multiple priors in discrete time is developed in Riedel (2009). It shows that each adapted optimal stopping problem under multiple priors has a minimizing measure that reduces the problem to a single prior optimal stopping problem. One of the conditions that a set of priors has to fulfill in order to be used in an optimal stop-

[2]Extensions and applications are numerous; for a recent review of ambiguity aversion theory see Gilboa and Marinacci (2016).

ping problem under multiple priors is *time consistency*. It can be viewed as a mechanism that ensures that backward induction procedure gives the same optimal behavior as ex-ante optimization along all possible paths, thus avoiding dynamic inconsistencies.

Among the few optimal stopping problems completely solved in the multiple priors setting is the *secretary problem* (Chudjakow and Riedel (2013)), a better known, yet simpler predecessor to the FIBC problem. After identifying the minimizing measure, the authors proceed to solve the single prior optimal stopping problem. As will be seen, due to the complexity of the FIBC problem, this approach is not viable in our case. Indeed, several advances/generalizations in the theory of optimal stopping under multiple priors were necessary for the problem to be solved in some generality: construction of the set of priors, identifying certain extremal measures and adaptation of non-adapted problems among others.

Time-consistency is explicitly taken into account in the construction of the set of priors that we propose. We start with a set of measures that contains a uniform distribution. It can be thought of as "marginal ambiguity" due to the fact that it describes uncertainty about uniform measure in each period. Using certain predictable processes, we paste these single period measures using a dynamic product of Radon-Nykodim derivatives to obtain a set of priors for the whole process. Random variables X_t are not independent nor identically distributed under each measure in the set of priors. However, as marginal ambiguity remains constant, they can be considered as having *identical and independent ambiguity* in each period.

Identifying the minimizing measure in optimal stopping problems under multiple priors is not always easy. We use ideas from first order stochastic dominance[3] to identify certain *extremal measures* within the set of priors. Extremal measures can facilitate solving the problem and characterizing the minimizing measure, as will be the case in the solution of the FIBC problem.

We note that, to the best of our knowledge, all of the so far solved prob-

[3]See Levy (2015).

lems of optimal stopping under multiple priors use sets of priors that can be considered special cases of our construction (in particular the *exponential neighborhood* in Riedel (2009) and the set of multiple priors in Chudjakow and Riedel (2013)). Indeed, our construction is quite general and allows for complex sets of priors that cannot be parametrized by a real parameter, or even countably many of them. Furthermore, the extremal measures we introduce play an important role in all of the already available solutions.

The probability of stopping at the maximum value in the FIBC problem depends on future observations. This means that the problem is not adapted, hence the theory of optimal stopping under multiple priors cannot be applied directly. We formulate an equivalent and adapted version of the problem by conditioning on currently available information and minimizing over all priors. This was already done in Chudjakow and Riedel (2013) and we show that the same approach works with our more general construction, while offering additional details that allow the procedure to be potentially used in other applications.

Even with a deep understanding of the classical FIBC problem it is not immediately clear how multiple priors affect the solution. If one thinks in terms of minimizing measures the "opposing effects of ambiguity" in FIBC appear: if the agent stops, the worst that can happen are high outcomes, and if she continues, low outcomes would be the worst. This makes identifying the minimizing measure difficult. We initially avoid it altogether by finding suitable representations for values of stopping and continuing. Somewhat surprisingly, the representations are just monotone functions of a single variable. Naturally, it is optimal for the agent to stop once the value of stopping exceeds the value of continuing; this leads to the decreasing thresholds, as in the classical case.

Once we have the solution, we are able to identify the minimizing measure. It is history dependent: the agent's observations and actions up to a certain moment influence what she perceives to be the worst probability measure from that moment on. In particular, under the minimizing measures variables

X_t are not independent. This has a technical consequence that the FIBC problem under the minimizing measure is not equivalent to a single prior version of the FIBC problem[4].

Our theoretical results are in accordance with experimental studies of the FIBC problem which is, due to its simple formulation, suitable for behavioral research. The oldest study on the subject (Kahan et al. (1967)) shows that agents do not recognize the underlying probability distribution well; this may be another reason to consider multiple priors in models of human behavior. Corbin et al. (1975) shows that agents do not consider the observations as independent, even when they are informed that they actually are. More recently, Lee (2006) demonstrates that the observed behavior of participants in the study is best described by threshold rules. Overall, this is a positive indication that optimal stopping under multiple priors can be a viable model for real human behavior in optimal stopping problems.

We also establish a connection with the theory of coherent measures of risk, which at its core also has multiple priors (Artzner et al. (1999) , Föllmer and Schied (2011)). One could interpret the behavior of the FIBC agent operating in the multiple prior setting as follows: she considers her investment opportunity as a financial position, and chooses behavior that is optimal with respect to a certain risk measure. We introduce a locally constant ambiguity neighborhood (LCAn) that we use as the set of marginal ambiguity. This way we, effectively, describe marginal ambiguity by a risk measure wich turns out to have connections with the well known *Average Value at Risk*[5].

We finally investigate the way ambiguity affects the optimal behavior by considering two examples: exponential neighborhood introduced in Riedel (2009) and LCAn. Both examples can be considered robust neighborhoods around the initial measure. By deriving the explicit equations for the values of the optimal thresholds we are able to peform numerical calculations in both cases. Calculations offer two interesting conclusions. First, for "small sets of priors" the threshold values converge to the classical FIBC solution,

[4]See discussion after the formulation of the theorem 2.3.1 below.

[5]See ch. 4 in Föllmer and Schied (2011).

establishing its robustness. Second, in both settings, the ambiguity averse agent stops earlier. This is different than the conclusions of the similar analysis on the secretary problem Chudjakow and Riedel (2013), where the agent could stop both earlier and later, depending on parameters that describe the set of priors.

We revisit the classical FIBC problem in section 2. In section 3 we first present the generalized way of constructing the set of priors for problems of optimal stopping and identify extremal measures within it. We then formulate and solve the FIBC problem under multiple priors. Examples with numerical calculations can be found in section 4, where we introduce the LCAn neighborhood and identify its extremal measures. Proofs and valuable additional material are in the appendix.

2.2 The Original FIBC Problem

For the sake of completeness we formulate the classical FIBC problem and briefly revisit its solution.

At each period $t \in \{1, 2, \ldots, T\}$ the agent observes the process X_t which consists of random variables independently and, without loss of generality[6], uniformly distributed on the interval $[0, 1]$. Let $(\Omega, (\mathcal{F}_t)_{0 \leq t \leq T}, P_0)$ be a filtered probability space with $\Omega = [0, 1]^T$ being the product space, $\mathcal{F}_t$ being the σ-algebra generated by random variables X_1, X_2,...,X_t, and P_0 being the product of the (given) uniform marginal measures. We denote the set of all stopping times with $\mathcal{T}$ and the running maximum of the process with $M_t = \max(X_1, X_2, \ldots, X_t)$.

The agent is interested in detecting the maximum: finding the optimal stopping time τ that maximizes the probability $P_0(X_\tau = M_T)$ of stopping at X_t with the highest valued realization. If we define the reward process $Y_t = P(X_t = M_T)$ we can formulate the FIBC problem as the following

[6]Indeed, if the distribution $F = F_{X_t}$ was not uniform, a simple transformation would suffice: $X'_t = F^{-1}(X_t)$.

optimal stopping problem.

Problem 1 (FIBC problem – non-adapted version)**.** Find $\tau^* \in \mathcal{T}$ such that:

$$E[Y_{\tau*}] = \max_{\tau \in \mathcal{T}} E[Y_\tau] = \max_{\tau \in \mathcal{T}} P_0(X_\tau = M_n).$$

The process Y_t is not adapted to the filtration $\mathcal{F}_t$, so the theory of optimal stopping cannot be applied. We define the adapted payoff process $\widehat{Y}_t = E[Y_t \mid \mathcal{F}_t]$ and the problem can be equivalently formulated as:

Problem 2 (FIBC problem – adapted version)**.** Find $\tau^* \in \mathcal{T}$ such that

$$E[\widehat{Y}_{\tau*}] = \max_{\tau \in \mathcal{T}} E[\widehat{Y}_\tau].$$

The problems are equivalent in the sense that the same stopping time solves both problems. Indeed, using the law of iterated expectation, one can easily prove that $E[\widehat{Y}_\tau] = E[Y_\tau]$ holds for any $\tau \in \mathcal{T}$[7].

In the classical FIBC problem it is optimal to stop if the current value X_t is a candidate (i.e. $X_t = M_t$) that exceeds a certain threshold a_t that decreases with time: the less time remains the lower valued candidate the agent is willing to accept. More precisely the optimal stopping time is given with $\tau* = \min\{t \mid X_t = M_t \geq a_{T-t}\}$ where the numbers a_n satisfy the equations:

$$a_0 = 0; \quad \sum_{j=1}^{n} \frac{1}{j a_n^j} = 1 + \sum_{j=1}^{n} \frac{1}{j}, \; n \in \mathbb{N}.$$

For details see equation 1.2 in Samuels (1982) and original papers Gilbert and Mosteller (1966) and Bojdecki (1978).

[7]It is worth pointing out that, although this equivalence is straightforward in the single prior case, the process of adaptation of payoff will be somewhat more complex in the multiple prior setting, as will be seen in the lemma 2.C.2 below.

2.3 FIBC Problem under Multiple Priors

In the classical formulation of the FIBC problem above (Problems 1 and 2) the agent was maximizing the probability of stopping at the highest value of the process. The probability measure used for calculating the expectation was the one given prior P_0. In a multiple prior setting the agent performs all her calculations over a set of priors and then makes the most cautious/pessimistic decision.

Before the FIBC problem under multiple priors can be formulated, some technical preparation is needed when it comes to the set of priors. We do so in the first subsection, while the problem's formulation and solution are left for the second subsection.

2.3.1 The Set of Multiple Priors

We present a relatively general construction of the set of priors for problems of optimal stopping under multiple priors. The idea is to first introduce, for each period, the "marginal set of priors", and then to paste those in a time consistent manner to form the set of priors for the whole proces (X_t).

Let $(S, \mathcal{S}, v_0)$, $S \subset \mathbb{R}$, be a given probability space. Without loss of generality we assume that v_0 is strictly positive; this is clearly the case when v_0 is uniform. We furthermore assume that v_0 has a positive and bounded density. We define a set $\Omega = S^T$, for $T \in \mathbb{N}$, a sigma field $\mathcal{F} = \otimes_{t=0}^{T} \mathcal{S}$ (generated by projections $X_t : \Omega \to S$) and a probability measure $P_0 = \otimes_{t=1}^{T} v_0$ (under which the projections X_t are i.i.d[8]).

Let

$$\mathcal{V}^A = \{v_\alpha : \mathcal{F} \to (0,1) \,|\, \alpha \in A\}$$

be a set of probability measures on S indexed by some fixed set A. Since sets of multiple priors are used to model ambiguity one can, analogously,

[8]Although we use i.i.d. random variables the arguments that follow can readily be adjusted to the case when random variables are not identically distributed (but are still independent).

think of the set $\mathcal{V}^A$ as the set of *marginal ambiguity*. We note that marginal ambiguity will remain constant, i.e. the set $\mathcal{V}^A$ is fixed and does not change with the flow of time[9].

The set of priors on Ω is obtained by pasting together measures from $\mathcal{V}^A$. In order to do the pasting in a time-consistent manner we define a set $\mathcal{A}$ of all predictable processes with values in A:

$$\mathcal{A} = \left\{ a = (a_t)_{t \leq T} \;\middle|\; \begin{array}{l} a_{t+1} = a_{t+1}(x_1, x_2, \ldots, x_t) \in A, \;\; x_s \in S, s \leq t < T; \\ \dfrac{dv_{a_1}}{dv_0} \cdot \dfrac{dv_{a_2}}{dv_0} \cdot \ldots \cdot \dfrac{dv_{a_t}}{dv_0} \in L^0(\Omega, \mathcal{F}_t, P_0|\mathcal{F}_t), \;\; t \leq T \end{array} \right\},$$

where L^0 is the set of all measurable functions. As can be seen, $\mathcal{A}$ contains all processes that are predictable in a sense that their value at time $t+1$ depends on past realizations of random variables. The second requirement in the definition is technical, but revealing: requiring a "dynamic product" of Radon-Nykodim derivatives to be measurable with respect to $\mathcal{F}_t$ allows us to assign a measure to each process in $\mathcal{A}$; we do so below.

For each $a \in \mathcal{A}$ we can *define* a probability measure P^a on $(\Omega, (\mathcal{F}_t))$ by defining its density process:

$$\left.\frac{dP^a}{dP_0}\right|_{\mathcal{F}_t} = \prod_{s=1}^{t} \frac{dv_{a_s}}{dv_0}, \tag{2.1}$$

and, finally, we set

$$\mathcal{P} = \mathcal{P}(\mathcal{V}^A) = \{P^a \,|\, a \in \mathcal{A}\}.$$

All measures in the set $\mathcal{P}$ are equivalent; this is due to the definition of the set $\mathcal{V}^A$ within which all the measures are equivalent. We note that random variables X_t are not independent under every measure $P \in \mathcal{P}$. In fact, the only measures under which they are independent are those that correspond to direct products, i.e. processes $a \in \mathcal{A}$ such that for each t the function a_t is constant.

[9] Again, careful reading of the arguments that follow shows that one does not lose on generality by fixing an identical set of beliefs at every step.

Although we use the set $\mathcal{P}$ as the set of priors to formulate and solve the FIBC problem under multiple priors, the generality of its construction allows for other applications: under mild assumptions on the set $\mathcal{V}^A$ the theory of optimal stopping under multiple priors from Riedel (2009) can be applied. We prove this result in Appendix 2.A, where we also offer some further details on optimal stopping under multiple priors that are relevant for our solution.

Extremal Measures In order to be able to solve and reduce a multiple prior problem to a single prior one, we define "extremal measures" within the set $\mathcal{P}$. The ideas we use are those of the theory of (first order) stochastic dominance (Levy (2015)).

Let us denote by $\underline{v} \in \mathcal{V}^A$ the measure (if it exists) such that:

$$\underline{v}(X_1 \leq x) \geq v_a(X_1 \leq x), \quad \text{for any } x \in \mathbb{R} \text{ and any } v_a \in \mathcal{V}^A. \tag{2.2}$$

As can be seen the minimizing measure $\underline{v} \in \mathcal{V}^A$ is the one that puts the most weight on the lowest valued outcomes.

This allows us to single out the measure $\underline{P} = \otimes_{t=0}^{n} \underline{v} \in \mathcal{P}$ (under which the variables X_t are independent!). Measures $\overline{v} \in \mathcal{V}^A$ and $\overline{P} = \otimes_{t=0}^{n} \overline{v}$ are defined analogously. Characterizations of extremal measures and additional useful results are offered in Appendix 2.B.

2.3.2 FIBC Problem under Multiple Priors

Let $\mathcal{P}$ be a set of multiple priors obtained by pasting together one-period multiple priors sets $\mathcal{V}^A$ indexed by some set A. We assume $\mathcal{V}^A$ satisfies all the conditions of lemma 2.A.1, with v_0 being the uniform distribution. We also assume that the set $\mathcal{V}^A$ contains both a measure $\overline{v}$ that satisfies the lower extremal property and a measure $\overline{v}$ that satisfies the upper extremal property.

Let X, Y, $\mathcal{F}$ be as in section 2.2. Note that while X_t are independent under the reference measure P_0, they are not independent under every $P \in \mathcal{P}$.

The agent is solving the problem:

Problem 3 (FIBC problem under multiple priors – non-adapted version)**.** Find $\tau^* \in \mathcal{T}$ such that

$$\min_{P\in\mathcal{P}} E^P[Y_{\tau^*}] = \max_{\tau\in\mathcal{T}} \min_{P\in\mathcal{P}} E^P[Y_\tau].$$

Since the duality equality

$$\max_{\tau\in\mathcal{T}} \min_{P\in\mathcal{P}} E^P[Y_\tau] = \min_{P\in\mathcal{P}} \max_{\tau\in\mathcal{T}} E^P[Y_\tau] \tag{2.3}$$

holds[10], the following interpretation is plausible: the agent maximizes the probability of stopping at the maximum of the given process under the "worst possible" measure in the set $\mathcal{P}$.

Similarly as in Problem 1, this problem needs to be reduced to an adapted problem; this will allow us to solve it using the theory of optimal stopping under multiple priors. For that purpose we define the adapted payoff process under multiple priors:

$$Z_t = \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[Y_t|\mathcal{F}_t].$$

Problem 4 (FIBC problem under multiple priors – adapted version)**.** Find $\tau^* \in \mathcal{T}$ such that

$$\inf_{P\in\mathcal{P}} E^P[Z_{\tau*}] = \sup_{\tau\in\mathcal{T}} \inf_{P\in\mathcal{P}} E^P[Z_\tau].$$

Problems 3 and 4 are equivalent – the same stopping time solves both problems. The equivalence in the multiple priors setting is less clear than it was in the single prior setting; we prove it in Appendix 2.C. The proof does not depend on the definition of the payoff process, hence it holds for any non-adapted process under multiple priors.

The following theorem completely characterizes the solution of the FIBC problem under multiple priors.

[10] See theorem 2 in Riedel (2009).

Theorem 2.3.1. *1. There is a decreasing sequence of thresholds $(b_t)_{t=1,\dots T}$ such that the optimal stopping time τ^* that solves the BCIF problem under multiple priors is:*

$$\tau^* = \min\{t \,|\, X_t = \max(X_1, X_2, \dots, X_t) > b_t\}. \tag{2.4}$$

2. Thresholds b_t are the unique solutions of equations $w_t(x) = \overline{r}_t(x)$, $t < T$, where functions $\overline{r}_t$ and w_t are defined recursively:

$$\overline{r}_T(m) \equiv 1, \quad \overline{r}_t(m) = E^{\overline{P}}\left[\prod_{s>t} \mathbb{1}_{X_s \le m}\right]; \quad w_T(m) \equiv 0,$$

$$w_t(m) = \operatorname*{ess\,inf}_{v \in \mathcal{V}^A} \Big(\int_m^1 \max(\overline{r}_{t+1}(x), w_{t+1}(x))\, dv(x) + w_{t+1}(m) \int_0^m dv(x) \Big).$$

Specially, $b_T = 0$.

3. The minimizing measure $P^* = P^{a^*}$ is given by the predictable process

$$a_t^*(x_1, \dots, x_t) = \begin{cases} a_t^c(x_1, \dots, x_t), & t < \tau^* \\ \overline{\alpha}, & t \ge \tau^* \end{cases}, \tag{2.5}$$

where $v_{\overline{\alpha}} = \overline{v}$ and

$$a_t^c(x_1, \dots, x_t) = \arg\min_{a \in A} \Big(\int_{\max(x_1,\dots,x_t)}^1 \max(\overline{r}_{t+1}(x), w_{t+1}(x))\, dv_a(x) + w_{t+1}(\max(x_1, \dots, x_t)) \int_0^{\max(x_1,\dots,x_t)} dv_a(x) \Big). \tag{2.6}$$

The duality equation (2.3) holds for the process Z_t, too. A reasonable attempt at solving Problem 4 would be to identify the minimizing measure and solve the classical, single prior optimal stopping problem under the minimizing measure. One could even hope that one of the measures $\underline{P}$ or $\overline{P}$ would turn out to be the minimizing measure, thus allowing the problem to

be reduced to the classical FIBC problems 1 and 2. The theorem shows that the minimizing measure is significantly more complicated than that. This is due to the fact that the multiple priors setting creates opposing effects about what is "pessimistic": when the agent stops the worst measures are those that put the most weight on high outcomes, and when she continues the worst measures are those that put the most weight on the lowest outcomes, while accounting for future behavior.

As can be seen, the minimizing measure P^* is highly history dependent and even depends on the act of stopping. This implies that random variables X_t are not independent under P^*. Hence, an agent operating in a multiple prior setting views the FIBC problem in a way that is substantially different from that of an agent making decisions under the single prior. This is true on a technical level, too: the reduction to the classical FIBC problem via the probability integral transform (as indicated on pp.51-52 in Gilbert and Mosteller (1966)) is not possible.

The proof ultimately relies on several careful backward inductions and can be considered a multiple priors version of the proof offered in Samuels (1982). The details are available in the appendix. Although tedious, the proof offers significant insight into the FIBC problem.

The proof reveals that, if at time t the agent observes value x_t that is a running maximum, then the expected value (under multiple priors) of continuing is $w_t(x_t)$ while the expected value of stopping is $\overline{r}(x_t)$. Hence, the stopping rule prescribed by τ^* merely says that the agent stops if payoff of stopping exceeds the payoff of continuing; this is in accordance with the theory of optimal stopping (under multiple priors). Furthermore, the multiple prior Snell envelope of the adapted version of the FIBC problem under multiple priors can be expressed in terms of functions $\overline{r}_t$ and w_t:

$$U_t = \max(\overline{r}_t(X_t)\mathbb{1}_{X_t=M_t}, w_t(M_t)).$$

The (classical) FIBC problem is "end-invariant" in the following sense: "optimum decision numbers depend only upon the number of remaining

draws", as noted in Gilbert and Mosteller (1966). The proof reveals that, once the set $\mathcal{P}$ is fixed, the same is true for the FIBC problem under multiple priors. Indeed, its solution was derived by backward induction and was shown to depend only on the values of the current observation and the running maximum. Naturally, the cutoff points b_t depend on the on the set of priors $\mathcal{P}$; we explore this dependency numerically in the next section.

2.4 Examples

What is the effect of introducing multiple priors to the FIBC problem? How does the optimal stopping time change once ambiguity is introduced? We try to give some answers to these and related questions in this section.

The sequence of cutoff points that define the optimal stopping time in the classical version of the problem has been well studied (already in Gilbert and Mosteller (1966)) and their asymptotic behavior is well understood (Samuels (1982)). However, due to the complexity of the minimizing measure and recursive equations in 2.3.1, that kind of analysis is not trivial in our setting. We focus our attention on the simple case when $T = 3$; it will be seen below that even this case is computationally cumbersome and leads to highly nonlinear equations. Given the comments about the end invariance of the FIBC problem in the previous section, what follows is effectively an analysis of the final three periods of any FIBC problem with the horizon $T \geq 3$; the notations we use reflect this fact.

2.4.1 Classical Case

For the sake of completeness we briefly review the numerical values of the optimal stopping time in the classical FIBC problem. In our context, it corresponds to the case when $\mathcal{V}^{\mathcal{A}}$ is a singleton with its only element being the uniform measure. Omitting the straightforward calculations we present the interesting parts of the result.

Functions $\overline{r}_t$ and w_t take the following form:

$$\overline{r}_{T-1}(x) = x, \qquad \overline{r}_{T-1}(x_1, \ldots, x_{T-2}) = x^2$$
$$w_{T-1}(m) = 1 - m, \quad w_{T-2}(m) = \int_m^1 \max(x, 1-x)\, dx + m(1-m).$$

The cutoff points that define the stopping time are

$$b_T = 0, \quad b_{T-1} = \frac{1}{2}, \quad b_{T-2} = \frac{2+\sqrt{24}}{10} \approx 0.6899.$$

Note that b_{T-1} and b_{T-2} are the solutions of the equations $m = w_{T-1}(m)$ and $m^2 = w_{T-2}(m)$ respectively.

2.4.2 Exponential Neighborhood

The *exponential neighborhood* is the set of priors $\mathcal{P}^{\alpha,\beta}$ introduced in Riedel (2009). It has important connections to Girsanov theory and arises naturally in statistics where it is referred to as *exponential family*. It has been used to model uncertainty in optimal stopping problems related to finance (option pricing), as well as ambiguous versions of the *house pricing problem* and the *parking problem*[11].

In the context of this paper it can be introduced by setting $A = [\alpha, \beta]$ and:

$$\mathcal{V}^A_{EXP} = \left\{ v_a \,\middle|\, \frac{dv_a}{dv_0}(x) = \frac{e^{ax}}{\int_0^1 e^{at}\, dt},\ a \in A \right\},$$

where v_0 is the uniform measure on the interval $[0, 1]$. It is known[12] that $\underline{v} = v_\alpha$ and $\overline{v} = v_\beta$. The exponential neighborhood is simply: $\mathcal{P}^{\alpha,\beta} = \mathcal{P}(\mathcal{V}^A_{EXP})$.

We have already seen that $\overline{r}_T(x) = 1$ and $w_T(x) = 0$ for any $x \in [0, 1]$. Direct calculations yield:

$$\overline{r}_{T-1}(x) = \frac{e^{\beta x} - 1}{e^\beta - 1}, \qquad w_{T-1}(x) = \frac{e^\beta - e^{\beta x}}{e^\beta - 1}.$$

[11] See section 4 in Riedel (2009).
[12] For details see section 4 in Riedel (2009).

α	β	b_{T-1}	b_{T-2}	α	β	b_{T-1}	b_{T-2}
0	0	0.5000	0.6899	-2	2	0.7169	0.8084
-0.01	0.01	0.5013	0.6905	-5	5	0.8627	0.9048
-0.1	0.1	0.5125	0.6958	-10	10	0.9307	0.9518
-0.25	0.25	0.5312	0.7047	-1	2	0.7169	0.8144
-0.5	0.5	0.5619	0.7200	-1	3	0.7851	0.8582
-1	1	0.6201	0.7512	-2	1	0.6201	0.7406
-1.5	1.5	0.6722	0.7811	-3	1	0.6201	0.7335

Table 2.1: Exponential neighborhood – Values of the cutoff points b_{T-1} and b_{T-2} for different values of α and β.

Equating $\overline{r}_{T-1}(x) = w_{T-1}(x)$ we obtain

$$b_{T-1} = \frac{1}{\beta} \ln \frac{e^{\beta} + 1}{2}.$$

The expression for w_{T-2} is more cumbersome:

$$w_{T-2}(m) = \frac{1}{e^{\beta} - 1} \min_{a \in [\alpha, \beta]} \left(\frac{1}{\int_0^1 e^{at}\, dt} \left(a \int_m^1 e^{ax} \max(e^{\beta x} - 1, e^{\beta} - e^{\beta x})\, dx \right. \right.$$
$$\left. \left. + (e^{\beta} - e^{\beta m})(e^{am} - 1) \right) \right),$$

while $\overline{r}_{T-2} = (\overline{r}_{T-1})^2$.

As can be seen, explicitly calculating the cutoff point b_{T-2} (i.e. solving the equation $w_{T-2}(m) = r_{T-2}(m)$) is not computationally easy and to obtain the approximations of its value one can resort to mathematical software[13]. Table 1 provides approximations of values of cutoff points b_{T-1} and b_{T-2} for different values of parameters α and β. If the difference $\beta - \alpha$ is interpreted as the "amount of ambiguity" one can notice that the increase in ambiguity causes later stopping. The last four rows seem to imply greater sensitivity of cutoff point values to the change in β, than in α. This is somewhat expected

[13] All the graphs and data for the tables were made using Wolfram Matematica, Research (2015).

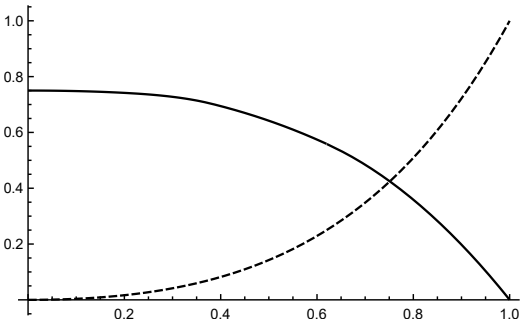

Figure 2.1: Exponential Neighborhood - Graphs of functions $w_{t-2}(x)$ (decreasing) and $\overline{r}_{T-2}(x)$ (increasing) for $\alpha = -1$ and $\beta = 1$. The point of intersection is b_{T-2}.

given the shape of the exponential function.

2.4.3 Local Constant Ambiguity Neighborhood

The *locally constant ambiguity neighborhood* (LCAn) is the set whose "marginal ambiguity" is:

$$\mathcal{V}^A_{CLA} = \left\{ v_a \,\middle|\, \frac{1}{\lambda} \leq \frac{dv_a}{dv_0} \leq \lambda \right\}.$$

The constant $\lambda \geq 1$ describes the "amount of ambiguity" – greater values of λ imply greater ambiguity about the "correct measure" that drives the payoff process. Case $\lambda = 1$ corresponds to the case where there is no ambiguity and the set $\mathcal{V}^A_{CLA}$ reduces to a singleton containing v_0. The LCAn is simply: $\mathcal{Q}^\lambda_{CLA} = \mathcal{P}(\mathcal{V}^A_{CLA})$. As can be seen the set $\mathcal{V}^A$ cannot be parametrized by a real parameter, nor even countably many real parameters. In that sense, it differs substantially from the exponential neighborhood, or any other analogously created neighborhood that depends on a fixed family of distributions.

One can interpret the marginal ambiguity of LCAn as follows: the agent is certain about which events are possible/impossible (described by measure v_0), but she allows for the possibility that for any sufficiently "small event" it's probability is up to λ-times overestimated or underestimated by v_0.

LCAn bears some resemblance to the well known ε-contamination from Huber (1981), which was already used in the context of ambiguty in the well

known paper Maccheroni et al. (2006). In our context, ε-contamination could be described as the set of measures, the range of densities of which lies within the interval $[1-\varepsilon, 1+\varepsilon]$. Arguably, this is a less natural model of ambiguity than LCAn when it comes to describing belief by a set of priors. Beyond the obvious fact that ε cannot be greater than one (which discounts for the possibility of any event being more than twice underestimated) there seem to be indications that humans innately think logarithmically, rather than linearly (Dehaene (2003)). In particular, to put it in more plastic terms, this may mean that it is more natural to think of [1/2,2] as a neighborhood around the point 1 than [0.5,1.5]; this corresponds to the way in which the ambiguity around "small events" is modeled by LCAn.

We note that the set V^{λ}_{CLA} is related to certain sets that appear in the theory of risk measures. In particular, the well known risk measure known as *average value at risk* can be characterized by a similarly defined set (chapter 4 in Föllmer and Schied (2011)). It is well known that there are mathematical connections between risk measures and the theory of multiple priors. It is also well known that ambiguity (in the sense of multiple priors) could be viewed as a way to describe model uncertainty. The same is true for risk measures and model uncertainty in finance.

Due to the similarities between the set that characterizes AVaR and LCAn, one could argue that LCAn introduces robustness to the dynamic of the FIBC problem in a way that is closer to robustness in finance. Indeed, at each moment t the agent evaluates the values of all her possible present and future actions, then chooses the least risky one with respect to the risk measure induced by the set LCAn. Arguably, this makes the LCAn an attractive option for future (dynamic) models in economics and finance where uncertainty needs to be introduced.

With the set $\mathcal{Q}^{\lambda}_{CLA}$ defined, we can turn to the question of the existence of extremal measures within it. We answer this question in our context, i.e. with the reference measure v_0 being the uniform measure on the interval $[0,1]$, and we do so by focusing on the monotone function characterization

of extremal measures (see equation (2.9)). It would seem plausible, that measures $\underline{v}$ and $\overline{v}$ are the ones that put the most weight on the right and, respectively, left end of the interval $[0,1]$; we prove this result in lemma 2.4.1 below. We actually prove that the densities of extremal measures are given with:

$$\frac{d\underline{v}}{dv_0} = \frac{1}{\lambda}\mathbb{1}_{[0,\frac{\lambda}{\lambda+1}]} + \lambda\mathbb{1}_{[\frac{\lambda}{\lambda+1},1]} =: \underline{\varphi}, \qquad \frac{d\overline{v}}{dv_0} = \frac{1}{\lambda}\mathbb{1}_{[0,\frac{1}{\lambda+1}]} + \lambda\mathbb{1}_{[\frac{1}{\lambda+1},1]} =: \overline{\varphi}.$$

By similar logic as above, the worst measure for U-shaped payoffs should be the one that puts the most weight on an interval that contains the minimum of the payoff function; this result related to extremal measures is included in the lemma. The formulation of the lemma requires us to define the following set of densities of measures in $\mathcal{V}^{\lambda}_{CLA}$:

$$\begin{aligned} D^{\lambda}_{CLA} &= \left\{ \frac{dv}{dv_0} \middle| v \in \mathcal{V}^{\lambda}_{CLA} \right\} \\ &= \left\{ \varphi : [0,1] \to \mathbb{R} \middle| \int_0^1 \varphi(x)\,dx = 1, \tfrac{1}{\lambda} \leq \varphi(x) \leq \lambda \right\}. \end{aligned}$$

Lemma 2.4.1. *1. For every increasing, bounded measurable function $g : [0,1] \to \mathbb{R}$ and every $\varphi \in D^{\lambda}_{CLA}$ the following inequality holds:*

$$\int_0^1 g(x)\varphi(x)\,dx \geq \int_0^1 g(x)\underline{\varphi}(x)\,dx.$$

2. For every function $h : [0,1] \to \mathbb{R}$ which is decreasing on $[0,k]$ and increasing on $[k,1]$ for some $k \in [0,1]$ and for every function $\varphi \in D^{\lambda}_{CLA}$ there exists a function $\psi \in D^{\lambda}_{U}$ such that:

$$\int_0^1 h(x)\varphi(x)\,dx \geq \int_0^1 h(x)\psi(x)\,dx,$$

where

$$D^{\lambda}_{U} = \left\{ \lambda\mathbb{1}_{[c,c+\frac{1}{\lambda+1}]} + \frac{1}{\lambda}\mathbb{1}_{[0,c]\cup[c+\frac{1}{\lambda+1},1]} \middle| c \in \left[0, \frac{\lambda}{\lambda+1}\right] \right\}. \tag{2.7}$$

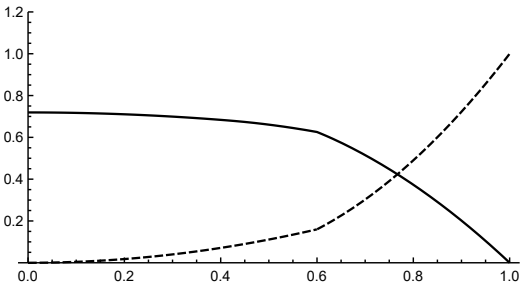

Figure 2.2: LCAn - Graphs of functions $w_{t-2}(x)$ (decreasing) and $\overline{r}_{T-2}(x)$ (increasing) for $\lambda = 3/2$. The point of interesection is b_{T-2}.

It can be seen that the set $D_U^\lambda \subset D_{CLA}^\lambda$ is the set of densities that put the most weight on the interval $\left[c, c + \frac{1}{\lambda+1}\right]$, which is in accordance with the considerations preceding the formulation of the lemma.

Analogous results can be formulated about decreasing functions and the inverted-U-shaped functions.

Definition of the function w_{T-2} and the monotonicity of functions w_{T-1} and r_{T-1} imply that the function w_{T-2} is U-shaped. Lemma 2.4.1 allows us to narrow down the search for the minimizing measure within the set D_U^λ, which in turn allows for mathematical software to be used to identify the minimizing measure, plot the graph of the function w_{T-2} (see figure 2.2) and calculate the value of the cutoff point b_{T-2}. Similarly as before, we provide a table with the approximate values of cutoff points b_{T-1} and b_{T-2} for different values of λ.

λ	b_{T-1}	b_{T-2}	λ	b_{T-1}	b_{T-2}
1	0.5	0.6899	2	0.7500	0.8182
1.01	0.5050	0.6916	3	0.8333	0.8754
1.1	0.5455	0.7073	4	0.8750	0.9057
1.25	0.6000	0.7318	8	0.9375	0.9524
1.5	0.6667	0.7671			

Table 2.2: LCAn – Values of the cutoff points b_{T-1} and b_{T-2} for different values of λ.

It is worth noting that in both examples we presented the agent stops

later than in the classical case. It is not hard to see that this is true for the period $T-1$ for any set of multiple priors, but is less obvious for periods $t < T-1$, hence it remains a conjecture. This is different from the results of Chudjakow and Riedel (2013) where it was found the agent could stop both earlier and later than the agent not facing ambiguity, depending on the shape of the set of multiple priors.

2.5 Conclusion

We formulated and solved the multiple priors version of the classical *full information best choice* problem under rather general conditions. We showed that the solution can be fully characterized via a set of decreasing thresholds, just as in the classical case. Instead of identifying the minimizing measure and then solving the single prior problem, we solve the problem with a more direct approach using the theory of optimal stopping under multiple priors.

More generally, we have demonstrated that the theory of optimal stopping under multiple priors can accommodate complex problems, hopefully paving the way for even harder problems to come. In this context, of interest is our result about adapting any non-adapted optimal stopping problem under multiple priors.

Our results fit into a wider setting of dynamic problems under multiple priors: we described a construction of a set of priors for the whole process using only a single-period set of priors. The construction ensures that the resulting set of priors is time consistent, thus allowing for "variables with independent and identical ambiguity" to be used practically and in some generality to model uncertainty in multi-period models, even beyond the theory of optimal stopping.

Although the theory of maxmin expected utility is a mature one, non-trivial examples of the sets of multiple priors in dynamic settings are rare. We introduced one such example using ideas from the theory of risk measures: *locally constant ambiguity neighborhood* is a set of priors in which ambiguity

of probability about th 'small' events remains constant. The set itself has promising interpretations in terms of model uncertainty and invites future research in the context of risk measures. It also opens possiblities in the other direction – to explicitly use sets of priors related to established risk measures in the context of dynamic economic problems under ambiguity.

As it is becoming increasingly evident that economic models with a single probability measure are not capturing the reality in a satisfactory way, it becomes necessary to investigate robust models that manage to take into account Knightian uncertainty of economic problems; we hope this paper convincingly presents one such model.

Appendix 2.A Applicability of the Theory of Optimal Stopping under Multiple Priors

For the theory of optimal stopping to be applied to processes with bounded payoffs the set of priors $\mathcal{P}$ has to satisfy three assumptions. It should be L^1 weakly closed and all the measures within the set $\mathcal{P}$ should be equivalent. The set $\mathcal{P}$ should also be time consistent: for any two measures, the measure that allows the agent to "switch" between them at some (possibly random) time must also be in the set $\mathcal{P}$; see assumptions $A2-4$ in Riedel (2009). The following lemma shows that the set $\mathcal{P}$ satisfies those assumptions once we impose mild conditions on the set $\mathcal{V}^A$.

Lemma 2.A.1. *Assume the set $\mathcal{V}^A$ satisfies:*

1. $v_0 \in \mathcal{V}^A$
2. *All the densities $\dfrac{dv_a}{dv_0}$, $a \in A$, are strictly positive and bounded*
3. *The set $\mathcal{V}^A$ is weakly closed in $L^1(S, \mathcal{S}, v_0)$*

Then the set of measures $\mathcal{P}(\mathcal{V}^A)$ satisfies assumptions A2, A3 and A4 in Riedel (2009).

Proof. The assumption A2 is satisfied because all the densities in $\mathcal{V}^A$ are strictly positive and bounded.

For the weak compactness it is sufficient to show that the set $\mathcal{P}$ is closed and bounded by a uniformly integrable random variable. Since all the densities are bounded, the latter is obvious. Closedness is a consequence of the third assumption in the formulation of the lemma: weakly closed sets are also strongly closed, thus the closedness is inherited from weak closedness in each period by pasting. To see this, it suffices to recall that a sequence of positive functions convergent in L_1 has a subsequence that converges pointwise (almost everywhere). With this the closedness can be proven using the classical argument (that a limit of a sequence of elements of the set also belongs to the set) by exploiting the previous remarks.

It remains to prove the time consistency. Due to the predictability of each of the functions a_k this is straightforward: Let P^a and P^b be two measures with densities $\frac{dP^a}{dP_0}\Big|_{\mathcal{F}_t} = \prod_{s=1}^{t} \frac{dv_{a_s}}{dv_0}$ and $\frac{dP^b}{dP_0}\Big|_{\mathcal{F}_t} = \prod_{s=1}^{t} \frac{dv_{b_s}}{dv_0}$ respectably and let τ be a stopping time. Define $c_t = a_t$ when $t \leq \tau$ and $c_t = b_t$ when $t > \tau$. The resulting measure from the property A4 coincides with the measure P^c with density $\frac{dP^c}{dP_0}\Big|_{\mathcal{F}_t} = \prod_{s=1}^{t} \frac{dv_{c_s}}{dv_0}$ which obviously belongs to $\mathcal{P}$; this is exactly what was supposed to be proven[14]. □

The theory of optimal stopping under multiple priors guarantees the existence of the stopping time $\tau^* \in \mathcal{T}$ such that:

$$\max_{\tau \in \mathcal{T}} \min_{P \in \mathcal{P}} E^P[\mathcal{E}_\tau] = \min_{P \in \mathcal{P}} E^P[\mathcal{E}_{\tau^*}],$$

where $\mathcal{E}_t$ is a bounded payoff process adapted to the filtration $\mathcal{F}_t$. The min-

[14]This lemma could alternatively be proven by showing that the set $\mathcal{P}$ coincides with the time-consistent hull "around" the set of all direct product measures from $\mathcal{V}^A$; see pp. 868 in Riedel (2009), or Riedel (2004).

imal optimal stopping time τ^* is given with

$$\tau^* = \min\{t \geq 0 \mid U_t = \mathcal{E}_t\},$$

where U is the recurcively defined *multiple priors value process*:

$$U_T = \mathcal{E}_T,\ U_t = \max\left(\mathcal{E}_t, \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[U_{t+1} \mid \mathcal{F}_t]\right).$$

Furthermore, the theory guarantees the existence of the measure $Q^* \in \mathcal{P}$ such that the value process under multiple priors of the optimal stopping problem under multiple priors coincides with the value process of the (single-prior) optimal stopping problem of the process $\mathcal{E}_t$ under the measure Q^*; this allows the possibility of reducing the multiple priors problems to the classical ones. For further details see Theorems 1 and 2 in Riedel (2009).

Appendix 2.B Details on Extremal measures

It is easy to prove that the inequality:

$$\underline{P}(X_{t+1} \leq x|\mathcal{F}_t) \geq P(X_{t+1} \leq x|\mathcal{F}_t)) \tag{2.8}$$

holds for any $t > 0$, $x \in \mathbb{R}$ and $P \in \mathcal{P}$, and a characterization in terms of monotone functions is straightforward along the lines of the classical proofs of theorems on first order stochastic dominance (Levy (2015)). Specifically, the measure $\underline{P} \in \mathcal{P}$ satisfies the inequality

$$E^{\underline{P}}[h(X_{t+1}) \mid \mathcal{F}_t] \leq E^P[h(X_{t+1}) \mid \mathcal{F}_t] \tag{2.9}$$

for each $t > 0$, each $P \in \mathcal{P}$ and each bounded, increasing real function h.

We note an immediate consequence of the monotone characterization of the extremal measures (2.9):

Lemma 2.B.1. *For any function $g_t : S^t \to \mathbb{R}$ that is bounded, measurable*

and increasing in its last argument the following equality holds:

$$\operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[g(X_1,...,X_t,X_{t+1}) \mid \mathcal{F}_t] = E^{\underline{P}}[g(X_1,...,X_t,X_{t+1}) \mid \mathcal{F}_t]$$

Proof. Since the filtration $\mathcal{F}$ is generated by $X_1,\dots,X_t$ it suffices to show that, for an arbtirary history $X_1 = x_1$, $X_2 = x_2$, ..., $X_t = x_t$, the following inequality holds:

$$\begin{aligned} E^P[g(X_1,...,X_{t+1}) \mid X_1 = x_1,\dots,X_t = x_t] \\ \geq E^{\underline{P}}[g(X_1,...,X_{t+1}) \mid X_1 = x_1,\dots,X_t = x_t]. \end{aligned}$$

This, however, is true because of the monotone characterization of the extremal measures (2.9). Indeed, once we fixed the values of random variables X_1, X_2, ..., X_t, the function g_t can be interpreted as a function of a single variable X_{t+1} and the inequality follows directly from the inequality (2.9). □

An analogous result holds for the decreasing functions.

We can now prove a general result on optimal stopping under multiple priors that allows one to explicitly reduce certain multiple prior problems to classical ones.

Lemma 2.B.2. *Suppose the set of priors $\mathcal{P}$, obtained by pasting as above, satisfies the conditions of lemma 2.A.1 and contains the measure $\underline{P}$. Then, for a sequence of functions $f_t(\epsilon_1,\dots,\epsilon_t)$, $t = 1,\dots,T$, each of which is bounded, measurable, and increasing in every argument on S, the following equality holds:*

$$\max_{\tau\in\mathcal{T}} \min_{P\in\mathcal{P}} E^P[f_\tau(\epsilon_1,\dots,\epsilon_\tau)] = \max_{\tau\in\mathcal{T}} E^{\underline{P}}[f_\tau(\epsilon_1,\dots,\epsilon_\tau)]. \tag{2.10}$$

This result identifies the measure $\underline{P}$ as the minimizing measure in the optimal stopping problem (under multiple priors) with payoff at time t given by $\mathcal{E}_t = f_t(\epsilon_1,\dots,\epsilon_t)$. It can be considered a twofold[15] generalization of

[15]Twofold in the sense that it considers both a more general set of priors and a more

the Theorem 5 in Riedel (2009). Analogous result can be formulated for decreasing functions and the measure $\overline{P}$.

Lemma 2.B.2 is a result on optimal stopping under multiple priors interesting in its own right, but it won't be directly used in the solution of the FIBC problem under multiple priors, However, the structure of the proof is conceptually similar to the significantly more involved proof of the main result on the optimal stopping in the FIBC problem under multiple priors. Hence, the lemma above can be considered a preparatory/introductory result.

Proof of lemma 2.B.2 : The value process of the optimal stopping problem under multiple priors is defined recursively with

$$U_T = f_T(\epsilon_1, \epsilon_2, \ldots, \epsilon_T), \quad U_t = \max\left(f_t(\epsilon_1, \epsilon_2, \ldots, \epsilon_t), \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[U_{t+1} \mid \mathcal{F}_t]\right).$$

As was already mentioned in the section 3 the optimal stopping time is given with $\tau^* = \min\{t \geq 0 \mid U_t = f_t(\epsilon_1, \epsilon_2, \ldots, \epsilon_t)\}$. Thus, it is sufficient to prove that the value proces U coincides with the value process $\underline{U}$ of the classical (single prior) optimal stopping process under the measure $\underline{P} \in \mathcal{P}$. Indeed, since the latter is defined with

$$\underline{U}_T = f_T(\epsilon_1, \epsilon_2, \ldots, \epsilon_T); \; \underline{U}_t = \max\left(f_t(\epsilon_1, \epsilon_2, \ldots, \epsilon_t), E^{\underline{P}}[\underline{U}_{t+1} \mid \mathcal{F}_t]\right), \; t < T,$$

and it's optimal stopping time is given with[16]: $\underline{\tau}^* = \min\{t \mid \underline{U}_t = f_t(\epsilon_1, \ldots, \epsilon_t)\}$ it is clear that the optimal stopping times will coincide if the value processes do, too.

general class of monotone functions.

[16]The cornerstone result of the optimal stopping theory is that it is optimal to stop the first time the value process and payoff are equal; see Peskir and Shiryaev 2006 or Ferguson 2006.

We can define, for $t < T$, the sequence of functions u_t:

$$u_t(e_1, \ldots, e_t) = \max\left(f_t(e_1, \ldots, e_t), \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[U_{t+1} \mid \epsilon_1 = e_1, \ldots, \epsilon_t = e_t]\right), \tag{2.11}$$

with $u_T(e_1, \ldots, e_T) = f_T(e_1, \ldots, e_T)$. Similarly, we can define

$$\underline{u}_T(e_1, \ldots, e_T) = f_T(e_1, \ldots, e_T)$$

and

$$\begin{aligned}\underline{u}_t(e_1, \ldots, e_t) &= \max\left(f_t(e_1, \ldots, e_t), E^{\underline{P}}[\underline{U}_{t+1}] \mid \epsilon_1 = e_1, \ldots, \epsilon_t = e_t]\right) \\ &= \max\left(f_t(e_1, \ldots, e_t), E^{\underline{P}}[\underline{U}_{t+1}]\right), \end{aligned} \tag{2.12}$$

where the second equality holds due to independence of variables ϵ_t under the measure $\underline{P}$.

As in lemma 2.B.1 we note that the filtration $\mathcal{F}_t$ is generated by $\epsilon_1, \ldots, \epsilon_t$, hence $U_t = u_t(\epsilon_1, \ldots, \epsilon_t)$ and $\underline{U}_t = \underline{u}_t(\epsilon_1, \ldots, \epsilon_t)$. Thus, processes U and $\underline{U}$ coincide if and only if the functions u_t and $\underline{u}_t$ concide. We prove the latter assertion by backward induction.

The assertion is satisfied at time T by the definition of the functions u_t and $\underline{u}_t$.

For $t < T$, the equality $u_{t+1}(\epsilon_1, \epsilon_2, \ldots, \epsilon_{t+1}) = \underline{u}_{t+1}(\epsilon_1, \epsilon_2, \ldots, \epsilon_{t+1})$ holds by assumption. This allows us to write the function u_t as:

$$\begin{aligned}&u_t(e_1, \ldots, e_t) = \\ &\quad = \max\left(f_t(e_1, \ldots, e_t), \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[u_{t+1}(\epsilon_1, \ldots, \epsilon_{t+1}) \mid \epsilon_1 = e_1, \ldots, \epsilon_t = e_t]\right) \\ &\quad = \max\left(f_t(e_1, \ldots, e_t), \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[\underline{u}_{t+1}(\epsilon_1, \ldots, \epsilon_{t+1}) \mid \epsilon_1 = e_1, \ldots, \epsilon_t = e_t]\right)\end{aligned}$$

Given the definition of $\underline{u}_t$ in the equation (2.12) it suffices to show that the infimum in the second term in the maximum above is attained for $\underline{P}$. For

that purpose we define a sequence of functions

$$w_t(e_1,\dots,e_t) = \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[\underline{u}_{t+1}(\epsilon_1,\dots,\epsilon_{t+1}) \mid \epsilon_1 = e_1,\dots,\epsilon_t = e_t], \quad (2.13)$$

for $t = 0,\dots,T-1$. To complete the proof of the lemma we will demonstrate that the following claim holds:

Claim: For each $t \in \{0,\dots,T-1\}$

i) the function w_t is increasing in every argument,

ii) the equality $w_t(e_1,\dots,e_t) = E^{\underline{P}}[\underline{u}_{t+1}(e_1,\dots,,e_t,\epsilon_{t+1})]$ holds.

Indeed, the definition of functions $\underline{u}_t$ implies that the functions w_t satisfy the recursion:

$$w_t(e_1,\dots,e_t) = \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\left[\max\left(f_{t+1}(e_1,\dots,e_{t+1}), w_{t+1}(e_1,\dots,e_{t+1})\right) \mid \epsilon_1 = e_1,\dots,\epsilon_t = e_t\right] \quad (2.14)$$

for $t = 0,\dots,T-2$ which allows us to prove the claim by backward induction.

For the induction base, note that

$$w_{T-1}(e_1,\dots,e_{T-1}) = \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[f_T(\epsilon_1,\dots,\epsilon_T) \mid \epsilon_1 = e_1,\dots,\epsilon_{T-1} = e_{T-1}].$$

The function f_T is increasing in every argument by assumption. Hence, using lemma 2.B.1 we conclude that

$$\begin{aligned} w_{T-1}(e_1,\dots,e_{T-1}) &= E^{\underline{P}}[f_T(\epsilon_1,\dots,\epsilon_T) \mid \epsilon_1 = e_1,\dots,\epsilon_{T-1} = e_{T-1}] \\ &= E^{\underline{P}}[f_T(\epsilon_1,\dots,\epsilon_T)] \end{aligned}$$

where the last equality holds due to the fact that $\epsilon_1,\dots,\epsilon_t$ are random variables that are independent under the measure $\underline{P}$. Hence, part *ii*) of the claim holds and part *i*) follows from basic properties of integration[17].

[17] See also Lemma 2.D.2 bellow.

Assume now that $t < T-1$. Since the function $\max(f_{t+1}, w_{t+1})$ is increasing by assumptions on f_{t+1} and w_{t+1} we can apply lemma 2.B.1 to w_t (as written in (2.14)) to obtain:

$$w_t(e_1,\ldots,e_t) = E^{\underline{P}}\left[\max\left(f_{t+1}(e_1,\ldots,e_t,\epsilon_{t+1}), w_{t+1}(e_1,\ldots,e_t,\epsilon_{t+1})\right)\right],$$

where, similarly as before, we used the independence under $\underline{P}$. Again, basic integration properties imply that part i) of the claim holds. Using the part ii) of the inductive assumption we can rewrite the last equation as:

$$\begin{aligned} w_t(e_1,\ldots,e_t) &= E^{\underline{P}}\left[\max\left(f_{t+1}(e_1,\ldots,e_t,\epsilon_{t+1}), E^{\underline{P}}[\underline{u}_{t+2}(e_1,\ldots,e_t,\epsilon_{t+1},\epsilon_{t+2})]\right)\right] \\ &= E^{\underline{P}}[\underline{u}_{t+1}(e_1,\ldots,e_t,\epsilon_{t+1})]. \end{aligned}$$

Thus, the part ii) of the claim holds, too. This completes the proof of the claim and, hence, the proof of the lemma. □

Appendix 2.C Equivalence of Problems 3 and 4

A version of this result appears in Chudjakow and Riedel (2013); the analysis we offer contains additional details and, due the generality of the construction of the set of priors $\mathcal{P}$, applies to a broader class of problems.

We begin by proving an auxiliary result based on Lemma 8 in Riedel (2009).

Lemma 2.C.1 (Iterated version of Lemma 8 in Riedel 2009 and corollaries)**.** *Let $P_1, P_2, \ldots, P_n$ be measures in $\mathcal{P}$, $\tau \in \mathcal{T}$ a stopping time, and $A_1, A_2, \ldots, A_n$ sets in $\mathcal{F}_\tau$ that form a partition of Ω.*

1. *There exists a measure $P \in \mathcal{P}$ such that for any r.v. Z:*

$$E^P[Z \mid \mathcal{F}_\tau] = \sum_{k=1}^{n} E^{P_k}[Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau]. \tag{2.15}$$

2. *For any r.v.* Z *and any* $k \in \{1, \ldots, n\}$ *the equality* $E^P[Z\mathbb{1}_{A_k}] = E^{P_k}[Z\mathbb{1}_{A_k}]$ *holds.*

3. *The equality* $E^P[Z] = \sum_{k=1}^n E^{P_k}[Z\mathbb{1}_{A_k}]$ *holds for any r.v.* Z *and any* $k \in \{1, \ldots, n\}$.

Proof. 1. This claim is just an iterated version of Lemma 8 in Riedel (2009).

2. First, we note that for any set A_k and any r.v. Z, by plugging $Z\mathbb{1}_{A_k}$ in (2.15) we have:

$$E^P[Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau] = \sum_{i=1}^n E^{P_i}[Z\mathbb{1}_{A_k}\mathbb{1}_{A_i} \mid \mathcal{F}_\tau] = E^{P_k}[Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau] \tag{2.16}$$

In particular for an arbitrary set $B \in \mathcal{F}_\tau$ we have:

$$E^{P_k}[\mathbb{1}_{A_k B}] = E^{P_k}[\mathbb{1}_{A_k B} \mid \mathcal{F}_\tau] = E^P[\mathbb{1}_{A_k B} \mid \mathcal{F}_\tau] = E^P[\mathbb{1}_{A_k B}], \tag{2.17}$$

Since measures P and P_k are both in $\mathcal{P}$ the Radon-Nykodim derivative $\frac{dP_k}{dP}$ is well defined. Thus, using (2.16), for an arbitrary set $B \in \mathcal{F}_\tau$, the following holds:

$$E^P\left[E^P\left[\frac{dP_k}{dP}\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right]\mathbb{1}_B\right] = E^P\left[\frac{dP_k}{dP}\mathbb{1}_{A_k}\mathbb{1}_B\right] = E^{P_k}[\mathbb{1}_{A_k B}]. \tag{2.18}$$

Combining (2.17) and (2.18) we obtain[18]:

$$E^P\left[\frac{dP_k}{dP}\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right] = \mathbb{1}_{A_k}. \tag{2.19}$$

Now, multiplying the well known identity

$$E^P\left[\frac{dP_k}{dP} \mid \mathcal{F}_\tau\right] E^{P_k}[Z \mid \mathcal{F}_\tau] = E^P\left[\frac{dP_k}{dP}Z \mid \mathcal{F}_\tau\right]$$

[18] If M is $\mathcal{F}$ measurable then $E[Z \mid \mathcal{F}] = M$ iff for any $B \in \mathcal{F}$ the equality $E[M\mathbb{1}_B] = E[Z\mathbb{1}_B]$ holds.

with $\mathbb{1}_{A_k}$ (which is $\mathcal{F}_\tau$-measurable) we obtain:

$$E^P\left[\frac{dP_k}{dP}\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right] E^{P_k}\left[Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right] = E^P\left[\frac{dP_k}{dP} Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right].$$

Using the equations (2.16) and (2.19) the last equality can be rewritten as:

$$E^P\left[Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right] = E^P\left[Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right] = E^P\left[\frac{dP_k}{dP} Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right].$$

Finally, taking expectation over P in the last equality, we obtain the desired:

$$\begin{aligned} E^P\left[Z\mathbb{1}_{A_k}\right] &= E^P\left[E^P\left[Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right]\right] = E^P\left[E^P\left[\frac{dP_k}{dP} Z\mathbb{1}_{A_k} \mid \mathcal{F}_\tau\right]\right] = E^P\left[\frac{dP_k}{dP} Z\mathbb{1}_{A_k}\right] \\ &= E^{P_k}\left[Z\mathbb{1}_{A_k}\right] \end{aligned}$$

3. Direct consequence of 2. . □

With this we are prepared for the following lemma:

Lemma 2.C.2. *For Y_t, Z_t and $\mathcal{P}$ as defined above the following equality holds:*

$$\min_{P\in\mathcal{P}} E^P\left[Z_\tau\right] = \min_{P\in\mathcal{P}} E^P\left[Y_\tau\right] \tag{2.20}$$

Proof of Lemma 2.C.2. Let us, for each t, denote by Q_t and R_t measures that minimize the adapted and non-adapted payoffs at time $\tau = t$, i.e. $Z_t\mathbb{1}_{\{\tau=t\}}$ and $Y_t\mathbb{1}_{\{\tau=t\}}$, respectfully (the existence of these measures is guaranteed by Riedel (2009); see Lemma 10 therein). Using the law of iterated expectation for multiple priors (Lemma 4 in Riedel (2009)) we obtain:

$$\begin{aligned} E^{Q_t}\left[Z_t\mathbb{1}_{\{\tau=t\}}\right] &= \min_{P\in\mathcal{P}} E\left[Z_t\mathbb{1}_{\{\tau=t\}}\right] \\ &= \min_{P\in\mathcal{P}}\left[\operatorname*{ess\,inf}_{P'\in\mathcal{P}} E^{P'}[Y_t \mid \mathcal{F}_t]\mathbb{1}_{\{\tau=t\}}\right] \\ &= \min_{P\in\mathcal{P}}\left[\operatorname*{ess\,inf}_{P'\in\mathcal{P}} E^{P'}[Y_t\mathbb{1}_{\{\tau=t\}} \mid \mathcal{F}_t]\right] \\ &= \min_{P\in\mathcal{P}}\left[Y_t\mathbb{1}_{\{\tau=t\}}\right] = E^{R_t}\left[Y_t\mathbb{1}_{\{\tau=t\}}\right]. \end{aligned} \tag{2.21}$$

By the third claim of lemma 2.C.1 above there exist the measures $Q, R \in \mathcal{P}$ such that $\sum_{t=1}^{T} E^{Q_t}[Z_t \mathbb{1}_{\{\tau=t\}}] = E^{Q}[Z_\tau]$ and $\sum_{t=1}^{T} E^{R_t}[Z_t \mathbb{1}_{\{\tau=t\}}] = E^{R}[Z_\tau]$. Combining the second claim of the same lemma with the equation (2.21) above we have, for each t:

$$E^{Q}[Z_t \mathbb{1}_{\{\tau=t\}}] = E^{Q_t}[Z_t \mathbb{1}_{\{\tau=t\}}] = E^{R_t}[Z_t \mathbb{1}_{\{\tau=t\}}] = E^{R}[Z_t \mathbb{1}_{\{\tau=t\}}] \tag{2.22}$$

Furthermore, $\arg\min_{P\in\mathcal{P}} E^{P}[Z_t \mathbb{1}_{\{\tau=t\}}] = Q$ for each t, which allows us to write:

$$\min_{P\in\mathcal{P}} E^{P}\left[Z_\tau\right] = \min_{P\in\mathcal{P}} \sum_{t=1}^{T} E^{P}\left[Z_t \mathbb{1}_{\{\tau=t\}}\right] = \sum_{t=1}^{T} E^{Q}\left[Z_t \mathbb{1}_{\{\tau=t\}}\right] = E^{Q}\left[Z_\tau\right]. \tag{2.23}$$

Similarly, we conclude that:

$$\min_{P\in\mathcal{P}} E^{P}\left[Y_\tau\right] = \min_{P\in\mathcal{P}} \sum_{t=1}^{T} E^{P}\left[Y_t \mathbb{1}_{\{\tau=t\}}\right] = \sum_{t=1}^{T} E^{R}\left[Y_t \mathbb{1}_{\{\tau=t\}}\right] = E^{R}\left[Y_\tau\right]. \tag{2.24}$$

Finally, the left hand side of equation (2.23) and (2.24) are equal because of (2.22), hence the right hand sides are also equal, which completes the proof. □

Appendix 2.D Proof of Theorem 2.3.1

For the sake of convenience, we begin by defining a sequence of functions

$$i_{t+1}(x_1, \ldots, x_t, x_{t+1}) = \mathbb{1}_{x_{t+1} \leq \max(x_1, \ldots, x_t)}$$

for $t < T$. Note that this allows the random variable $\mathbb{1}_{X_{t+1}\leq M_t}$ to be written in terms of the function i_{t+1} as follows:

$$\mathbb{1}_{X_{t+1}\leq M_t} = i_{t+1}(X_1, \ldots, X_t, X_{t+1}).$$

As a preparation for the proof of theorem 2.3.1 we prove a result on the representation of the payoff process Z_t.

Since $X_1, \ldots, X_t$ are independent under $\overline{P}$ we can derive the following representation for functions r_t:

$$\overline{r}_t(m) = \prod_{s>t} E^{\overline{P}}\left[\mathbb{1}_{X_s \leq m}\right] = \left(\overline{v}(X_1 \leq m)\right)^{T-t},$$

where the second equality is due to the definition of the measure $\overline{v}$. It is now obvious that $\overline{r}_t$ is an increasing function.

The next lemma describes the expected (ambiguous) Z_t in terms of the function $\overline{r}_t$:

Lemma 2.D.1. *For each $t \in \{1, \ldots, T\}$ the following representation holds:*

$$Z_t = \mathbb{1}_{\{X_t = M_t\}} \overline{r}_t(X_t) \tag{2.25}$$

Proof. Note that:

$$Z_t = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P\left[\mathbb{1}_{\{X_t = M_t = M_T\}} \mid \mathcal{F}_t\right] = \mathbb{1}_{\{X_t = M_t\}} \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P\left[\mathbb{1}_{\{M_t = M_T\}} \mid \mathcal{F}_t\right]. \tag{2.26}$$

Define the process:

$$R_t = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P\left[\mathbb{1}_{\{M_t = M_T\}} \mid \mathcal{F}_t\right],$$

and the function:

$$r_t(x_1, \ldots, x_t) = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P\left[\mathbb{1}_{\{M_t = M_T\}} \mid X_1 = x_1, \ldots, X_t = x_t\right].$$

Clearly, the following equalities hold:

$$Z_t = R_t \mathbb{1}_{\{X_t = M_t\}} = r_t(X_1, \ldots, X_t) \mathbb{1}_{\{X_t = M_t\}}. \tag{2.27}$$

Thus it suffices to show the following:

Claim: For each t the equality $R_t = \overline{r}_t(M_t)$ holds almost surely.

The claim is proven by backward induction.

Since $R_T = \overline{r}_T(M_T) = 1$ the claim trivially holds in the last period so we turn to the case $t < T$.

We begin by deriving a recursive expression for R_t (using the law of iterated expectations for multiple priors[19]) as follows:

$$\begin{aligned} R_t &= \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\left[\mathbb{1}_{M_t=M_{t+1}=M_T} \mid \mathcal{F}_t\right] \\ &= \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\left[\operatorname*{ess\,inf}_{Q\in\mathcal{P}} E^Q\left[\mathbb{1}_{M_{t+1}=M_T} \mid \mathcal{F}_{t+1}\right]\mathbb{1}_{M_t=M_{t+1}} \mid \mathcal{F}_t\right] \\ &= \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\left[R_{t+1}\mathbb{1}_{M_t\geq X_{t+1}} \mid \mathcal{F}_t\right]. \end{aligned}$$

If we denote the realization of M_t with m_t (i.e. $m_t = \max(x_1,\ldots,x_t)$) we can rewrite the last equality in terms of the functions r_t and $\overline{r}_t$ using (2.27) and the induction hypothesis as follows:

$$\begin{aligned} r_t(x_1,\ldots,x_t) &= \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\left[\overline{r}_{t+1}(M_{t+1})\mathbb{1}_{X_{t+1}\leq m_t} \mid X_1 = x_1,\ldots,X_t = x_t\right] \\ &= \overline{r}_{t+1}(m_{t+1}) \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\left[\mathbb{1}_{X_{t+1}\leq m_t} \mid X_1 = x_1,\ldots,X_t = x_t\right]. \end{aligned}$$

In the last equality above we used the fact that on the set $\{X_{t+1} \leq m_t\}$ the equality $M_{t+1} = M_t$ holds.

Since $\mathbb{1}_{X_{t+1}\leq M_t} = i_{t+1}(X_1,\ldots,X_t,X_{t+1})$ and the function i_{t+1} is decreasing in its last variable we can use lemma 2.B.1 to identify $\overline{P}$ as the minimizing measure in the last expression:

$$\begin{aligned} r_t(x_1,\ldots,x_t) &= \overline{r}_{t+1}(m_t)E^{\overline{P}}\left[\mathbb{1}_{X_{t+1}\leq m_t} \mid X_1 = x_1,\ldots,X_t = x_t\right] \\ &= \left(\prod_{s>t+1} E^{\overline{P}}\left[\mathbb{1}_{X_s\leq m_t}\right]\right) E^{\overline{P}}\left[\mathbb{1}_{X_{t+1}\leq m_t}\right] = \overline{r}_t(m_t); \end{aligned}$$

the last equality is due to the definition of $\overline{r}_t$. □

[19] Lemma 4 in Riedel (2009).

We note that Lemma 2.D.1 proves that infimum in the definition of the adapted payoff Z_t is attained for $\overline{v}$.

For the sake of convenience we also state a simple result about monotonicity of integral functions in the setting of our problem.

Lemma 2.D.2. *Let $g(x_1, \ldots, x_t, x_{t+1})$, $t < T$, be a function increasing (decreasing) in each of the first t arguments. For any $P \in \mathcal{P}$ the function*

$$h^P(x_1, \ldots, x_t) = E^P[g(X_1, \ldots, X_t, X_{t+1}) \mid X_1 = x_1, \ldots X_t = x_t]$$

is increasing (decreasing) in every argument, as is the function

$$h(x_1, \ldots, x_t) = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[g(X_1, \ldots, X_t, X_{t+1}) \mid X_1 = x_1, \ldots X_t = x_t].$$

Proof. The elementary proof of the first part of the lemma is omitted. Once one notices that $h = \operatorname{ess\,inf}_{P \in \mathcal{P}} h^P$ the second part follows immediately from the first part and the properties of the essential infimum. □

We turn to proving the core of the theorem and for that purpose we define the value process U of the FIBC optimal stopping problem under multiple priors:

$$U_T = Z_T; \quad U_t = \max(Z_t, \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[U_{t+1} \mid \mathcal{F}_t]),\ t < T.$$

The analysis will focus on the properties of the second argument in the maximum above so we define:

$$W_t = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[U_{t+1} \mid \mathcal{F}_t],\ 0 \leq t < T.$$

As can be seen from the value process, the random variable W_t describes the expected value (under multiple priors) of the payoff the agent will receive if she does not stop at time t given the available information. The definition above implies:

$$W_{T-1} = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[Z_T \mid \mathcal{F}_{T-1}], \quad W_t = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[\max(Z_{t+1}, W_{t+1}) \mid \mathcal{F}_t].$$

If we introduce the sequence of functions:

$$w_t^*(x_1,\dots,x_t) = \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[U_{t+1} \mid X_1 = x_1,\dots,X_t = x_t],\ 0 \le t < T.$$

the equality $W_t = w_t^*(X_1,\dots,X_t)$ clearly holds. Furthermore:

Lemma 2.D.3. *For each $t \in \{0,1,\dots,T-1\}$ the function w_t^* is decreasing in every variable.*

Proof. The proof is by backward induction.

We first consider w_{T-1}^*. Notice that:

$$\begin{aligned} w_{T-1}^*(x_1,\dots,x_{T-1}) &= \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[Z_T \mid \mathcal{F}_{T-1}] \\ &= \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[\mathbb{1}_{X_T=M_T} \mid X_1 = x_1,\dots,X_{T-1} = x_{T-1}] \\ &= \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P[i_T(x_1,\dots,x_{T-1},X_T) \mid X_1 = x_1,\dots,X_{T-1} = x_{T-1}] \end{aligned}$$

Since i_T is obviously decreasing in first $T-1$ variables, we can use the above lemma 2.D.2 to conclude that w_{T-1}^* is decreasing in every variable.

For $t < T-1$ we have:

$$\begin{aligned} w_t^*(x_1,\dots,x_t) = \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\Big[&\max\big(\mathbb{1}_{X_{t+1}=M_{t+1}}\overline{r}_{t+1}(X_t+1), \\ &\qquad w_{t+1}^*(x_1,\dots,x_t,X_{t+1})\big) \mid X_1 = x_1,\dots,X_t = x_t\Big] \\ = \operatorname*{ess\,inf}_{P\in\mathcal{P}} E^P\Big[&\max\big(i_{t+1}(x_1,\dots,x_t,X_{t+1})\overline{r}_{t+1}(X_t+1), \\ &\qquad w_{t+1}^*(x_1,\dots,x_t,X_{t+1})\big) \mid X_1 = x_1,\dots,X_t = x_t\Big]. \end{aligned}$$

The function i_{t+1} is decreasing in its first t arguments and the function $\overline{r}_{t+1}$ is decreasing in every argument. The function w_{t+1}^* is decreasing in every argument by assumption. Thus, the result now follows from the fact that the maximum of decreasing function is a decreasing function and the lemma 2.D.2.

□

The last result allows us to formulate a simple representation of the process W_t:

Lemma 2.D.4. *For each $t \in \{0, 1, \ldots, T-1\}$ there exists a decreasing function $w_t(m)$ such that $W_t = w_t(M_t)$.*

Proof. We begin the proof by backward induction by noting that, since $Z_T = \mathbb{1}_{X_T = M_T}$ and

$$W_{T-1} = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[Z_T \mid \mathcal{F}_{T-1}] = \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[\mathbb{1}_{X_T = M_T} \mid \mathcal{F}_{T-1}],$$

we have, due to the definition of w^*_{T-1},

$$\begin{aligned} w^*(x_1, \ldots, x_{T-1}) &= \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[i_T(x_1, \ldots, x_{T-1}, X_T) \mid X_1 = x_1, \ldots, X_{T-1} = x_{T-1}] \\ &= \underline{E^P}[i_T(x_1, \ldots, x_{T-1}, X_T) \mid X_1 = x_1, \ldots, X_{T-1} = x_{T-1}] \\ &= \underline{P}(X_T \geq M_{T-1}), \end{aligned}$$

where the second equality is due to lemma 2.B.1. It thus suffices to define $w_{T-1}(m) = \underline{P}(X_T \geq m)$. Indeed, the function w_{T-1} is clearly decreasing and the equality $w_{T-1}(M_t) = W_T$ holds because of the previous considerations.

Suppose that for $t < T$ there exists a decreasing function w_{t+1} such that $w_{t+1}(M_{t+1}) = W_{t+1}$. This allows us to rewrite W_t in terms of w_{t+1} and $\overline{r}_{t+1}$:

$$\begin{aligned} W_t &= \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P[\max(Z_{t+1}, W_{t+1}) \mid \mathcal{F}_t] = \\ &= \operatorname*{ess\,inf}_{P \in \mathcal{P}} E^P\left[\max\left(\overline{r}_{t+1}(M_{t+1}) \cdot \mathbb{1}_{X_{t+1} = M_{t+1}}, w_{t+1}(M_{t+1})\right) \mid \mathcal{F}_t\right] \\ &= \operatorname*{ess\,inf}_{P \in \mathcal{P}} \Big(E^P\left[\max\left(\overline{r}_{t+1}(X_{t+1}), w_{t+1}(X_{t+1})\right) \cdot \mathbb{1}_{X_{t+1} \geq M_t} \mid \mathcal{F}_t\right] \\ &\qquad + w_{t+1}(M_{t+1}) \cdot E^P[\mathbb{1}_{X_{t+1} < M_t} \mid \mathcal{F}_t]\Big), \end{aligned}$$

where the last equality is due to:

$$\begin{aligned} &\max\left(\overline{r}_{t+1}(M_{t+1}) \cdot \mathbb{1}_{X_{t+1} = M_{t+1}}, w_{t+1}(M_{t+1})\right) = \\ &\max\left(\overline{r}_{t+1}(X_{t+1}), w_{t+1}(X_{t+1})\right) \cdot \mathbb{1}_{X_{t+1} \geq M_t} + w_{t+1}(M_{t+1}) \cdot \mathbb{1}_{X_{t+1} < M_t}. \end{aligned}$$

Since $W_t = w_t^*(X_1, \ldots, X_t)$, and $M_t = M_{t+1}$ on the set X_{t+1}, we can write:

$$
\begin{aligned}
&w_t^*(x_1, \ldots, x_t) = \\
&\operatorname*{ess\,inf}_{P \in \mathcal{P}} \Big(E^P \left[\max(\overline{r}_{t+1}(X_{t+1}), w_{t+1}(X_{t+1})) \cdot \mathbb{1}_{X_{t+1} \geq M_t} \mid X_1 = x_1, \ldots, X_t = x_t \right] \\
&\qquad\qquad + w_{t+1}(M_t) \cdot E^P[\mathbb{1}_{X_{t+1} < M_t} \mid X_1 = x_1, \ldots, X_t = x_t] \Big) \\
&= \operatorname*{ess\,inf}_{v \in \mathcal{V}^A} \Big(\int_{m_t}^1 \max(\overline{r}_{t+1}(x), w_{t+1}(x))\, dv + w_{t+1}(m_t) \int_0^{m_t} dv \Big),
\end{aligned}
$$

where $m_t = \max(x_1, \ldots, x_t)$ and the last equality is due to the definition of the set $\mathcal{P}$ in section 3. Thus, by setting

$$
w_t(m) = \operatorname*{ess\,inf}_{v \in \mathcal{V}^A} \Big(\int_m^1 \max(\overline{r}_{t+1}(x), w_{t+1}(x))\, dv + w_{t+1}(m) \int_0^m dv \Big), \quad (2.28)
$$

for $t < T$, we get $w_t(m_t) = w_t^*(x_1, \ldots, x_t)$ which, due to the definition of w_t^*, implies $w_t(M_t) = W_t$.

Finally, since $w_t(\max(x_1, \ldots, x_t)) = w_t^*(x_1, \ldots, x_t)$, the function w_t^* is symmetric; thus, the monotonicity of the function w_t is a consequence of the monotonicity of the function w_t^* as described by the lemma 2.D.3. □

We now turn to proving that the stopping time is of the threshold type.

The proof of the last lemma reveals that the functions w_t are defined by the recursion $w_{T-1}(m) = \underline{P}(X_T \geq m)$ and, for $t < T - 1$, the equation (2.28). Equivalently, we can expand the definition to include the final period by setting $w_T(m) = 0$ and w_t as defined by the expression in the equation (2.28) for $t < T$.

It is clear that, for each $t < T$, the equalities $w_t(1) = \overline{r}_t(0) = 0$ hold and that the functions r_t are strictly increasing, while the functions w_t are (weakly) decreasing. Thus, for $t < T$, there exists a unique $b_t \in [0, 1)$ such that $w_t(b_t) = \overline{r}_t(b_t)$. Additionally, we define $b_T = 0$. We record the previous considerations, along with the proof of the monotonicity of the sequence (b_t), in the following lemma.

Lemma 2.D.5. *For each $t < T$ there exists a unique $b_t \in [0,1]$ such that the equality $w_t(b_t) = \overline{r}_t(b_t)$ holds. Furthermore, for each $t < T$ the inequality $b_t > b_{t+1}$ holds.*

Proof. Suppose $t < T$. Note that, due to the definition of the sequence (b_t) and the fact that the function $\overline{r}_{t+1}$ is strictly increasing, the following (in)equalities hold

$$\max(\overline{r}_{t+1}(x), w_{t+1}(x)) = \overline{r}_{t+1}(x) > \overline{r}_{t+1}(b_{t+1}) = w_{t+1}(b_t),$$

for each $x \in (b_{t+1}, 1]$. Hence:

$$\begin{aligned} w_t(b_{t+1}) &= \operatorname*{ess\,inf}_{v \in \mathcal{V}^A} \left(\int_{b_{t+1}}^{1} \max(\overline{r}_{t+1}(x), w_{t+1}(x))\, dv + w_{t+1}(b_{t+1}) \int_{0}^{b_{t+1}} dv \right) \\ &> \operatorname*{ess\,inf}_{v \in \mathcal{V}^A} \left(\int_{b_{t+1}}^{1} w_{t+1}(b_{t+1})\, dv + w_{t+1}(b_{t+1}) \int_{0}^{b_{t+1}} dv \right) = w_{t+1}(b_{t+1}). \end{aligned} \tag{2.29}$$

Note, also, that the definition of $\overline{r}_t$ implies $\overline{r}_t(x) < \overline{r}_{t+1}(x)$ for any $x \in (0,1)$. Thus, given the previously obtained inequality (2.29), we get:

$$w_t(b_{t+1}) > w_{t+1}(b_{t+1}) = \overline{r}_{t+1}(b_{t+1}) > \ \overline{r}_t(b_{t+1}). \tag{2.30}$$

With the inequality (2.30) proven we can turn to proving the inequality stated in the formulation of the lemma.

Suppose the opposite: $b_t \leq b_{t+1}$. The definition of b_t and the monotonicity of r_t imply: $w_t(b_t) = \overline{r}_t(b_t) \leq \overline{r}_t(b_{t+1}) < w_t(b_{t+1})$, where the last inequality is due to the previously proven inequality (2.30). This, however, is in contradiction with the monotonicity of w_t.

□

To complete the proof of the first two parts of the theorem it remains to prove the equality (2.4); we do so in the following lemma:

Lemma 2.D.6.

$$\tau^* = \min\{t \,|\, X_t = M_t > b_t\}$$

Proof. In the context of FIBC the optimal stopping time is given with:

$$\tau^* = \min\{t \mid Z_t = U_t\} = \min\{t \mid Z_t \geq W_t\}.$$

Using the representations for Z_t and W_t obtained in lemmas 2.D.1 and 2.D.4 respectfully, the inequality $Z_t \geq W_t = w_t(M_t)$ can only be satisfied when $X_t = M_t$ (in which case $Z_t = \overline{r}_t(M_t)$), hence:

$$\tau^* = \min\{t \mid X_t = M_t,\, \overline{r}_t(M_t) \geq w_t(M_t)\}.$$

Finally, due to the monotonicity of $\overline{r}_t$ and w_t and lemma 2.D.5, the inequality $\overline{r}_t(X_t) \geq w_t(X_t)$ is satisfied only when $b_t \leq M_t = X_t$.

□

It remains to note that the essential infimum in (2.28) is attained (see Lemma 10 in Riedel (2009)). This, with the definitions of W_t and U_t, and Lemma 2.D.1 proves the third part of the theorem. Indeed, before stopping the minimizing measure is the one attained in (2.28), and once the agent stops the her payoff is Z_t, and lemma 2.D.1 implies that the minimizing measure is $\overline{v}$.

Appendix 2.E Proof of Lemma 2.4.1

Proof of claim 1 of Lemma 2.4.1. We define an operator $G : L^1([0,1]) \to \mathbb{R}$ with

$$G\varphi = \int_0^1 g(x)\varphi(x)\,dx,$$

and note that it is (Lipschitz) L^1-continous. Indeed, using the fact that g is increasing and bounded:

$$|G\varphi_1 - G\varphi_2| = \left| \int_0^1 g(x)(\varphi_1(x) - \varphi_2(x))\, dx \right| \leq C \left| \int_0^1 \varphi_1(x) - \varphi_2(x)\, dx \right|,$$

where C is a positive constant that bounds $|g(x)|$.

Let D_S^λ be the set of all the step functions within the set D_{CLA}^λ. We will prove that D_S^λ is dense[20] in D_{CLA}^λ. For an arbitrary $\varphi \in D_S^\lambda$ and an arbitrary $\varepsilon > 0$ one can choose a step function φ_1 such that $\frac{1}{\lambda} \leq \varphi_1(x) \leq \varphi(x) \leq \lambda$ and:

$$\left| \int_0^1 \varphi_1(x) - \varphi(x)\, dx \right| < \frac{\varepsilon}{2}. \tag{2.31}$$

If one defines $I = \int_0^1 \varphi_1(x)\, dx \leq 1$ and:

$$\gamma = \frac{1 - \int_B \varphi_1(x)\, dx}{\int_A \varphi_1(x)\, dx}$$

for $A = \{\varphi_1(x) \leq I\}$ and $B = [0,1] \backslash A$ it is easy to check that, for sufficiently small ε, the function $\varphi_S = \gamma\varphi_1 \mathbb{1}_A + \varphi_1 \mathbb{1}_B$ is a function that belongs to D_{CLA}^λ. Furthermore, direct calculations show that the inequality

$$\left| \int_0^1 \varphi_1(x) - \varphi_S(x)\, dx \right| < \frac{\varepsilon}{2} \tag{2.32}$$

holds[21]. Combining the inequalities (2.31) and (2.32) gives:

$$\left| \int_0^1 \varphi_S(x) - \varphi(x)\, dx \right| < \varepsilon,$$

which proves the density.

As the operator G is continuous and the set D_S^λ (which contains $\underline{\varphi}$) is dense in D_{CLA}^λ, for the claim to hold it suffices to show that for any $\varphi \in D_S^\lambda$

[20] With respect to L^1 metric.

[21] The inequality (2.32) is in fact equivalent to (2.31).

the inequality $G\varphi \geq G\underline{\varphi}$ holds. We do so in the remainder of the proof.

Let us fix $\varphi \in D_S^\lambda$:

$$\varphi = \sum_{i=1}^{n} d_i \mathbb{1}_{[c_{i-1},c_i]}, \in D_{CLA}^\lambda. \tag{2.33}$$

Without loss of generality we can assume that there is an index $m \in \{1, \ldots, n\}$ such that $c_m = 1/(1+\lambda)$.

Set $\varphi_0 = \varphi$. The idea is to create a finite sequence of functions (φ_i) in which the last element is $\underline{\varphi}$ with the inequality $G\varphi_i \leq G\varphi_{i-1}$ being satisfied for any $i > 0$.

If $\varphi_0 = \underline{\varphi}$ the proof is done. If not, we choose the step function φ_1 such that it differs from φ_0 by the value it takes on two appropriately chosen intervals. For that purpose we define:

$$j = \min\{i \mid d_i < \lambda\}, \quad j' = \max\{i \mid d_i > 1/\lambda\}.$$

Note that since $\varphi_0 \neq \underline{\varphi}$ we have $j < m < j'$. We now focus on the intervals $[c_{j-1}, c_j]$ and $[c_{j'-1}, c_{j'}]$ and set the value of φ_1 to be λ on the former interval or $1/\lambda$ on the latter by "repositioning the weight" of φ_0.

If $(\lambda - d_j)(c_j - c_{j-1}) \leq (d_{j'} - \frac{1}{\lambda})(c_{j'-1} - c_{j'})$ we "reposition the excess weight" from the interval $[c_{j'-1}, c_{j'}]$ to the interval $[c_{j-1}, c_j]$, that is we define:

$$\begin{aligned}\varphi_1 =& \varphi_0 \mathbb{1}_{[0,1]\setminus([c_{j-1},c_j]\cup[c_{j'-1},c_{j'}])} + \\ & \left(d_j + (d_{j'} - \tfrac{1}{\lambda})\frac{c_{j'-1} - c_{j'}}{c_j - c_{j-1}}\right)\mathbb{1}_{[c_{j-1},c_j]} + \frac{1}{\lambda}\mathbb{1}_{[c_{j'-1},c_{j'}]}.\end{aligned}$$

The inequality $G\varphi_o \leq G\varphi_1$ is satisfied. Indeed, direct calculation yields

$$\begin{aligned}G\varphi_o - G\varphi_1 &= \int_{c_{j-1}}^{c_j} g(x)(\varphi_0(x) - \varphi_1(x))\,dx + \int_{c_{j'-1}}^{c_{j'}} g(x)(\varphi_0(x) - \varphi_1(x))\,dx \\ &= -(d_{j'} - \tfrac{1}{\lambda})\frac{c_{j'-1} - c_{j'}}{c_j - c_{j-1}}\int_{c_{j-1}}^{c_j} g(x)\,dx + (d_{j'} - \tfrac{1}{\lambda})\int_{c_{j'-1}}^{c_{j'}} g(x)\,dx,\end{aligned}$$

and one can use the monotonicity of the function g and the inequalities

$j < m < j'$ to make the following estimation:

$$G\varphi_o - G\varphi_1 \geq (d_{j'} - \tfrac{1}{\lambda}) \left(g(c_{j'-1})(c_{j'-1} - c_{j'}) - \frac{c_{j'-1} - c_{j'}}{c_j - c_{j-1}} g(c_j)(c_{j-1} - c_j) \right)$$
$$= (d_{j'} - \tfrac{1}{\lambda})(c_{j'-1} - c_{j'})(g(c_{j'-1}) - g(c_j) \geq 0.$$

When the inequality $(\lambda - d_j)(c_j - c_{j-1}) > (d_{j'} - \frac{1}{\lambda})(c_{j'-1} - c_{j'})$ holds one can construct the function φ_1 using an analogous "weight repositioning".

If $\varphi_1 = \underline{\varphi}$ the proof is done. If not, one can create φ_2 from φ_1 as above. As the step function φ has finitely many steps the procedure ends after finitely many iterations.

□

Proof of claim 2 of Lemma 2.4.1. We begin by fixing $\varphi \in D^{\lambda}_{CLA}$ and defining:

$$\mu_1 = \int_0^k \varphi(x)\, dx, \quad \mu_2 = \int_k^1 \varphi(x)\, dx.$$

We will identify two functions ψ_1 and ψ_2, defined on $[0, k]$ and $[k, 1]$ respectively, such that the function $\psi := \psi_1 \mathbb{1}_{[0,k]} + \psi_2 \mathbb{1}_{(k,1]}$ is the one that satisfies the claim. These will be the functions that "reposition the weight" μ_1 and μ_2 within the intervals $[0, k]$ and $[k, 1]$, such that most weight is on the upper part of the former and the lower part of the latter.

First we focus on the interval $[0, k]$. The first claim showed how to identify the step function $\overline{\varphi}$ that, for a fixed, decreasing and bounded function g, minimizes the integral on the right hand side of (2.E) among all the functions φ whose range is within the interval $[1/\lambda, \lambda]$ and whose total weight is equal to 1. Note that $\overline{\varphi}$ was simply the function that put the most weight possible on the upper part of the interval $[0, 1]$. Focusing on the interval [0,k], where the function h is decreasing, we are in a similar situation: among all the functions with a range within $[1/\lambda, \lambda]$ and whose integral is equal to μ_1 we are looking for a function ψ_1 that minimizes the integral $\int_0^k h(x)\psi_1(x)\, dx$. Analogous reasoning to the one in the proof of the first claim[22] will lead us

[22]Note that the first claim could have been formulated for functions on any interval

to the conclusion that ψ_1 has to be the function that puts the most weight as possible on the upper part of the interval $[0, k]$:

$$\psi_1 = \frac{1}{\lambda}\mathbb{1}_{[0,c_1]} + \lambda\mathbb{1}_{(c_1,k]}, .$$

for an appropriately chosen c_1. Identifying the precise value of c_1 is not difficult: since the inequalities clearly $\frac{k}{\lambda} \leq \mu_1 \leq k\lambda$ hold, there exists $c_1 \in [0, k]$ such that:

$$\frac{c_1}{\lambda} + \lambda(k - c_1) = \mu_1.$$

This proves the inequality:

$$\int_0^k h(x)\varphi(x)\,dx \geq \int_0^k h(x)\psi_1(x)\,dx. \tag{2.34}$$

Similarly, by focusing on the interval $[k, 1]$ one can identify the function ψ_2 (and the corresponding c_2) which puts the most weight on the lower part of the interval, such that:

$$\int_k^1 h(x)\varphi(x)\,dx \geq \int_k^1 h(x)\psi_2(x)\,dx = \int_k^1 h(x)\left(\lambda\mathbb{1}_{[k,c_2]} + \frac{1}{\lambda}\mathbb{1}_{(c_2,1]}\right)\,dx. \tag{2.35}$$

Direct calculations show that $D_U^\lambda \ni \psi := \psi_1\mathbb{1}_{[0,k]} + \psi_2\mathbb{1}_{(k,1]}$ and the claim follows by combining (2.34) and (2.35). □

$[a, b]$, and with total weight of densities being equal to any number (as opposed to 1); we chose not to do so for the sake of readability.

References

Artzner, Philippe et al. (1999). "Coherent measures of risk". In: *Mathematical finance* 9.3, pp. 203–228.

Bojdecki, Tomasz (1978). "On optimal stopping of a sequence of independent random variables — probability maximizing approach". In: *Stochastic Processes and their Applications* 6.2, pp. 153–163.

Bruss, F Thomas, Thomas S Ferguson, et al. (2002). "High-risk and competitive investment models". In: *The Annals of Applied Probability* 12.4, pp. 1202–1226.

Chudjakow, Tatjana and Frank Riedel (2013). "The best choice problem under ambiguity". In: *Economic Theory* 54.1, pp. 77–97.

Corbin, Ruth M, Chester L Olson, and Mona Abbondanza (1975). "Context effects in optional stopping decisions". In: *Organizational Behavior and Human Performance* 14.2, pp. 207–216.

Dehaene, Stanislas (2003). "The neural basis of the Weber–Fechner law: a logarithmic mental number line". In: *Trends in cognitive sciences* 7.4, pp. 145–147.

Ellsberg, Daniel (1961). "Risk, ambiguity, and the Savage axioms". In: *The quarterly journal of economics*, pp. 643–669.

Ferguson, Thomas S. (1989). "Who Solved the Secretary Problem?" In: *Statistical Science* 4.3, pp. 282–289.

Ferguson, Thomas S. (2006). *Optimal Stopping and Applications*. Electronic text: https://www.math.ucla.edu/ tom/Stopping/Contents.html, Mathematics Department, UCLA.

Föllmer, Hans and Alexander Schied (2011). *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

Gilbert, John P. and Frederick Mosteller (1966). "Recognizing the Maximum of a Sequence". In: *Journal of the American Statistical Association* 61.313, pp. 35–73.

Gilboa, Itzhak and Massimo Marinacci (2016). "Ambiguity and the Bayesian paradigm". In: *Readings in Formal Epistemology*. Springer, pp. 385–439.

Gilboa, Itzhak and David Schmeidler (1989). "Maxmin expected utility with non-unique prior". In: *Journal of mathematical economics* 18.2, pp. 141–153.

Huber, Peter J (1981). *Robust statistics*. John Wiley, New-York.

Kahan, James P, Amnon Rapoport, and Lyle V Jones (1967). "Decision making in a sequential search task". In: *Perception & Psychophysics* 2.8, pp. 374–376.

Lee, Michael D (2006). "A Hierarchical Bayesian Model of Human Decision-Making on an Optimal Stopping Problem". In: *Cognitive Science* 30.3, pp. 1–26.

Levy, Haim (2015). *Stochastic dominance: Investment decision making under uncertainty*. Springer.

Lippman, Steven A and John McCall (1976). "The economics of job search: A survey". In: *Economic inquiry* 14.2, pp. 155–189.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006). "Ambiguity aversion, robustness, and the variational representation of preferences". In: *Econometrica* 74.6, pp. 1447–1498.

Peskir, Goran and Albert Shiryaev (2006). *Optimal stopping and free-boundary problems*. Springer.

Porteus, Evan L (2002). *Foundations of stochastic inventory theory*. Stanford University Press.

Research, Wolfram (2015). *Mathematica 10.3*. Champaign, Illinios.

Riedel, Frank (2004). "Dynamic coherent risk measures". In: *Stochastic processes and their applications* 112.2, pp. 185–200.

Riedel, Frank (2009). "Optimal stopping with multiple priors". In: *Econometrica* 77.3, pp. 857–908.

Samuels, SM (1982). "Exact solutions for the full information best choice problem, Purdue Univ". In: *Stat. Dept. Mimea Series*, pp. 82–17.

Chapter 3

Locally Constant Model Uncertainty Risk Measure

Abstract

This paper introduces a (coherent) risk measure that describes the uncertainty of the model (represented by a probability measure P_0) by a set $\mathcal{P}_\lambda$ of probability measures each of which has a Radon-Nikodym's derivative (with respect to P_0) that lies within the interval $[\lambda, \frac{1}{\lambda}]$ for some constant $\lambda \in (0, 1]$. Economic considerations are discussed and an explicit representation is obtained that gives a connection to both the expected loss of the financial position and its *average value-at-risk*. Optimal portfolio analysis is performed – different optimization criteria lead to Merton portfolio. Comparison with related problems reveals examples of extreme sensitivity of optimal portfolios to model parameters and the choice of risk measure.

3.1 Introduction

The theory of (coherent) risk measures[1] allows one to describe the risk of a financial position in monetary terms: the value of a risk measure of a certain position is the amount of numeraire that needs to be added to the position to make it safe. As the future value of a financial position is not deterministic it is classically modelled by a random variable on a probability space that is assumed to be given, the implicit assumption being that one is somehow able to deduce the "correct" probability measure that drives the prices of the underlying assets. In practice, there is always going to be some model uncertainty – one can never be sure that the measure in use is the one that really drives the world.

Ideally, a risk measure should "take into account" both the model uncertainty and the "genuine" uncertainty (due to the randomness of the world). Arguably, coherent risk measures achieve just that: the well known result on robust representations of coherent risk measures proves that each coherent risk measure ρ can be completely characterized by a set $\mathcal{P}$ of probability measures[2]. The characterization allows the (monetary) risk of any financial position X to be calculated as the maximal expected loss of the said position with respect to measures that belong to $\mathcal{P}$:

$$\rho(X) = \max_{P \in \mathcal{P}} E^P[-X]. \tag{3.1}$$

Unsurprisingly, sets of measures that represent many of the well known coherent risk measures have been characterized explicitly. It is however worth noting that coherent risk measures are usually not defined via the set of probability measures that represents them, but rather via an explicit expression

[1] The literature begins with Artzner et al. (1999). A serious introduction to coherent and convex risk measures can be found in Föllmer and Schied (2011). Although we will point out some important results, for further details we refer the reader to that text and the references offered therein.

[2] Generally, the probability measures in the set $\mathcal{P}$ are finitely additive, but under reasonable technical assumptions they are sigma-additive. For details see chapters 4.2 and 4.3 in Föllmer and Schied (2011).

that is somehow economically motivated.

In this paper we take the opposite approach: we introduce a coherent risk measure, the *Locally Constant Model Uncertainty* ($LCMU_\lambda$), via the set of measures that represents it. We will impose conditions on the set of priors so that the risk measure $LCMU_\lambda$ describes the uncertainty of the model (described by a given probability measure P_0) in a way that is, in a sense that will soon become clear, locally constant and quantified by the constant $\lambda \in (0, 1]$[3].

In order to formalize the idea of locally constant model uncertainty we will use ideas that are closely related to *ambiguity theory*, and, in particular, *maxmin expected utility theory*[4]. We assume that we are given a probability measure P_0 on the event space that, for now, we assume is a subset of the real line. One might assume that the measure describes the randomness of the world well, but not ideally. More precisely, we assume that for each "small" interval/event $[a, b]$ its probability p prescribed by the measure P_0 could be wrong, but still a good approximation – the "true probability" of the event lies within an interval that contains p. The first approach would be to consider the interval $(p - \varepsilon, p + \varepsilon)$; this resembles the ε-contamination model of Maccheroni et al. (2006) in ambiguity theory. Possible reservations when it comes to this approach would be that we are immediately limited to situations where $\varepsilon < 1$. Arguably even more important than that, when one considers an event's probability one does not necessarily think in terms of whether something is more or less likely for a certain amount of percents; it may be more natural to think in terms of how many more (or less) *times* something is more (or less) likely to happen. For example, for a quite bad model, one may decide that the prescribed probability p could be wrong in either direction: it could be up to twice as likely, or up to two times

[3]This approach somewhat resembles the definition of a risk measure known as the *superhedging price*, but the choice of the set of measures that appear in its robust characterization is completely different, both formally and in motivation.

[4]The theory was introduced in the seminal paper Gilboa and Schmeidler (1989). For a recent review of ambiguity theory and the place of maxmin expected utility theory within it see Gilboa and Marinacci (2016).

overestimated.

We choose to describe the model uncertainty by the interval $[p\lambda, p/\lambda]$ for some constant $\lambda \in (0, 1]$. One can think of λ as the *model uncertainty level*: the greater the value, the lower the model uncertainty. In order to be able to use this idea more generally, we would have to consider "infinitesimaly small events". This is the reason why we introduce the set $\mathcal{P}_\lambda$, that defines the risk measure $LCMU_\lambda$, via Radon-Nikodym derivatives: it will contain all the measures P such that $\lambda \leq dP/dP_0 \leq 1/\lambda$.

One of the main results of this paper is a representation theorem: the $LCMU_\lambda$ of a financial position can be represented as a convex combination of its expected future loss (with respect to the given measure P_0) and its *average value-at-risk* calculated at an appropriately chosen level.

Value-at-risk at level λ (VaR_λ) of a financial position is simply a negative value of its λ-quantile; it is a risk measure that is not coherent and has several undesirable properties. Average value-at-risk at level λ ($AVaR_\lambda$) is an average of all the values of value-at-risk at levels between zero and λ. It is a coherent risk measure with technical and economic properties superior to VaR. Given the difference in motivations for introducing $LCMU$ and $AVaR$ it is quite curious that there is a deeper connection between the two measures. The connection is due to the resemblance of the set $\mathcal{P}_\lambda$ to the set that appears in the the representation of ($AVaR_\lambda$)[5]. An agent estimating their risk using $LCMU$ ends up with an estimation that is, in a very precise sense, a mixture of estimations of a risk-neutral agent and an agent utilizing $AVaR$.

Once the $LCMU_\lambda$ risk measure is introduced, optimal portfolio analysis is performed. We consider a continuous-time frictionless financial market with a numeraire the value of which evolves deterministically, and several risky assets. Risky assets are assumed to be a "time dependent version" of geometrical Brownian motion: the drift coefficient and the diffusion matrix are not constants, but rather deterministic functions. This admittedly simple

[5]See, for example, chapter 4.4 in Föllmer and Schied (2011) and equation (3.9) in the appendix.

model has already been studied in the context of risk measures[6].

Problems of choosing the optimal portfolio that minimizes risk (possibly under constraints) or maximizes expected reward under a risk constraint have been solved for $AVaR$. In Gambrah and Pirvu (2014) it was proven that it is optimal to distribute ones wealth between a numeraire and what is essentially a Merton portfolio (Merton (1969)). We prove the same result for $LCMU$, and give an example where the optimal portfolios for $AVaR$ and $LCMU$ coincide.

We also analyze a surprising example where optimizing with respect to the two risk measures leads to completely different optimal portfolios: optimizing with respect to $AVaR$ leads to a portfolio without risky assets, and optimizing with respect to $LCMU$ gives a portfolio with only risky assets! We offer both technical and theoretical explanation as to why this happens[7].

In the next section we formally introduce the risk measure $LCMU_\lambda$ and provide its representation that connects it to $AVaR_\lambda$. We also identify the minimizing measure for each financial position. We conclude the section with some numerical examples and simple comparisons between the two measures. In the third section we introduce the model of the financial market, formulate the results on optimal portfolios and perform the sensitivity analysis. A review of relevant facts about coherent risk measures, a corollary of a generalized version of the Neyman-Pearson lemma, proofs of theorems and additional relevant details are in the appendices.

[6]See Gambrah and Pirvu (2014) and references offered therein.
[7]See the discussion after theorem 3.3.1.

3.2 Representation of the LCAN Risk Measure

3.2.1 Definition

Let $(\Omega, \mathcal{F}, P_0)$ be a given probability space, where P_0 is a probability measure support of which is the whole set Ω. We denote the set of probability measures defined on $(\Omega, \mathcal{F})$ with $\mathcal{M}$. For any $P \in \mathcal{M}$ and a random variable $X \in L^\infty(\Omega, \mathcal{F}, P_0)$ we denote the expectation of the random variable X with respect to probability measure P with $E^P[X]$, and, particularly, we write $E[X]$ for $E^{P_0}[X]$. We define a set of probability measures using Radon-Nikodym derivatives:

$$\mathcal{P}_\lambda = \left\{ P \in \mathcal{M} \,\middle|\, 0 < \lambda \leq \frac{dP}{dP_0} \leq \frac{1}{\lambda} \right\}, \tag{3.2}$$

where $\lambda \in (0, 1]$ is a given constant. As was mentioned earlier, one can think of λ as the level of model uncertainty: the closer the value of λ is to one there is less model uncertainty, i.e. we have greater confidence that the model is "good". Note that, due to the definition of the set P_λ, all the measures in $\mathcal{P}_\lambda$ are equivalent to P_0.

We can now define the *locally constant model uncertainty* risk measure:

$$LCMU_\lambda(X) = \sup_{P \in \mathcal{P}_\lambda} E^P[-X]. \tag{3.3}$$

Clearly[8], this is a coherent risk measure.

[8] Due to representation theorems for coherent risk measures; see (3.1).

3.2.2 Connection with Avarege Value-at-Risk

In this subsection we will relate the coherent measure $LCMU$ to the well known *average value-at-risk* ($AVaR$), also known as *expected shortfall*:

$$AVaR_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda VaR_t(X)\,dt = -\frac{1}{\lambda}\int_0^\lambda q_X^+(t)\,dt$$

where $q_X^+(t) = \inf\{x \mid P_0(X \leq x) > t\} = -VaR_t(X)$ is the upper quantile function of the random variable X with respect to the measure P_0 that appears in the definition of value-at-risk.

Relevant additional details on coherent measures of risk, and $AVaR$ in particular, are available in appendix 3.A. We are now ready to formulate one of the main results of this paper.

Theorem 3.2.1. *The coherent risk measure $LCMU_\lambda$ allows the following representation:*

$$\mathrm{LCMU}_\lambda(X) = \lambda E[-X] + (1-\lambda)AVaR_{\frac{\lambda}{1+\lambda}}(X). \tag{3.4}$$

The proof of the theorem is along the lines of the proof of robust representation for $AVaR$[9]. It relies on using the generalized version of the well known Neyman-Pearson lemma which we reformulate to fit our context in appendix 3.B. The proof of the theorem can be found in appendix 3.C.

We note that if the distribution of the random variable X has density then $LCMU_\lambda(X)$ can be written as follows:

$$LCMU_\lambda(X) = -E\left[\lambda E[X] + (1-\lambda)X \;\middle|\; X < VaR_{\frac{\lambda}{1+\lambda}}(X)\right];$$

this is due to the representation of $AVaR_\lambda$ for random variables with density (see equation (3.9) in appendix 3.A). Hence, an agent estimating the risk of a financial position using $LCMU$ calculates an expectation of a mixture of the position and its expected value, conditioned on the fact that there will

[9] See also theorem 4.47 in Föllmer and Schied (2011).

be losses. The value obtained is the amount of numeraire that makes the position safe.

3.2.3 Maximizing Measure

Careful reading of the proof of the theorem 3.2.1 shows that the supremum in the definition (3.3) of $LCMU_\lambda$ is attained. In particular, if q is a $\frac{\lambda}{1+\lambda}$-quantile of X with respect to P_0 and

$$\psi_X = \mathbb{1}_{\{X<q\}} + k\mathbb{1}_{\{X=q\}},$$

one can write:

$$LCMU_\lambda(X) = \int -X \left(\lambda + (1-\lambda)\frac{1+\lambda}{\lambda}\psi_X\right) dP_0.$$

The measure Q_X defined via its Radon-Nikodym's derivative:

$$\frac{dQ_X}{dP_0} = \lambda + (1-\lambda)\frac{1+\lambda}{\lambda}\psi_X \tag{3.5}$$

belongs to the set $\mathcal{P}_\lambda$ and is the maximizing measure in (3.3); we record this fact in the following theorem:

Theorem 3.2.2. *For any random variable $X \in L^\infty(\Omega, \mathcal{F}, P_0)$, the measure $Q_X \in \mathcal{P}_\lambda$ as defined in (3.5) is the maximizing measure in the defintion (3.3) of the risk measure $LCMU_\lambda$, i.e.:*

$$\mathrm{LCMU}_\lambda(X) = E^{Q_X}[-X].$$

The proof follows from the preceding theorem and we comment on it briefly in appendix 3.C.

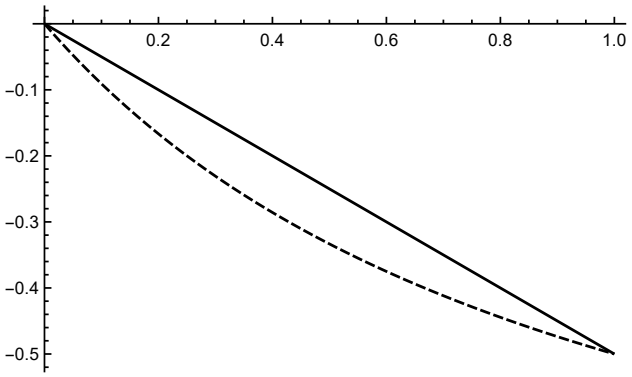

Figure 3.1: Case $X \sim \mathcal{U}([0,1])$. The full line represents the function $AVaR_\lambda(X)$, while the dashed line represents $LCMU_\lambda(X)$ (as functions of λ)

.

3.2.4 Comparison with Average Value at Risk

Once the connection between $AVaR$ and $LCMU$ have been established it is worthwhile to explore (numerical) similarities and differences between the two measures. First, one easily notes that both measures satisfy:

$$LCMU_1[X] = AVaR_1[X] = E[-X],$$
$$\lim_{\lambda \to 0+} LCMU_\lambda(X) = \lim_{\lambda \to 0+} AVaR_\lambda(X) = \operatorname{ess\,sup} -X.$$

Furthermore, due to the fact that the set $\mathcal{P}_\lambda$ that represents $LCMU_\lambda$ is clearly a subset of the set that gives the robust representation of $AVaR_\lambda$ (see equation (3.9)) the inequality $AVaR_\lambda(X) \geq LCMU_\lambda(X)$ holds for any financial position X. This means that, from a regulatory point of view, $LCMU$ is the less conservative of the two measures.

To get a clearer insight into the way that the different risk measures value risk differently we will focus on two simple examples with positions distributed uniformly and log-normally.

Uniform distribution

Suppose a random variable X is uniformly distributed on the interval $[a, b]$. Straightforward computations yield:

$$VaR_\lambda(X) = AVaR_\lambda(X) = -a - \frac{\lambda}{2}(b-a)$$
$$LCMU_\lambda(X) = -a - \frac{\lambda}{1+\lambda}(b-a).$$

Figure 3.1 contains the graphs of $AVaR_\lambda(X)$ and $LCMU_\lambda(X)$ as functions of λ. The figure confirms that, indeed, $LCMU$ prescribes substantially lower values of numeraire than AVaR.

Log-normal distribution

Suppose now that the random variable $X \sim \ln\mathcal{N}(\mu, \sigma^2)$ is log-normally distributed. After some computations one can see that:

$$VaR_\lambda(X) = -\exp(\mu + \sigma\Phi^{-1}(\lambda)),$$
$$AVaR_\lambda(X) = -\frac{1}{\lambda}\int_0^\lambda \exp(\mu + \sigma\Phi^{-1}(t))\,dt = -\exp\left(\mu + \frac{\sigma^2}{2}\right)\frac{\Phi(\Phi^{-1}(\lambda) - \sigma)}{\lambda}$$
$$\begin{aligned} LCMU_\lambda(X) &= -\lambda\exp\left(\mu + \frac{\sigma^2}{2}\right) - (1-\lambda)\frac{(\lambda+1)}{\lambda}\int_0^{\frac{\lambda}{\lambda+1}} \exp(\mu + \sigma\Phi^{-1}(t))\,dt \\ &= -\exp\left(\mu + \frac{\sigma^2}{2}\right)\left(\lambda + \frac{1-\lambda^2}{\lambda}\Phi\left(\Phi^{-1}\left(\tfrac{\lambda+1}{\lambda}\right) - \sigma\right)\right). \end{aligned}$$

The second equality for $AVaR$ is the only one that is slightly more involved; we prove it in appendix 3.D.

Figure 3.2 contains the graphs of $AVaR_\lambda(X)$ and $LCMU_\lambda(X)$ as functions of λ. As can be seen, the less conservative $LCMU$ can prescribe substantially lower values.

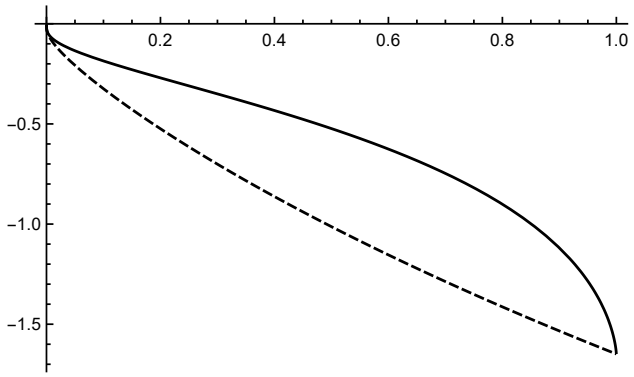

Figure 3.2: Case $X \sim \ln \mathcal{N}(0,1)$. The full line represents the function $AVaR_\lambda(X)$, while the dashed line represents $LCMU_\lambda(X)$ (as functions of λ)

.

3.3 Optimal Portfolio Analysis

3.3.1 Model

Let $(\Omega, \{\mathcal{F}_t\}_{0\le t\le T}, \mathcal{F}, P)$ be a filtered probability space which accommodates a standard m-dimensional Brownian motion $W(t) = (W^j(t))_{j=1,\dots,m}$. We consider a financial market with a numeraire S_0 and m risky assets S_i which are traded continuously over a finite time horizon $[0,T]$ in a frictionless market. The dynamics of the assets are:

$$\begin{aligned} dS_0(t) &= r(t)\,dt, \\ dS_i(t) &= S_i(t)\left(b_i(t)\,dt + \sum_{j=1}^{m} \sigma_{ij}(t)\,dW^j(t)\right), \quad i = 1,\dots,m, \end{aligned}$$

where $r(t)$ is the deterministic interest rate, the functions $b_i(t)$ are deterministic and denote the drift of the stock, and the volatility matrix $\sigma(t) = (\sigma_{ij}(t))_{i,j=1,\dots,m}$ is deterministic and invertible. We assume that functions r, b_i and σ_{ij} are square integrable and that the inequalities $0 < r(t) < b(t)$ are satisfied for each t.

Self financing strategies are described by a deterministic vector $\pi(t) =$

$(\pi_1(t), \ldots, \pi_m(t)) \in \mathbb{R}^m$ such that

$$\sum_{i=1}^{m} \pi_i(t) \leq 1, \text{ and } \pi_i(t) \geq 0, \quad i = 1, \ldots, m. \tag{3.6}$$

An agent following the strategy π invests a fraction π_i of their wealth in the risky stock S_i, while the remainder is invested in the bond (represented by the numeraire S_0). As can be seen, no borrowing or short selling is allowed. Hence, if we denote the wealth at time t by $X^\pi(t)$ and the number of shares of the asset i held in portfolio by $N_i(t)$, we have

$$\pi_i(t) = N_i(t)S_i(t)/X^\pi(t),\ i \geq 1, \text{ and } X^\pi(t) = \sum_{i=0}^{m} N_i(t)S_i(t).$$

Dynamics of S_i imply that the agent's wealth satisfies:

$$dX^\pi(t) = X^\pi(t)\left((r(t) + B(t)'\pi(t))\,dt + \sigma(t)'\pi(t)\,dW(t)\right),$$

where $B(t) = (b_1(t) - r(t), \ldots, b_m(t) - r(t))$ and $'$ is the transposition operator.

Using Ito's lemma, direct calculations yield:

$$X^\pi(T) = X^\pi(0)\exp\left(\int_0^T r(s) + B(s)'\pi(s) - \frac{1}{2}||\sigma(s)'\pi(s)||^2\,ds + \int_0^T ||\sigma(s)'\pi(s)||\,dW(s)\right).$$

If we introduce the following notation:

$$R = \exp\left(\int_0^T r(s)\,ds\right),\ x_\pi = X^\pi(0),$$
$$\mu(\pi) = \int_0^T B(s)'\pi(s)\,ds,\ \psi(\pi) = \int_0^T ||\sigma(s)'\pi(s)||^2\,ds$$

then

$$E[X^\pi(T)] = x_\pi R\exp\left(\mu(\pi)\right). \tag{3.7}$$

3.3.2 Loss and Risk Measures

We define loss as $L(\pi) = X^{\pi}(T) - X^{\pi}(0)$; it is simply a difference between the wealth at the end and at the beginning of the time period. This is the quantity that will be involved in different optimization problems that we solve. In particular, we will consider the quantity $LCMU_{\lambda}(L(\pi))$ and, for comparison purposes, $AVaR_{\lambda}(L(\pi))$. Considering a risk measure of a random variable that depends only on the final and, possibly, initial value of the stochastic process is standard in literature (see Schied and Wu (2005)). It is also in the spirit of the classical stochastic control problem in financial mathematics – *Merton's portfolio problem* (Merton (1969)) where the utility of the terminal wealth is considered.

We note that, although the dynamics of the process in question are acknowledged, this approach can be considered static, as we only consider two points in time and do not impose constraints on the financial positions in between the two time endpoints. An alternative would be to consider dynamic versions of risk measures; this is, on a technical level, significantly more involved. Reasons for the complications include having to do with the time consistency of dynamic risk measures and the non-time consistency of AVaR (Cheridito and Stadje (2009)). Considering only deterministic (instead of predictable) trading strategies, as we do here, somewhat offsets the need for dynamic measures as the agent effectively makes a decision about trading throughout the whole period. In any case, when analyzing the results of models that only involve the final time point, one should be aware of the limitations of models of this kind and therefore careful in the interpretations.

It can be shown[10] that:

$$AVaR_{\lambda}(L(\pi)) = x_{\pi}\left(1 - \frac{R}{\lambda}\Phi\left(\Phi^{-1}(\lambda) - \sqrt{\psi(\pi)}\right)\exp(\mu(\pi))\right).$$

Combining the expressions for $E[X^{\pi}(T)]$ and $AVaR_{\lambda}(L(\pi))$, and using the

[10]The derivation is quite similar to the derivation in appendix 3.D for $AVaR_{\lambda}(X)$ for a log-normally distributed position X. For more details see the proof of proposition 3.1.2.1 in Gambrah and Pirvu (2014).

representation of $LCMU$ from theorem 3.2.1, we obtain:

$$LCMU_\lambda(L(\pi)) = \\ x_\pi \left(1 - R\left(\lambda + \tfrac{1-\lambda^2}{\lambda}\Phi\left(\Phi^{-1}\left(\tfrac{\lambda}{1+\lambda}\right) - \sqrt{\psi(\pi)}\right)\right)\right)\exp(\mu(\pi)).$$

3.3.3 Optimization Problems and Merton portfolio

Let $\mathcal{Q}$ be the set of all the trading strategies π which are Borel measurable, deterministic and satisfy the conditions of equation (3.6). We will consider three problems that lead to different optimal portfolios in $\mathcal{Q}$.

First we consider the unconstrained problem of choosing the portfolio for which the risk measure $LCMU$ prescribes the lowest risk:

$$(P1) \qquad \min_{\pi\in\mathcal{Q}} LCMU_\lambda(L(\pi)).$$

The second problem we consider is finding the lowest risk portfolio among all the portfolios with fixed expected return:

$$(P2) \qquad \min_{\pi\in\mathcal{Q}} LCMU_\lambda(L(\pi)) \text{ such that } E[X^\pi(T)] = M.$$

Finally, we consider the problem of maximizing the expected returns while requiring the risk to be above some positive boundary C:

$$(P3) \qquad \max_{\pi\in\mathcal{Q}} E[X^\pi(T)] \text{ such that } LCMU_\lambda(L(\pi)) \geq C.$$

All three problems have been explicitly solved for risk measures VaR and $AVaR$ in Gambrah and Pirvu (2014). Optimal portfolios for both risk measures and for all three problems are closely related to the trading strategy:

$$\pi_M(t) = (\sigma(t)\sigma(t)')^{-1}B(t); \tag{3.8}$$

in each case the optimal portfolio is just a multiple of π_M, the well known

Merton portfolio from Merton (1969) and numerous related problems[11]. Furthermore, due to the strength of the constraint in the problem (P2), the optimal portfolios for both risk measures coincide for that problem. This leads to an interpretation similar to the well known *mutual fund theorem*: if there is a hedge fund with portfolio π_M it is optimal for the agent to distribute their wealth between the hedge fund and the bonds no matter what the optimization criterion is. However, different optimization criteria can lead to different proportions of investments between the hedge fund and the bond.

It turns out that the solutions for problems (P1-P3) are also multiples of π_M.

Theorem 3.3.1. *For each of the problems (P1), (P2) and (P3) there are constants c_1, c_2, c_3 such that the solutions to the problems are:*

$$\pi_1^* = c_1\pi_M, \quad \pi_2^* = c_2\pi_M, \quad \pi_3^* = c_3\pi_M.$$

Furthermore, the same portfolio solves the, appropriately reformulated, optimization problem (P2) for the risk measures *LCMU*, *AVaR*, and *VaR*; see theorem 3.E.1 in appendix 3.E.

The fact that optimal portfolios when one optimizes with respect to risk measures (as the theorems 3.3.1 and 3.E.1 show) and with respect to utility functions (as the classical literature Merton (1969) shows) is in some ways surprising. We offer some comments that explain why this is the case in this model, but we also comment on the modelling approach to optimal portfolios in general.

On a technical level this result is driven by strong assumptions: log-normally distributed returns or risky assets, deterministic trading strategies, a frictionless market, and constraints on borrowing and short selling. In a sense, if the market conditions are close to ideal then the conclusions of the classical theory remain valid.

[11] See, for example, Rogers (2013).

However, the theory of risk measures was developed because, among other things, the markets are not ideal: the returns of investments are not distributed log-normally and tails of the "actual distributions" driving the world are heavy. Thus, while the analysis offered in Gambrah and Pirvu (2014) and in our work sheds valuable insight into the optimal portfolio choice with respect to risk measures, it could be considered a mere first step in optimization problems of this kind and further investigation into more robust and realistic models is needed.

The derivations in the proof of theorem 3.3.1 are closely related to the ones offered in Gambrah and Pirvu (2014); details are in Appendix 3.E.

3.3.4 Sensitivity of Optimal Portfolios to the Choice of Risk Measures

Let us consider problems (P1-3) for the risk measures $AVaR_\lambda$ and $LCMU_\lambda$. We have already seen that the solutions of the optimization problems for both risk measures belong to the same class. In this subsection we further explore how does the choice of risk measure influence the optimal allocation between the numeraire and Merton porfolio π_M.

We begin by analyzing optimal portfolios for problem (P1). Due to the close connection between the risk measures $AVaR$ and $LCMU$ (as established by theorem 3.2.1) and similarity of results in theorems 3.3.1 and 3.E.1 it would be expected that the optimal portfolios when optimizing with respect to the two measures behave similarly. Somewhat surprisingly, this is not the case, as we demonstrate below.

If we solve the problem (P1) for $AVaR_\lambda$ and $LCMU_\lambda$ the optimal portfolios are:

$$AVaR_\lambda : \pi_A = c_A \pi_M, \quad LCMU_\lambda : \pi_L = c_L \pi_M,$$

for some constants c_A and c_L; see theorem 3.E.1. The analysis in subsection 3.2.4 shows that $LCMU_\lambda(L(\pi) \geq AVaR_\lambda(L(\pi))$. This implies that $c_A \leq c_B$: constants c_A and c_L determine the amount of numeraire to be kept in the

optimal portfolio, hence $AVaR$, being the less risky of the two risk measures, prescribes less risky assets in the optimal portfolio and more numeraire.

The proof reveals how the constants c_A and c_L are calculated. Let us introduce functions:

$$g_\lambda(\varepsilon) := G^A_\lambda(\Theta\varepsilon, \varepsilon^2) = \Theta\varepsilon + \ln(\varphi_\lambda(\varepsilon^2)),$$
$$f_\lambda(\varepsilon) := G^L_\lambda(\Theta\varepsilon, \varepsilon^2) = \Theta\varepsilon + \ln(\lambda + (1-\lambda)\varphi_{\frac{\lambda}{1+\lambda}}(\varepsilon^2)),$$

where

$$\varphi_\lambda(y) = \frac{1}{\lambda}\Phi(\Phi^{-1}(\lambda) - \sqrt{y}) \text{ and } \Theta = \sqrt{\int_0^T ||\sigma(s)^{-1}B(s)||^2\, ds}.$$

Let ε_A and ε_L be the solutions of optimization problems

$$(\varepsilon P) \qquad \max_{\varepsilon\in I} g_\lambda(\varepsilon) \text{ and } \max_{\varepsilon\in I} f_\lambda(\varepsilon),$$

where

$$I = \left[0, \int_0^T ||\sigma(s)||^2\, ds\right].$$

Then $c_A = \varepsilon_A/\Theta$ and $c_L = \varepsilon_L/\Theta$.

For different values of the parameters the solutions c_A and c_L can be on the boundaries of the interval I. In general, the equality $c_A = c_L$ does not hold. Furthermore, there are examples where $c_A = 0$ and $c_L = I$. This means that for certain reasonable values of the parameters of the model it can happen that optimal portfolios with respect to closely related risk measures $AVaR$ and $LCMU$ are completely different: it is optimal with respect to $AVaR$ to not invest in the risky assets, while with respect to $LCMU$ it is optimal to invest *only* in risky assets!

To illustrate that this can indeed be the case we will consider a simple special case of the model we introduced: a market with one risky asset in which the risk rate, the drift coefficient and the diffusion coefficient are all constant. For simplicity we also assume that the time horizon satisfies $T = 1$.

Even in this simplified setting, solving optimization problems (εP) is technically cumbersome. We will avoid the complications by making appropriate approximations.

Direct calculations show that under the simplified assumptions the interval I becomes $[0, \sigma^2]$. Thus for $\sigma < 1$ the interval I becomes "small". Once one notices that $f_\lambda(0) = g_\lambda(0) = 0$, we can approximate the functions f_λ and g_λ with their tangents at 0:

$$f_\lambda(\varepsilon) \approx \varepsilon f'_\lambda(0), \qquad g_\lambda(\varepsilon) \approx \varepsilon g'_\lambda(0).$$

As we are demonstrating that the solutions of problems (εP) are on the boundary of the interval I these approximations will suffice. Indeed, it is sufficient to establish "opposite" monotonicities of the functions f_λ and g_λ on the interval I.

Direct calculations show:

$$\begin{aligned}
\varphi'_\lambda(\varepsilon) &= -\frac{1}{\lambda}\Phi'(\Phi(\lambda) - \varepsilon), \\
g'_\lambda(\varepsilon) &= \Theta + \frac{\varphi'_\lambda(\varepsilon)}{\varphi_\lambda(\varepsilon)}, \\
f'_\lambda(\varepsilon) &= \Theta + \frac{(1-\lambda)\varphi'_{\lambda/(1+\lambda)}(\varepsilon)}{\lambda + (1-\lambda)\varphi_{\lambda/(1+\lambda)}(\varepsilon)}.
\end{aligned}$$

It follows that:

$$g'_\lambda(0) = \Theta + \varphi'_\lambda(0) \text{ and } f'_\lambda(0) = \Theta + (1-\lambda)\varphi'_{\lambda/(1+\lambda)}(0).$$

For example, if we choose $\lambda = 0.2$, $\mu - r = 0.4$ and $\sigma = 0.32$ then $\theta = 1.25$, $I = [0, 1.5625]$ and $g'_\lambda(0) < 0$ and $f'_\lambda(0) > 0$. The approximations we introduced are good enough; see figure 3.3. Indeed, $f_\lambda(\varepsilon)$ achieves its maximum on the right hand side of the interval, and $g_\lambda(\varepsilon)$ achieves its maximum on the left hand side of the interval I; this implies $c_A = 0$ and $c_L = 1.5625$.

Furthermore, figure 3.4 shows that the difference $f'_\lambda(0) - g'\lambda(0)$ (as a function of λ) is always positive. Thus, for any value of λ we can always

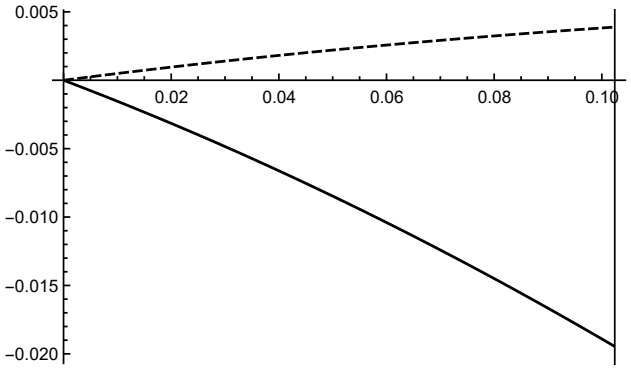

Figure 3.3: Graphs of functions $f_{0.2}(\varepsilon)$ (dashed) and $g_{0.2}(\varepsilon)$ (full) for $\varepsilon \in I$ and $\mu - r = 0.4$, $\sigma = 0.32$. The vertical line denotes the end of interval I.

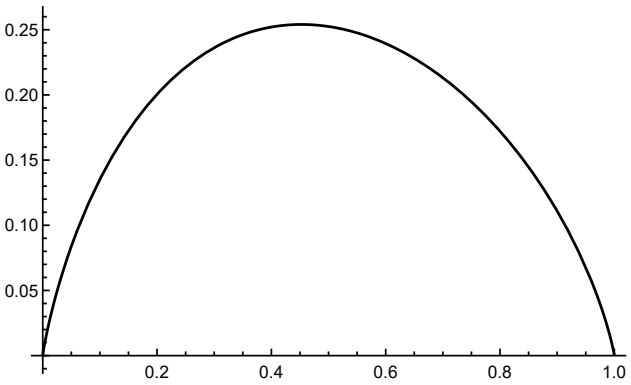

Figure 3.4: Graph of the function $f'_\lambda(0) - g'_\lambda(0)$ (as a function of λ).

choose a value of Θ such that $g'_\lambda(0) < 0$ and $f'_\lambda(0) > 0$. This can be achieved by choosing the appropriate values of μ, r, and σ such that $\Theta \in [g'_\lambda(0), f'_\lambda(0)]$, and that interval I is "small enough".

We conclude this section by briefly turning to the problems (P2) and (P3).

We note that optimal portfolios with respect to both measures coincide for the problem (P2) (see the solution of (ρP2) in the proof of theorem 3.E.1 in appendix 3.E).

As for the problem (P3), the situation is quite similar to sensitivity analysis performed for the problem (P1): there are situations in which optimizing with respect to different measures prescribes radically different optimal be-

havior. This is due to the similarities of problems (εP) and the optimization problem that the problems (P3) and (ρP3) are reduced to; see the part of the proof of theorem 3.E.1 related to the problems (P3) and (ρP3).

3.4 Conclusion

Motivated by ideas from ambiguity theory we have introduced a new coherent risk measure: locally constant model uncertainty (LCMU). It is explicitly defined via its set of probability measures in a way that makes uncertainty about the probabilities of "small" events constant – the Radon-Nikodym derivative lies within a fixed interval.

We have derived a representation of $LCMU$ as a convex combination of the expected loss of the position and its average value-at-risk (AVaR) calculated at an appropriately chosen interval. We have thus demonstrated a viable connection between ambiguity theory and well established risk measures.

We have considered and solved optimal investment problems in continuous time related to $LCMU$ in a frictionless market with m-assets that evolve following a time dependent version of the multi-dimensional geometric Brownian motion with no-borrowing and no-short-selling constraints. We have proven a version of a *mutual fund theorem*: choosing portfolios that minimize risk or maximize profit with a risk constraint both lead to Merton portfolios; this result was already known for value-at-risk and $AVaR$ in this setting.

We have demonstrated that optimal portfolios can be radically different when optimizing with respect to $LCMU$ and AVaR. This surprising conclusion raises questions about dynamic models of optimal investment in continuous time that deal with risk measures. Our results also demonstrated the fragility of the solutions of optimization problems involving risk measures in dynamic settings, even in mathematically simple contexts.

Appendix 3.A Risk Measures

For convenience we collect the relevant definitions and results from the theory of risk measures that were used above; for details see chapter 4 and the appendix in Föllmer and Schied (2011).

A coherent risk measure is a functional $\rho : L^\infty(\Omega, \mathcal{F}, P_0) \to \mathbb{R}$ satisfying:

1. Monotonicity: $X \leq Y \to \rho(X) \geq \rho(Y)$,
2. Cash invariance: $\rho(X + c) = \rho(X) - c$, $c \in \mathbb{R}$
3. Subaditivity $\rho(X + Y) \leq \rho(X) + \rho(Y)$
4. Positive homogeniety: $\rho(\alpha X) = \alpha\rho(X)$, $\alpha > 0$.

We denote by $\mathcal{M}$ the set of all probability measures on $(\Omega, \mathcal{F})$. A robust representation theorem of coherent risk measures states that for every coherent risk measure ρ there exists a set of (possibly finitely additive) probability measures $\mathcal{P} \subset \mathcal{M}$ equivalent to P_0 such that

$$\rho(X) = \inf_{P \in \mathcal{P}} E[-X].$$

Under suitable technical conditions that can be formulated in terms of certain continuity properties of the functional ρ the set $\mathcal{P}$ can be concentrated on probability measures[12]. Similar results can be obtained for convex risk measures.

One of the best known examples of coherent measures of risk is *average value-at-risk*:

$$AVaR_\lambda(X) = -\frac{1}{\lambda}\int_0^\lambda q_X^+(t)\, dt$$

where $q_X^+(t) = \inf\{x \mid P(X \leq x) > t\}$ is the upper quantile function. It is closely related to the better known *value-at-risk*, but has superior theoretical properties: AVaR takes into account losses of all sizes and encourages

[12] For further details see chapters 4.2 and 4.3 in Föllmer and Schied (2011).

diversification, among others. It allows the following representations:

$$AVaR_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda VaR_t(X)\,dt = \max\left\{E^P[-X]\,\middle|\, Q \sim P_0, \frac{dQ}{dP_0} \leq \frac{1}{\lambda}\right\}$$

Furthermore, if the distribution of X is atomless then one can write:

$$AVaR_\lambda(X) = E[-X|X < VaR_\lambda(X)]. \tag{3.9}$$

Appendix 3.B Corollary of the Generalized Version of Neyman-Pearson Lemma

Lemma that follows is a direct corollary of the generalized Neyman-Pearson Lemma as formulated in theorem A.30 in Föllmer and Schied (2011).

Lemma 3.B.1. *If P and Q are given equivalent measures and $\alpha \in [0,1]$ is a given constant then:*

$$\max\left\{\int \psi\,dQ\,\middle|\, 0 \leq \psi \leq 1, \int \psi\,dP = \alpha\right\} = \alpha = \int \psi_X\,dQ \tag{3.10}$$

for

$$\psi_X = \mathbb{1}_{\{\frac{dQ}{dP} > c\}} + k\mathbb{1}_{\{\frac{dQ}{dP} = c\}} \tag{3.11}$$

where c is a $1-\alpha$-quantile of $\frac{dQ}{dP}$ with respect to P and

$$k = \begin{cases} 0, & P\left(\frac{dQ}{dP} = c\right) = 0 \\ \dfrac{\alpha - P\left(\frac{dQ}{dP} > c\right)}{P\left(\frac{dQ}{dP} = c\right)}, & P\left(\frac{dQ}{dP} = c\right) > 0. \end{cases} \tag{3.12}$$

Appendix 3.C Proofs of theorems 3.2.1 and 3.2.2

Proof. We begin by rewriting the left hand side of the equation (3.3) for a fixed random variable $X < 0$:

$$\begin{aligned}
\sup_{P\in\mathcal{P}_\lambda} E^P[-X] &= \sup\left\{ E^P[-X] \,\middle|\, P \in \mathcal{M}, \lambda \le \frac{dP}{dP_0} \le 1/\lambda \right\} \\
&= \sup\left\{ -X\frac{dP}{dP_0}\, dP_0 \,\middle|\, P \in \mathcal{M}, \lambda \le \frac{dP}{dP_0} \le 1/\lambda \right\} \\
&= \sup\left\{ -\int X\varphi\, dP_0 \,\middle|\, \int \varphi\, dP_0 = 1, \lambda \le \varphi \le 1/\lambda \right\} \\
&= \sup\left\{ -E[X]\int \frac{X}{E[X]}\varphi\, dP_0 \,\middle|\, \int \varphi\, dP_0 = 1, \lambda \le \varphi \le 1/\lambda \right\}.
\end{aligned}$$

So far we have only used the definitions of the set $\mathcal{P}_\lambda$ and basic properties of the expectation operator and Radon-Nikodym derivatives. We notice that $\int \frac{X}{E[X]}\, dP_0 = 1$, so the random variable $\frac{X}{E[X]} > 0$ is a Radon-Nikodym derivative for some measure Q that is equivalent to P_0. Hence, using the last expression above and the inequality $E[-X] > 0$, we have:

$$\begin{aligned}
\sup_{P\in\mathcal{P}_\lambda} E^P[-X] &= \sup\left\{ -E[X]\int \varphi\frac{dQ}{dP_0}\, dP_0 \,\middle|\, \int \varphi\, dP_0 = 1, \lambda \le \varphi \le 1/\lambda \right\} \\
&= E[-X]\sup\left\{ -E[X]\int \varphi\, dQ \,\middle|\, \int \varphi\, dP_0 = 1, \lambda \le \varphi \le 1/\lambda \right\}.
\end{aligned} \tag{3.13}$$

The following equivalence of inequalities:

$$\lambda \le \varphi \le 1/\lambda \Leftrightarrow 0 \le \frac{\lambda}{1-\lambda^2}(\varphi - \lambda) \le 1. \tag{3.14}$$

allows one to rewrite the right hand side of the equation (3.13) in terms of a new variable $\psi := \frac{\lambda}{1-\lambda^2}(\varphi - \lambda)$:

$$\sup_{P\in\mathcal{P}_\lambda} E^P[-X] =$$
$$= E[-X]\sup\left\{\int\left(\psi\tfrac{1-\lambda^2}{\lambda}+\lambda\right)\,dQ\,\middle|\,\int\left(\psi\tfrac{1-\lambda^2}{\lambda}+\lambda\right)\,dP_0 = 1, 0\le\psi\le 1\right\}$$
$$= E[-X]\left(\lambda + \tfrac{1-\lambda^2}{\lambda}\cdot\sup\left\{\int\psi\,dQ\,\middle|\,\int\psi\,dP_0 = \tfrac{\lambda}{1+\lambda}, 0\le\psi\le 1\right\}\right). \tag{3.15}$$

Applying lemma 3.B.1 one obtains:

$$\sup\left\{\int\psi\,dQ\,\middle|\,\int\psi\,dP_0 = \tfrac{\lambda}{1+\lambda}, 0\le\psi\le 1\right\} = \int\psi_X\,dQ, \tag{3.16}$$

where

$$\psi_X = \mathbb{1}_{\{\frac{dQ}{dP_0}>c\}} + k\mathbb{1}_{\{\frac{dQ}{dP_0}=c\}},$$

for

$$k = \begin{cases} 0, & P\left(\frac{dQ}{dP_0}=c\right)=0 \\ \dfrac{\alpha - P\left(\frac{dQ}{dP_0}>c\right)}{P\left(\frac{dQ}{dP_0}=c\right)}, & P\left(\frac{dQ}{dP_0}=c\right)>0, \end{cases}$$

and c a $1-\frac{\lambda}{1+\lambda}$-quantile of $\dfrac{dQ}{dP_0} = \dfrac{X}{E[X]}$ with respect to P_0. Keeping in mind that $X<0$, the inequality $\dfrac{dQ}{dP_0} > c$ can be written as:

$$\begin{aligned} X < E[X]\cdot c =& E[X]\cdot\inf\left\{t\,\middle|\,P_0\left(\frac{dQ}{dP_0}<t\right) > 1-\frac{\lambda}{1+\lambda}\right\} \\ =& \sup\left\{tE[X]\mid P_0\left(X>tE[X]\right)-1>-\frac{\lambda}{1+\lambda}\right\} \\ =& \sup\left\{t\mid 1-P_0\left(X>t\right)<\frac{\lambda}{1+\lambda}\right\} \\ =& \sup\left\{t\mid P_0\left(X\le t\right)<\frac{\lambda}{1+\lambda}\right\} =: q. \end{aligned}$$

We can see that q is a $\frac{\lambda}{1+\lambda}$-quantile of X with respect to P_0 and that inequalities $\frac{dQ}{dP_0} > c$ and $X < q$ are equivalent. Similarly, the equality $\frac{dQ}{dP_0} = c$ holds if and olny if $X = q$ holds. Hence:

$$\psi_X = \mathbb{1}_{\{X<q\}} + k\mathbb{1}_{\{X=q\}}. \tag{3.17}$$

Combining (3.15) and (3.16) one obtains:

$$\begin{aligned}
\sup_{P\in\mathcal{P}_\lambda} E^P[-X] &= E[-X]\left(\lambda + \frac{1-\lambda^2}{\lambda}\int \psi_X\, dQ\right) \\
&= E[-X]\left(\lambda + \frac{1-\lambda^2}{\lambda}\int \psi_X \frac{dQ}{dP_0}\, dP_0\right) \\
&= E[-X]\left(\lambda + \frac{1-\lambda^2}{\lambda}\int \psi_X \frac{X}{E[X]}\, dP_0\right) \\
&= \lambda E[-X] - \frac{1-\lambda^2}{\lambda}\int X\psi_X\, dP_0.
\end{aligned} \tag{3.18}$$

The integral that appears in the last expression can be rewritten (using 3.17) as follows:

$$\begin{aligned}
\int X\psi_X\, dP_0 &= \int_{\{X<q\}} X\, dP_0 + k\int_{\{X=q\}} X\, dP_0 \\
&= -\int_{\{X<q\}} (q-X)\, dP_0 + q\int_{\{X<q\}} dP_0 + k\int_{\{X=q\}} q\, dP_0 \\
&= -\int (q-X)^+\, dP_0 + q\left(P_0(X<q) + kP_0(X=q)\right).
\end{aligned} \tag{3.19}$$

If $P\left(\frac{dQ}{dP_0} = c\right) = P_0(X = q) = 0$ then clearly

$$P_0(X<q) + kP_0(X=q) = P_0(X<q) = \frac{\lambda}{1+\lambda}; \tag{3.20}$$

we used the definition of q in the last equality. If $P_0(X = q) > 0$ then, using

the definition of k, we have

$$P_0(X < q) + kP_0(X = q) = P_0(X < q) + \left(\frac{\lambda}{1+\lambda} - P_0(X < q)\right) = \frac{\lambda}{1+\lambda}. \tag{3.21}$$

Thus, in any case, combining (3.20) and (3.21) with (3.19), we obtain:

$$\int X\psi_X \, dP_0 = -\int (q - X)^+ \, dP_0 + q\frac{\lambda}{1+\lambda}.$$

Plugging this into (3.18), after some simplification, we obtain:

$$\begin{aligned}\sup_{P\in\mathcal{P}_\lambda} E^P[-X] &= \lambda E[-X] + \frac{1-\lambda^2}{\lambda}\int (q - X)^+ \, dP_0 - q(1-\lambda) \\ &= \lambda E[-X] + (1-\lambda)\left(\frac{1+\lambda}{\lambda}E[(q-X)^+] - q\right)\end{aligned} \tag{3.22}$$

Finally, given the representation of $AVaR$ from (3.9), the last expression is equal to the one from the formulation of the theorem.

It remains to note that the case when the inequality $X < 0$ is not satisfied follows directly from the boundedness of X, and cash invariance of $AVaR$, risk measure defined in the theorem, and $E[-X]$.

□

The proof of theorem 3.2.2 is a consequence of the preceding proof. Indeed, the assertion is clear for random variables $X < 0$. If, however, the inequality is not satisfied one only has to note that the equality $\psi_X = \psi_Y$ holds for all random variables X and Y such that $X - Y = c \in \mathbb{R}$ a.e. (see equation (3.17)); and the claim now follows from the cash invariance of $LCMU_\lambda$.

Appendix 3.D Calculations for AVaR and LCMU for a Log-Normally Distributed Position

We first introduce a new variable $y = \Phi^{-1}(t)$. We note that:

$$dt = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy,$$

which implies

$$\begin{aligned} AVaR_\lambda(X) &= -\frac{e^\mu}{\lambda} \int_0^\lambda \exp(\sigma\Phi^{-1}(t))\, dt \\ &= -\frac{e^\mu}{\lambda} \int_{-\infty}^{\Phi^{-1}(\lambda)} \frac{1}{\sqrt{2\pi}} \exp(\sigma y) \exp\left(-\frac{y^2}{2}\right) dy \end{aligned}$$

Now, completing the squares and introducing a new variable $z = y - \sigma$ we get:

$$\begin{aligned} AVaR_\lambda(X) &= -\frac{\exp\left(\mu + \frac{\sigma^2}{2}\right)}{\lambda} \int_{-\infty}^{\Phi^{-1}(\lambda)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-\sigma)^2}{2}\right) dy \\ &= -\frac{\exp\left(\mu + \frac{\sigma^2}{2}\right)}{\lambda} \int_{-\infty}^{\Phi^{-1}(\lambda)-\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\ &= -\frac{\exp\left(\mu + \frac{\sigma^2}{2}\right)}{\lambda} \Phi(\Phi^{-1}(\lambda) - \sigma). \end{aligned}$$

Appendix 3.E Details on Optimal Portfolios

Problems (P1-3) have been solved for risk measures VaR and $AVaR$ in theorems 3.2.1, 3.3.1 and 3.4.1 in Gambrah and Pirvu (2014). Careful reading of the proofs reveals technical conditions under which their techniques can be used for other risk measures. We offer slightly more general formulations of the aforementioned theorems from Gambrah and Pirvu (2014) that will

allow us to solve problems (P1-3) for $LCMU$.

Let us consider versions of problems (P1-3) where the risk measure $LCMU$ is replaced with a risk measure ρ: we will refer to those problems as (ρP1), (ρP1) and (ρP3). We will give sufficient conditions under which the solutions of the more general problems are multiples of π_M defined in (3.8). The key assumption is the following:

> *Assumption* (A): There are measurable functions functions $G : \mathbb{R}^2 \to \mathbb{R}$ and $h : \mathbb{R}^2 \to \mathbb{R}$ such that:
>
> 1. $G(\cdot, y)$ is increasing and $G(x, \cdot)$ is decreasing
> 2. $h(x)$ is decreasing.
> 3. $\rho(L(\pi)) = h\left(G\left(\mu(\pi), \psi(\pi)\right)\right).$

Measures $AVaR_\lambda$ and $LCMU_\lambda$ satisfy the assumption. Indeed, if we introduce a function

$$\varphi_\lambda(y) = \frac{1}{\lambda}\Phi(\Phi^{-1}(\lambda) - \sqrt{y})$$

it can easily be confirmed that:

$$\begin{aligned} AVAR_\lambda(L(\pi)) &= x_\pi - x_\pi R \exp(G^A_\lambda(\mu(\pi), \psi(\pi)), \\ &\quad \text{where } G^A_\lambda(x, y) = x + \ln(\varphi_\lambda(y)); \\ LCMU_\lambda(L(\pi)) &= x_\pi - x_\pi R \exp(G^L_\lambda(\mu(\pi), \psi(\pi)), \\ &\quad \text{where } G^L_\lambda(x, y) = x + \ln(\lambda + (1-\lambda)\varphi_{\frac{\lambda}{1+\lambda}}(y)). \end{aligned}$$

Theorem 3.E.1. *If the risk measure ρ satisfies the assumption (A) above, there are constants c_1, c_2 and c_3 such that strategies $\pi^*_1 = c_1\pi_M, \pi^*_2 = c_2\pi_M, \pi^*_3 = c_3\pi_M$ solve problems (ρP1), (ρP2) and (ρP3).*

For the sake of completeness we offer the proof of the theorem; it is essentially the proof offered in Gambrah and Pirvu (2014) with several small imprecisions and errors rectified.

We begin with proving two auxiliary results:

Lemma 3.E.1. *For a fixed $\kappa < 0$ the strategy*

$$\pi_\lambda(t) = -\frac{1}{2\kappa}(\sigma(t)\sigma(t)')^{-1}B(t) = -\frac{1}{2\kappa}\pi_M(t)$$

solves the maximization problem:

$$\max_{\pi\in\mathcal{Q}} \mu(\pi) + \kappa\psi(\pi).$$

Proof. Note that for any vectors $\pi, B \in \mathbb{R}^m\backslash\{(0,\dots,0)\}$ and any invertible matrix $\sigma \in \mathbb{R}^{m\times m}$ we have:

$$||\sigma'\pi||^2 + \frac{1}{\kappa}B'\pi = ||\sigma'\pi + \frac{1}{2\kappa}\sigma^{-1}B||^2 - \frac{1}{4\kappa^2}||\sigma^{-1}B||^2.$$

Indeed, by completing the squares:

$$\begin{aligned}
||\sigma'\pi||^2 + \frac{1}{\kappa}B'\pi &= \pi'\sigma\sigma'\pi + \frac{2}{2\kappa}(\sigma^{-1}B)'\sigma'\pi \\
&= \pi'\sigma\sigma'\pi + \frac{1}{2\kappa}\pi'\sigma\sigma^{-1}B + \frac{1}{2\kappa}(\sigma^{-1}B)'\sigma'\pi + \frac{1}{4\kappa^2}(\sigma^{-1}B)'\sigma^{-1}B \\
&\quad - \frac{1}{4\kappa^2}(\sigma^{-1}B)'\sigma^{-1}B \\
&= \pi'\sigma(\sigma'\pi + \frac{1}{2\kappa}\sigma^{-1}B) + \frac{1}{2\kappa}(\sigma^{-1}B)'(\sigma'\pi + \frac{1}{2\kappa}\sigma^{-1}B) \\
&\quad - \frac{1}{4\kappa^2}(\sigma^{-1}B)'\sigma^{-1}B \\
&= (\sigma'\pi + \frac{1}{2\kappa}\sigma^{-1}B)'(\sigma'\pi + \frac{1}{2\kappa}\sigma^{-1}B) - \frac{1}{4\kappa^2}(\sigma^{-1}B)'\sigma^{-1}B.
\end{aligned}$$

Hence:

$$\begin{aligned}
\mu(\pi) + \kappa\psi(\pi) &= \kappa\left(\int_0^T \frac{1}{\kappa}B(s)'\pi(s) + ||\sigma(s)'\pi(s)||^2\,ds\right) \\
&= \kappa\left(\int_0^T ||\sigma(s)'\pi(s) + \frac{1}{2\kappa}\sigma^{-1}(s)B(s)||^2\,ds\right) \\
&\quad - \frac{1}{4\kappa^2}\int_0^T ||\sigma^{-1}(s)B(s)||^2\,ds.
\end{aligned}$$

Note that only the first term in the last expression contains π. Thus, since $\kappa < 0$, the maximization problem from the formulation of the lemma reduces to:

$$\min_{\pi \in \mathcal{Q}} \int_0^T ||\sigma(s)'\pi(s) + \frac{1}{2\kappa}\sigma^{-1}(s)B(s)||^2 \, ds.$$

The last integral is non-negative. Furthermore, direct calculations show that it is equal to zero for π_κ, which proves the claim. □

Lemma 3.E.2. *Maximization problem:*

$$\max_{\pi \in \mathcal{Q}} \mu(\pi) \text{ subject to } \psi(\pi) = \varepsilon^2$$

is solved by

$$\pi_\varepsilon = \frac{\varepsilon}{\Theta}(\sigma(t)\sigma(t)')^{-1}B(t) = \frac{\varepsilon}{\Theta}\pi_M(t),$$

where

$$\Theta = \sqrt{\int_0^T ||\sigma(s)^{-1}B(s)||^2 \, ds}.$$

Proof. Direct calculations show that indeed $\psi(\pi_\varepsilon) = \varepsilon^2$.

Previous lemma established a mapping $\kappa \to \pi_\kappa$. Note that, by choosing $\kappa_\varepsilon = -\Theta(2\varepsilon)^{-1} < 0$ we have $\pi_\varepsilon = \pi_{\kappa_\varepsilon}$. The claim now follows directly by considering the Lagrangian: $\mathcal{L}(\pi, \kappa) = \mu(\pi) + \kappa(\psi(\pi) - \varepsilon^2)$. Indeed, for any strategy π satisfying the constraint $\psi(\pi) = \varepsilon^2$ we have:

$$\mu(\pi) = \mathcal{L}(\pi, \kappa_\varepsilon) \leq \mathcal{L}(\pi_\varepsilon, \kappa_\varepsilon) = \mu(\pi_\varepsilon),$$

where the inequality is the consequence of the previous lemma and the fact that, for a fixed $\kappa < 0$, the strategy π_κ maximizes $\mathcal{L}(\pi, \kappa)$.

□

Before we turn to proving the theorem we introduce some notation. For nonnegative ε we denote by $\mathcal{Q}_\varepsilon$ the set of all the strategies $\pi \in \mathcal{Q}$ such that $\psi(\pi) = \varepsilon^2$. Note that, due to the definition of ψ and the assumptions on σ

we have $\varepsilon \in I$ where:

$$I = \left[0, \int_0^T ||\sigma(s)||^2 \, ds\right],$$

and for every ε within that interval $\mathcal{Q}_\varepsilon \neq \emptyset$.

Clearly:

$$\bigcup_{\varepsilon \in I} \mathcal{Q}_\varepsilon = \mathcal{Q}.$$

Proof of 3.E.1 - ($\rho P1$) . Due to monotonicity of h we can reduce the problem to:

$$\min_{\pi \in \mathcal{Q}} G(\mu(\pi, \psi(\pi)).$$

We first solve the problem for a fixed $\varepsilon \in I$:

$$\min_{\pi \in \mathcal{Q}_\varepsilon} G(\mu(\pi), \varepsilon^2).$$

Due to monotonicity of $G(\cdot, x)$ this reduces to:

$$\min_{\pi \in \mathcal{Q}_\varepsilon} \mu(\pi).$$

By lemma 3.E.2 the solution is: $\pi_\varepsilon = \frac{\varepsilon}{\Theta}(\sigma(t)\sigma(t)')^{-1}B(t)$. Clearly, the problem ($\rho P1$) is now equivalent to:

$$\min_{\varepsilon \in I} G(\mu(\pi_\epsilon), \varepsilon^2).$$

Direct calculations show that $\mu(\pi_M(t)) = \Theta^2$, hence the continuous function:

$$g(\varepsilon) = G(\mu(\pi_\epsilon), \varepsilon^2) = G(\frac{\varepsilon}{\Theta}\mu(\pi_M(t)), \varepsilon^2) = G(\Theta\varepsilon, \varepsilon^2)$$

is defined on a closed interval and thus attains its maximum. This implies that there is a $\varepsilon_1 \in I$ such that $\pi_1 = \pi_{\varepsilon_1} =$ solves the prolem ($\rho P1$). In that case $c_1 = \varepsilon_1/\Theta$.

□

Proof of 3.E.1 - ($\rho P2$) . Similarly as in the proof regarding the problem ($\rho P1$) the minimization problem immediately reduces to:

$$\max_{\pi\in\mathcal{Q}} G(\mu(\pi,\psi(\pi)) \text{ such that } E[X^\pi(T)] = M$$

The condition $E[X^\pi(T)] = M$ is, due to (3.7), equivalent to:

$$\mu(\pi) = \ln\left(\frac{M}{x_\pi R}\right) =: \zeta. \tag{3.23}$$

Hence, the optimization problem can be rewritten as:

$$\max_{\pi\in\mathcal{Q}} G(\zeta,\psi(\pi)) \text{ such that } \mu(\pi) = \zeta,$$

which, due to monotonicity of G, further reduces to:

$$\min_{\pi\in\mathcal{Q}} \psi(\pi) \text{ such that } \mu(\pi) = \zeta.$$

We can now use lemma 3.E.1 to solve this problem. Indeed, for a fixed $\kappa < 0$, the maximization problem in the formulation of 3.E.1 is equivalent to the problem of minimizing $\psi(\pi) + \frac{1}{\kappa}\mu(\pi)$: the strategy π_κ solves both problems. Hence, for a fixed $\kappa < 0$, $\pi_{1/\kappa}$ solves the problem of minimizing $\psi(\pi) + \kappa\mu(\pi)$, and thus also the equivalent problem:

$$(P_\zeta) \qquad \min_{\pi\in\mathcal{Q}} \psi(\pi) + \kappa(\mu(\pi) - \zeta).$$

Thus, the solution of the initial optimization problem is:

$$\pi_2^* = \frac{\Theta^2}{\zeta}(\sigma(t)\sigma(t)')^{-1}B(t) = \pi_{1/\kappa_2^*} \text{ for } \kappa_2^* = -\Theta^2(2\zeta)^{-1}$$

Indeed, direct calculations show that $\mu(\pi_2^*) = \zeta$ and, for any π that satisfies $\mu(\pi) = \zeta$, we have

$$\psi(\pi) = \psi(\pi) + \kappa_2^*(\mu(\pi) - \zeta) \geq \psi(\pi_2^*) + \kappa_2^*(\mu(\pi_2^*) - \zeta) = \psi(\pi_2^*),$$

where the inequality is due to the fact that π_2^* solves the problem (P_ζ) for

$\kappa = \kappa_2^*$.

In this case $c_2 = \Theta^2/\zeta$. □

Proof of 3.E.1 - (ρP3) . We introduce the set $\mathcal{Q}' \subset \mathcal{Q}$ of all the strategies π satisfying the condition $\rho(L(\pi)) > C$, where C is the constant related to problems (P3) and (ρP3). We define $\mathcal{Q}'_\varepsilon = \mathcal{Q}_\varepsilon \cap \mathcal{Q}'$ and note that:

$$\mathcal{Q}'_\varepsilon = \left\{\pi \in \mathcal{Q} \mid \psi(\pi) = \varepsilon^2, \rho(L(\pi)) \leq C\right\} \text{ and } \bigcup_{\varepsilon \in I} \mathcal{Q}'_\varepsilon = \mathcal{Q}'.$$

Due to (3.7) the problem (ρP3) reduces to

$$\max_{\pi \in \mathcal{Q}'} \mu(\pi).$$

Let us consider a simpler problem:

$$(P'_\varepsilon) \qquad \max_{\pi \in \mathcal{Q}'_\varepsilon} \mu(\pi).$$

For a strategy $\pi \in \mathcal{Q}'_\varepsilon$ the constraint $\rho(L(\pi)) > C$ can be rewritten as:

$$\ln \frac{x_\pi - C}{x_\pi R} \leq G(\mu(\pi), \varepsilon^2).$$

The function $G(\cdot, \varepsilon^2)$ is increasing, hence it has an inverse that we denote with G_ε^{-1}. Thus the constraint can be rewritten as:

$$\mu(\pi) \geq G_\varepsilon^{-1}\left(\frac{x_\pi - C}{x_\pi R}\right) =: h(\varepsilon). \tag{3.24}$$

Let us consider the strategy π_ϵ from 3.E.2 that maximizes $\mu(\pi)$ over $\mathcal{Q}_\varepsilon$. Solving the problem P'_ε relies on noticing that the set $\mathcal{Q}'_\varepsilon$ is non-empty if and only if π_ε belongs to it. Indeed, if $\pi \in \mathcal{Q}'_\varepsilon$ then:

$$\mu(\pi_\epsilon) \geq \mu(\pi) \geq h(\varepsilon);$$

the first inequality is due to $\mathcal{Q}'_\varepsilon \subset \mathcal{Q}_\varepsilon$ and the second one is due to 3.E.2.

This allows us to rewrite the problem (ρP3) as follows:

$$\max_{\varepsilon \in I} \mu(\pi_\epsilon) \text{ such that } \mu(\pi_\varepsilon) \geq h(\varepsilon).$$

Due to the definition of μ, the value $\mu(\pi_\varepsilon)$ is increasing in ε and the problem reduces further to:

$$\max_{\varepsilon \in I} \varepsilon \text{ such that } \mu(\pi_\varepsilon) \geq h(\varepsilon).$$

Continuity of $\mu(\pi_\varepsilon)$ as a function of ε and monotonicity of $g(\varepsilon)$ ensure that the problem has a solution that we denote by ε_3.

This proves that $\pi_3^* = \pi_{\varepsilon_3}$ solves the optimization problem, in which case $c_3 = \varepsilon_3/\Theta$ □

References

Artzner, Philippe et al. (1999). "Coherent measures of risk". In: *Mathematical finance* 9.3, pp. 203–228.

Cheridito, Patrick and Mitja Stadje (2009). "Time-inconsistency of VaR and time-consistent alternatives". In: *Finance Research Letters* 6.1, pp. 40–46.

Föllmer, Hans and Alexander Schied (2011). *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

Gambrah, Priscilla Serwaa Nkyira and Traian Adrian Pirvu (2014). "Risk measures and portfolio optimization". In: *Journal of Risk and Financial Management* 7.3, pp. 113–129.

Gilboa, Itzhak and Massimo Marinacci (2016). "Ambiguity and the Bayesian paradigm". In: *Readings in Formal Epistemology*. Springer, pp. 385–439.

Gilboa, Itzhak and David Schmeidler (1989). "Maxmin expected utility with non-unique prior". In: *Journal of mathematical economics* 18.2, pp. 141–153.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006). "Ambiguity aversion, robustness, and the variational representation of preferences". In: *Econometrica* 74.6, pp. 1447–1498.

Merton, Robert C (1969). "Lifetime portfolio selection under uncertainty: The continuous-time case". In: *The review of Economics and Statistics*, pp. 247–257.

Rogers, Leonard CG (2013). *Optimal investment*. Springer.

Schied, Alexander and Ching-Tang Wu (2005). "Duality theory for optimal investments under model uncertainty". In: *Statistics & Decisions* 23.3/2005, pp. 199–217.

Chapter 4

A Note on the Perpetual American Straddle

Abstract

The value and the optimal exercise time of the perpetual American straddle is characterized by the unique solution of a *single* non-linear equation with one unknown variable.

4.1 Introduction

Perpetual American options can frequently be priced explicitly in the standard model of Black and Scholes (1973). This is the case with the *American straddle*: a portfolio consisting of a put and a call option on the same underlying asset with the same strike price. The pricing of the perpetual American straddle has been studied using different approaches and tools: in Alobaidi and Mallier (2002) by applying the theory of Laplace transforms, in Beibel and Lerche (1997) by transforming the problem to a "generalized parking problem", in Moraux (2009) by exploiting "an analogy with asymmetric rebates of double knock-out barrier options", in Gerber and Shiu (1994) "by means of the Esscher transform and the optional sampling theorem", and, more recently, by using a combination of several optimization techniques Lempa (2010) and Lamberton, Zervos, et al. (2013). In all of these papers the value function and the optimal exercise time are characterized by a solution of a non-linear system of equations consisting of (at least) two equations.

In this note, we show that the value function and the optimal exercise time of the perpetual American straddle can be characterized via a unique solution of a *single* one-variable equation; the solution lies in the interval $(0, 1)$. We do so by using one of the classical optimal stopping theory approaches: the Hamilton-Jacobi-Bellman equation and the smooth-fit principle in combination with a verification theorem. This leads to a system of non-linear equations that can, by appropriate transformations, be reduced to a single equation. To the best of our knowledge this is the first time that such one-equation characterization of the value and the optimal exercise time of the perpetual American straddle is obtained.

4.2 Result

Let the price process S_t be a a geometric Brownian motion,

$$dS_t = \alpha S_t dt + \sigma S_t dB_t,$$

where $\alpha \in \mathbb{R}$ and $\sigma \in \mathbb{R}$ are known constants. The American straddle yields a payoff

$$f(t, S_t) = e^{-rt}|S_t - I|$$

when exercised at time t, where $I > 0$ is the strike price and $r \leq \alpha$ is a given discount rate. (Inequality $r \leq \alpha$ is a standard assumption; see for example Shiryaev (1999).)

The value of the perpetual American straddle at time t is given by

$$V_t = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_t} E[e^{-r\tau}|S_\tau - I|], \tag{4.1}$$

where $\mathcal{T}_t$ is a set of all stopping times $\tau \geq t$. Our goal is to find a value function $v(t, x)$ such that $v(t, S_t) = V_t$ and an optimal stopping time τ^* such that $V_{\tau^*} = E[e^{-r\tau^*}|S_{\tau^*} - I|]$. Hamilton-Jacobi-Bellman (HJB) equation related to this problem (see e.g. Øksendal (2003, ch.11)) is:

$$\max_{(t,x)\in[0,+\infty]\times\mathbb{R}} \{f(t, x) - v(t, x), v_t(t, x) + \mathcal{L}v(t, x)\} = 0, \tag{4.2}$$

where $\mathcal{L} := \alpha\partial_x + \frac{1}{2}\sigma^2\partial_{xx}$ is a differential operator related to Ito's lemma.

A well known approach when dealing with time-discounted optimal stopping problems is to assume that the value function is of the form

$$v(t, x) = e^{-rt}\varphi(x);$$

this will later be confirmed using a verification theorem. The equality

$$v_t(t, x) + \mathcal{L}v(t, x) = 0$$

holds on the continuation region (due to the HJB equation). After canceling out e^{-rt} this gives:

$$r\varphi(x) - \alpha x\varphi'(x) - \frac{1}{2}\sigma^2 x^2\varphi''(x) = 0.$$

The last equation is a well known Cauchy-Euler ordinary differential equation and its solution is

$$\varphi(x) = Ax^{\lambda} + Bx^{\mu},$$

where A and B are two unknown constants and λ and μ are the solutions of the characteristic equation

$$r - \alpha m - \frac{1}{2}\sigma^2 m(m-1) = 0.$$

It can be easily verified that inequalities $\lambda > 1$ and $\mu < 0$ hold.

It is known that the optimal stopping time will be the first exit time from the interval $(x_1, x_2) \ni I$: it is optimal to exercise the put (call) option when the value of S_t goes beneath x_1 (above x_2). Furthermore, on the stopping region, the HJB equation implies that $f = v$. Thus, we assume that the function v should be of the form:

$$v(t,x) = \begin{cases} e^{-rt}(I - x), & 0 < x < x_1 \\ e^{-rt}(Ax^{\lambda} + Bx^{\mu}), & x_1 \leq x \leq x_2 \\ e^{-rt}(x - I), & x > x_2 \end{cases} \tag{4.3}$$

where A, B, x_1, x_2 are constants chosen in a way that makes the function v differentiable (smooth pasting conditions). In particular, we require continuity and differentiability in x_1 and x_2.

It is already clear that, should we find such constants, the above function $v(t,x)$ will be a value function. Indeed, conditions of any of the well known verification theorems for the optimal stopping of diffusions (e.g. ch. 3 in Krylov (2008) or ch. 10 in Øksendal (2003)) are easily satisfied for functions that coincide, piecewise, with (discounted) linear combinations of power functions. Furthermore, since the functions v and f coincide outside the interval (x_1, x_2), if v is indeed the value function, then the optimal stopping time is:

$$\tau^* = \inf\{t \geq 0 | S_t \notin (x_1, x_2)\}.$$

Smooth pasting conditions lead to a highly non-linear system of equations. We show that it can be reduced to a single equation:

Theorem 4.2.1. *The value process of the perpetual American Straddle V_t defined in (4.1) satisfies the equality $V_t = v(t, S_t)$ for the function v as defined in (4.3) where*

$$A = \frac{1}{\mu - \lambda}((1-\mu)x_1^{1-\lambda} + \mu x_1^{-\lambda}); \qquad x_2 = \frac{\mu I}{\mu - 1}\frac{1+\gamma^{-\lambda}}{1+\gamma^{1-\lambda}}$$
$$B = \frac{1}{\lambda - \mu}((1-\lambda)x_1^{1-\mu} + \lambda x_1^{-\mu}); \qquad x_1 = \gamma x_2$$

and $\gamma \in (0,1)$ is the unique number satisfying:

$$\frac{\mu}{\mu - 1}\frac{1+\gamma^{-\lambda}}{1+\gamma^{1-\lambda}} - \frac{\lambda}{\lambda - 1}\frac{1+\gamma^{-\mu}}{1+\gamma^{1-\mu}} = 0. \tag{4.4}$$

Proof. Smooth pasting conditions, after cancelling out e^{-rt}, can be written as:

$$\begin{aligned} I - x_1 &= Ax_1^\lambda + Bx_1^\mu, & -x_1 &= A\lambda x_1^\lambda + B\mu x_1^\mu, \\ x_2 - I &= Ax_2^\lambda + Bx_2^\mu, & x_2 &= A\lambda x_2^\lambda + B\mu x_2^\mu. \end{aligned} \tag{4.5}$$

In order to prove the theorem it is, by construction of the value function v, sufficient to prove that unique solution of the system (4.5) is the one given in the formulation of the theorem. The proof consists of reducing the system to equation (4.4), and proving that the solution of the latter is unique on the interval $(0,1)$.

First we comment on the uniqueness of the solution of the system of equations (4.5). Due to the uniqueness of the value function of the optimal stopping problems the solution of the system above must be unique. Indeed, two different solutions of the system (4.5) would lead to two functions v_1 and v_2 both of which would satisfy the verification theorem and the equation $v_1(t, S_t) = v_2(t, S_t)$ would holds almost surely, which is clearly impossible.

We now turn to proving the existence. We can eliminate variables A and

B in the two equations containing x_1 by treating them as a two dimensional linear system. Since determinant of that system is $D = x_1^{\lambda+\mu}(\mu - \lambda) \neq 0$, A and B are uniquely determined by it. We can do the same for the two equations containing x_2. If we introduce, for notational purposes, the function $Q(x; \mu, \lambda) = (\mu - \lambda)^{-1}(1 - \mu)x^{1-\lambda} + \mu x^{-\lambda}$, we can write the solutions of those two systems as:

$$A = Q(x_1; \mu, \lambda);\ B = Q(x_1; \lambda, \mu);\ A = -Q(x_2; \mu, \lambda);\ B = -Q(x_2; \lambda, \mu).$$

Equating the expressions for A and B we obtain the following nonlinear system with two equations and two variables, x_1 and x_2:

$$Q(x_1; \mu, \lambda) + Q(x_2; \mu, \lambda) = 0 \qquad Q(x_1; \lambda, \mu) + Q(x_2; \lambda, \mu) = 0 \quad (4.6)$$

Due to the nice form of the above system, we immediately see that if (x_1, x_2) is its solution so is (x_2, x_1). This means that there is a unique solution pair satisfying $x_1 < x_2$, and it will be the unique solution that we are looking for. We introduce a variable γ such that $x_1 = x_2\gamma$; since inequality $0 < x_1 < x_2$ holds, we have $\gamma \in (0, 1)$. The right hand side of the first equation of the system (4.6) can now, after some simple calculations, be written as:

$$Q(x_2\gamma; \mu, \lambda) + Q(x_2; \mu, \lambda) = (1 - \mu)x_2^{1-\lambda}(1 + \gamma^{1-\lambda}) + \mu x_2^{-\lambda}(1 + \gamma^{-\lambda}).$$

from which we obtain:

$$x_2 = \frac{\mu I}{\mu - 1}\frac{1 + \gamma^{-\lambda}}{1 + \gamma^{1-\lambda}}$$

Similarly, by changing $x_1 = x_2\gamma$ in $Q(x_1; \lambda, \mu) + Q(x_2; \lambda, \mu) = 0$ after multiplication with $x_1^{-\mu}$ we obtain:

$$x_2 = \frac{\lambda I}{\lambda - 1}\frac{1 + \gamma^{-\mu}}{1 + \gamma^{1-\mu}}.$$

Equating the two obtained expressions for x_2, after rearanging and cancelling out parameter I, we obtain the one-dimensional equation (4.4), stated in the

formulation of the theorem.

It remains to prove that there exists a unique solution of equation (4.4) in the interval $(0, 1)$. Indeed, if we denote the left hand side of the equation with $h(\gamma)$ it is obvious that function h is continuous on $(0, 1)$, and it is easy[1] to check that $h(1) < 0$, and $\lim_{\gamma \to 0+} h(\gamma) = +\infty$. We can thus conclude that a solution exists on the interval $(0, 1)$, and its uniqueness is a consequence of the argument from the beginning of the proof.

□

4.3 Conclusion

We have demonstrated that the perpetual American straddle, a classical and well studied portfolio of options, can be priced and fully characterized using a unique solution of a single non-linear equation on the unit interval. The solution itself gives a direct relation between two exercise boundaries of the American straddle. Beyond the theoretical relevance, this potentially has practical implications: estimating the value of this well known portfolio can be reduced to the estimation of the single parameter.

[1] Because $\lambda > 1$ and $\mu < 0$.

References

Alobaidi, G and R Mallier (2002). "Laplace transforms and the American straddle". In: *Journal of Applied Mathematics* 2.3, pp. 121–129.

Beibel, Martin and Hans Rudolf Lerche (1997). "A new look at optimal stopping problems related to mathematical finance". In: *Statistica Sinica* 7.1, pp. 93–108.

Black, Fischer and Myron Scholes (1973). "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3, pp. 637–654.

Gerber, Hans U and Elias SW Shiu (1994). "Martingale approach to pricing perpetual American options". In: *ASTIN Bulletin: The Journal of the IAA* 24.2, pp. 195–220.

Krylov, Nikolaj Vladimirovič (2008). *Controlled diffusion processes*. Vol. 14. Springer Science & Business Media.

Lamberton, Damien, Mihail Zervos, et al. (2013). "On the optimal stopping of a one-dimensional diffusion". In: *Electronic Journal of Probability* 18.

Lempa, Jukka (2010). "A note on optimal stopping of diffusions with a two-sided optimal rule". In: *Operations Research Letters* 38.1, pp. 11–16.

Moraux, Franck (2009). "On perpetual American strangles". In: *The Journal of Derivatives* 16.4, pp. 82–97.

Øksendal, Bernt (2003). "Stochastic differential equations". In: *Stochastic differential equations*. Springer, pp. 65–84.

Shiryaev, Albert N (1999). *Essentials of stochastic finance: facts, models, theory*. Vol. 3. World scientific.

Short Curriculum Vitae

Lazar Obradović

Born in 1987 in Podgorica, Montenegro.

2005 – 2009: BSc studies
Department of Applied Mathematics and Computer Science
Faculty of Sciences and Mathematics, University of Montenegro

2009 - 2011: Erasmus Mundus MSc studies
QEM – Models and Methods of Quantitative Economics .
Bielefeld University and Université Paris 1 Panthéon-Sorbonne

2012 - 2018: PhD studies
Center for Mathematical Economics
Faculty of Business Administration and Economics
Bielefeld University

2011 - present: employed at
Faculty of Science and Mathematics,
University of Montenegro