

Model uncertainty : the effect on robustness, estimation and stochastic optimisation

Citation for published version (APA):

Balter, A. G. (2016). *Model uncertainty : the effect on robustness, estimation and stochastic optimisation*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20160219ab>

Document status and date:

Published: 01/01/2016

DOI:

[10.26481/dis.20160219ab](https://doi.org/10.26481/dis.20160219ab)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Model Uncertainty

The Effect on Robustness, Estimation
and Stochastic Optimisation

Anne Balter



© Anne Balter, Maastricht 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

This book was typeset by the author using L^AT_EX.

Published by Universitaire Pers Maastricht

ISBN: 978-94-6159-517-1

Printed in The Netherlands by Datawyse Maastricht

Model Uncertainty

The Effect on Robustness, Estimation
and Stochastic Optimisation

DISSERTATION

to obtain the degree of Doctor at
Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. L.L.G. Soete,
in accordance with the decision of the Board of Deans,
to be defended in public
on Friday, 19 February, 2016 at 10.00 o'clock

by

Anne Gerarda Balter



Supervisors

Prof. dr. A.A.J. Pelsser

Prof. dr. P.C. Schotman

Assessment Committee

Prof. dr. F.C. Palm (Chair)

Prof. dr. W.F.M Bams

Prof. dr. R.J.A. Laeven (University of Amsterdam)

Prof. dr. B.J.M. Werker (Tilburg University)

This research was financially supported by The Netherlands Organisation for Scientific Research (NWO).

To my parents
Ronald and Anke

Acknowledgements

“You cannot be certain about uncertainty.”
-Frank Knight (1885-1972)

“You cannot be certain about uncertainty.”

-Frank Knight (1885-1972)

Risk stands for the uncertainty that an event might happen while knowing the probability whereas ambiguity represents the nescience of the probability. Any probability, about which one might be certain, indicates uncertainty. Therefore ambiguity can be interpreted as uncertainty of a second order since the additional uncertainty comes from the lack of certainty about the probability. As such, I believe that the future is ambiguous.

Similarly, the future path I followed was something I was ambiguous about three years ago. Or even earlier. The only but solid certainty in my life was that “I want to do something with mathematics”. This, in combination with the international atmosphere, made me choose to study Econometrics and Operations Research in Maastricht. After my study abroad in Australia, I started a part-time job at an Actuarial Department and I obtained the M.Sc. in this field. Subsequently, thanks to the initial approach from Antoon, I am now at the point of defending my thesis in order to obtain my Doctor of Philosophy (Ph.D.).

First of all, I thank my supervisor Antoon Pelsser for providing me the opportunity to enter the academic world. I learned a lot from all the meetings we had, which were seldom brief while his agenda was quite full. I would also like to thank Peter Schotman who was the co-author of my first paper and became my second supervisor. Especially in the first year, he guided me to find order in the chaos of subjects.

Furthermore, I would like to express my appreciation to Jean-Pierre Urbain. The brief talks we had once in a while and the fact that he always kept a sort of eye on me is something that has been very valuable to me. To his enjoyment, I would like to highlight the tips concerning broadening my understanding in Bayesian statistics by attending a workshop in Coimbra and a course at Université de Liège. Moreover, thanks to Jean-Pierre’s involvement I obtained the postdoc position that I currently possess. Similarly, I want to thank Eric Beutner and Stephan Smeekes for their approachability.

I would like to thank the reading committee, Franz Palm, Dennis Bams, Bas Werker and Roger Laeven, for their time and careful reading of the thesis. Also, I thank Karin van den Boorn and Yolanda Paulissen from the Quantitative Economics Department and Carina van der Velde from the Finance Department, for their help in the administrative support. Also I would like to thank my colleagues from both departments for the time we spent together, namely Marina Friedrich,

Verena Jung, Jan Lohmeyer, Hanno Reuvers, Sean Telg, Martijn van Brink, Marc Schröder, Thomas Götz, Hande Karabiyik, Greg Csapó, Artem Duplinsky, Yuliya Shapovalova, Kouma Gnameho, Ahmad Salahnejhad, Vincent Kreuzen, Anna Zseleva, Veerle Timmermans, Tim Oosterwijk, Andrej Winokurov, Andrey Kateshov, Rogier Quaedvlieg, Sally Shen and Janina Schweizer among others. With special thanks to my office mates Oksana Balabay and Alex Heinemann, who will be my paranympths.

Besides my colleagues and friends at work I would like to thank Sanne Willems, Martijn Schröder and Lieke Schreurs with whom I studied together.

Another task that I thoroughly enjoyed doing was being a tutor during the semi-weekly tutorials of the very first course for econometricians. At first hand this was not the most appealing course for “us”, though introducing students to university and sharing my own experiences with them led to a nice atmosphere during these meetings. For the more quantitative courses, I enjoy explaining mathematics. After both types of meetings I often headed back to my office with a big smile. The individual contacts with students writing their bachelor or master thesis or the more distant contact while giving a lecture were my pleasure. Related to this, I would like to thank Jeannette Hommes for her guidance during the procedure to obtain my University Teaching Qualification degree.

Moreover, I would like to mention Harmonie St. Caecilia Hoensbroek, the orchestra in which I play the saxophone, for the joy it gives me to rehearse together. Therefore I thank Esther Bonn  and Marlous de Jong. Also the running club, Atletiek Vereniging Oranje Nassau, gave me lots of relaxation, especially training for the marathon during the last year of my Doctor of Philosophy. With special credits to Bas Pierik.

Finally, I want to thank my family. I thank my parents, Ronald Balter and Anke Drielsma, for their support, and for raising me to be the person I am today with the background and ideas about life. Especially the characteristic of persistent determination, not only for achievements concerning education, which you obtain for yourself only, but also concerning sports and travelling. Moreover, the attempts to explain what I did in my thesis helped me tremendously in my own understanding. I am grateful to my sister Leonie Balter and my brother Benno Balter, with whom I share the spirit for extremes.

Anne Balter

Contents

*“It is good to have an end to journey toward;
but it is the journey that matters, in the end.”*

-Ernest Hemingway (1899-1961)

Contents		xi
1 Introduction		1
1.1 Literature Review: Pricing Methods		3
1.2 Literature Review: Uncertainty Sets		8
1.3 Outline		11
2 What does a Term Structure Model imply about very Long-Term Discount Rates?		13
2.1 Introduction		15
2.2 Data		17
2.3 Term Structure Models		19
2.3.1 Vasicek Model		20
2.3.2 Extrapolation		21
2.4 Econometric Model		23
2.5 Maximum Likelihood Estimates		26
2.6 Bayesian Analysis		28
2.6.1 Priors		28
2.6.2 Posterior Densities		30
2.7 Results		32
2.7.1 Vasicek Extrapolation		32
2.7.2 Extrapolation Errors		35
2.7.3 Alternative Extrapolations		36
2.8 Robustness		39
2.8.1 Maturities		39
2.8.2 Smoothness		40
2.8.3 Analysis United States		42
2.9 Conclusion		48
2.A Derivation of Yields		49
2.B Affine Relation Between \mathbb{P} and \mathbb{Q}		50
2.C Admissability Conditions for Σ		51
2.D Covariance Decomposition		53
2.E Forward Rate Extrapolation		53
2.F Sensitivity		55

3 Pricing and Hedging in Incomplete Markets with Model Ambiguity	61
3.1 Introduction	63
3.1.1 Literature Review	65
3.2 Optimal Pricing Result	66
3.2.1 The Agent's Problem	66
3.2.2 Financial Market with Model Ambiguity	68
3.2.3 Optimisation	70
3.3 Examples	79
3.3.1 Example I: Pure Hedgeable Risk	80
3.3.2 Example II: Pure Unhedgeable Risk	80
3.3.3 Example III: Correlated Risk	81
3.3.4 Example IV: Life Insurance Contract	83
3.4 Conclusion	84
3.A Optimisation	86
4 Sets of Indistinguishable Models for Robust Optimisation	89
4.1 Introduction	91
4.1.1 Literature Review	91
4.1.2 Intuition	92
4.2 Statistically Indistinguishable Models	93
4.2.1 Deterministic Drift Term	95
4.2.2 Stochastic Drift Term	97
4.3 Stochastic and Time-Consistent Set of Indistinguishable Models . .	98
4.3.1 Time-Consistency	98
4.3.2 Maximum Power Calculation	99
4.3.3 Sufficiency	103
4.4 Bounds on Divergences	104
4.5 Conclusion	107
4.A Variance Hyperbolic Cosine	108
4.B Power Hyperbolic Cosine	109
4.C Divergences	109
4.C.1 Kullback-Leibler	109
4.C.2 Burg Entropy	110
4.C.3 J-Divergence	111
4.C.4 χ^2 -Distance	112

4.C.5	Modified χ^2 -Distance	112
4.C.6	Hellinger Distance	113
4.C.7	Variation Distance	113
4.C.8	χ -Divergence of Order $\theta > 1$	114
4.C.9	Cressie-Read	114
5	Conclusion	117
6	Valorisation	121
6.1	Pension Funds	123
6.2	Insurance Companies	124
6.3	Misspecification	126
6.4	Climatology	127
	Bibliography	129
	Nederlandse Samenvatting	139
	Curriculum Vitae	145

Chapter 1

Introduction

“Reports that say that something hasn’t happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don’t know we don’t know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.”

-Donald Rumsfeld (1932)

Models can be wrong since models are simplifications of the real world. This simple fact did not only serve as a foundation for an entire branch of economic research, but also for this thesis. Admitting the limitation of models is important for policies. How financial and economic decision making can incorporate uncertainty is the focus of this dissertation. However, the generality of modelling and uncertainty goes beyond these fields. The same procedures and methods can be utilised in physics and engineering. Possible applications of this thesis are quantifying the uncertainty involved in the climate change effect, the probability of disasters happening such as earthquakes, and the safety of medicines. The recognition of uncertainty is important for any model-based decisions. The social relevance is described in the Valourisation (Chapter 6).

The first step we make towards a more realistic representation of the economy is that we assume that we are in an incomplete market. In a complete market all assets are assumed to be traded such that any cash flow can be replicated by a combination of these assets. However, not everything is traded or fairly priced. For instance, there is no liquid market in the number of people alive or in the possibility of an earthquake. For these two examples there are instruments available, mortality swaps and catastrophe bonds respectively. However, these assets are in limited amounts available and are based on a general index. Even more subtle the assumption of no transaction costs for liquid assets is too simplistic. The market incompleteness assumption incorporates all these frictions. This dissertation begins with exploring previous literature on pricing methods in *incomplete markets*. Chapter 2 and 3 contribute to this field.

Additional to the presence of unhedgeable risk factors is the acknowledgement of model uncertainty. In this thesis we mainly concentrate on robust decisions in financial and economic situations. This implies that the agent evaluates his behaviour against plausible alternative models and chooses the strategy that is least sensitive to perturbations of the model. Therefore we develop a method to construct a set of indistinguishable models *ex ante*. However, before we do so in Chapter 4, we first discuss the literature that deals with *model ambiguity* in the second review of this introductory chapter.

1.1 Literature Review: Pricing Methods

The First Fundamental Theorem of Asset Pricing (FFTAP) (Harrison and Kreps, 1979) states that the market is arbitrage free if and only if at least one equiva-

lent martingale measure (EMM) exists. Moreover, the market is complete if this EMM is unique (A)¹, whereas the market is incomplete if multiple EMM's exist (B) (Delbaen and Schachermayer, 1994). In complete markets derivatives are priced by replication. E.g. the price of plain vanilla options can be priced as the discounted expectation under the unique EMM. The Black-Scholes formula (1973) prices call and put options under additional assumptions. The uniqueness of the EMM corresponds to a unique price that can be obtained by replication of any payoff. However in incomplete markets many different methods are presented to quantify the “price” because multiple measures exist.

The literature about pricing in incomplete markets can be divided in methods that take all measures into account versus those that consider a subset of martingale measures. See Figure 1.1 for an overview of the pricing methods in the literature.

Firstly, a well-known method from the class that takes into account “all measures” (C) when pricing in an incomplete market is superreplication or superhedging (Kreps, 1981). The infimum is taken over all hedging strategies such that the price is larger than or equal to the liability. The minimised price is equal to the optimal objective value obtained by maximising the expected payoff over all measures such that the observed market prices are replicated. These two optimisation problems that lead to the same optimal solution are the primal and dual representation respectively. Duality means that an optimisation problem may be viewed from both perspectives. Subreplication or subhedging takes the infimum over all measures. A drawback of this pricing method are the extremely wide intervals for the price. The large amount of initial capital implied makes this an unpractical approach (Schied and Föllmer, 2011).

We can subdivide the class of “all measures” further into “penalty function on the measures” and “loss functions on the replication of the shortfall”.

Secondly, Ben-Tal et al. (2013) consider all possible measures on which a penalty function (D) is specified. Different penalty functions can be picked from the set of ϕ -divergence (or f -divergence). ϕ -Divergence functions measure the distance between two probability distributions weighted by the specific function. Mostly used is the Kullback-Leibler divergence, also known as entropy. This penalty term has gained a lot of attention because of Hansen and Sargent (2001, 2008).

Thirdly, there is the class of methods that do not state the penalty function

¹Letter refers to the corresponding box in Figure 1.1.

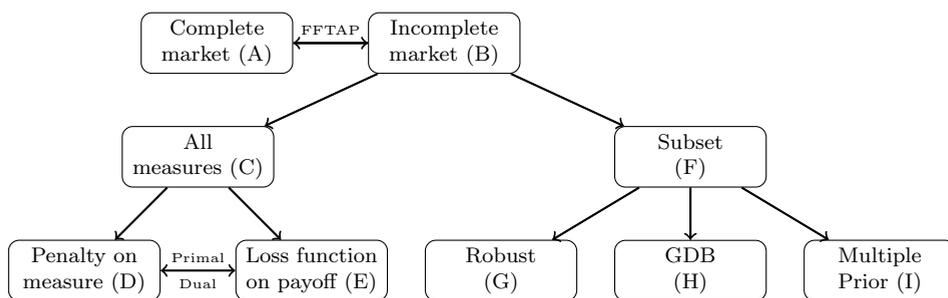


Figure 1.1: Overview of Pricing Methods

directly, but specify a loss function on the payoff instead (E). Most utility functions belong to this last category. Utility indifference pricing is based on the price of a claim such that an agent who is maximising his utility is indifferent between having the claim or not. The founder of this concept is Merton (1971, 1973). Hodges and Neuberger (1989) extended this to the dynamic setting. The choice among the different loss functions that has to be made is the challenging aspect of this class of methods. Also part of the class of “loss functions” is mean-variance hedging (Schweizer, 1992) which is based on the selection of a quadratic loss function. The strategy that maximises the probability that the uncertain payoff is hedged is called quantile hedging (Föllmer and Leukert, 1999), this corresponds with the indicator function as loss function. The strategy that maximises the probability of success, where success is the minimal shortfall is derived by Föllmer and Leukert (2000), which corresponds with a put option as loss function. Also here the challenge is to choose the appropriate loss function in whichever situation. The exponential utility function leads to a direct penalty function, and thus belongs to the class that considers all measures with a specific penalty term. For recent work on utility indifference pricing, in specific exponential utility functions, see Henderson (2002, 2005), Henderson and Hobson (2009), Miao and Wang (2007), Hu et al. (2005), Musiela and Zariphopoulou (2004), Young and Zariphopoulou (2002) and Zariphopoulou (2001).

The subclasses of those methods that consider a penalty on the measures and those that consider a loss function on the hedging errors can be linked through the fact that both subclasses are the primal and the dual of the same optimisation problems. For a general review of duality see Rogers (2003). Schweizer (1992) shows that the concept of mean-variance hedging corresponds with the minimal equivalent martingale measure. Via the dual problem the link between pricing

under the minimal entropy measure and exponential utility is shown by Delbaen et al. (2002).

We discussed the listed methods that can be labelled as those that consider “all measures”, thus we continue with the class of methods that consider a “subset of measures” (F).

Firstly, the path to the optimal hedge position of Good-Deal-Bounds (GDB) (H) (Cochrane and Saa-Requejo, 2000) is that the existence of assets with extreme market prices of risk are unlikely to be true (“too good to be true”). By the buy and sell imbalance these prices move towards an equilibrium, i.e. if the asset is too good, then demand will increase, which drives up the price and decreases the market price of risk and vice versa. Amongst others, Björk and Slinko (2006), Černý and Hodges (2002), Becherer (2009) and Klöppel and Schweizer (2007) extended this concept. Shen et al. (2014) consider an optimal control problem under parameter uncertainty, where the uncertainty set is based on a Good-Deal-Bound approach. The numerical results are obtained based on the utility function specified as the positive part of the surplus.

Secondly, additional to different methods of pricing in incomplete markets, the ambiguity that is involved when making decisions or during the process of pricing, can be implemented by a wide variety of different models. To define our concept of pricing under ambiguity note the difference between risk and ambiguity. Risk stands for uncertainty about an event that might happen while knowing the probability whereas ambiguity indicates the nescience of the probabilities itself. Since incompleteness corresponds to a set of martingale measures rather than one unique measure, ambiguity can be interpreted as the uncertainty among the different measures, i.e. ambiguity is equivalent to the set of martingale measures. The idea of robustness (G) we adhere to in this thesis is that the robust price and the robust optimal hedging strategy are the ones that are least sensitive to perturbations of the model. Barrieu and El Karoui (2005) consider a buyer and seller perspective in the two player game, where the pricing measure is known as the solution of an inf-convolution problem. Gundel and Weber (2007) and Goldfarb and Iyengar (2003) deal with robustness linked to portfolio optimisation. A literature review on ambiguity in asset pricing is performed by Guidolin and Rinaldi (2013).

Thirdly, the multiple prior (I) model from Gilboa and Schmeidler (1989) comes from a Bayesian approach where a set of priors are specified over a subset of measures. More robust control literature in mathematical finance can be found by

work of Garlappi et al. (2007) who extended the standard multiple prior approach. They consider a robust portfolio rather than the price. Although these three methods (robust optimisation, GDB and multiple priors) are in the literature presented as completely different methods they are actually based on one similar idea that given a subset one is not able to distinguish between the models within the set.

The question that pops up in this class of pricing methods is which subset should one take? Firstly, the model confidence set methodology introduced by Hansen et al. (2011) is a recursive method that tests which models to keep in the set of indistinguishable models and which ones to eliminate according to some loss function. By these decisional criteria the size of the set of alternatives can be chosen. Secondly, the size of the subset that consists of all plausible measures implied by the GDB methodology is determined by the bound on the market price of risk that is set. Those measures that are “too good to be true” are excluded from this set. Thirdly, the size determination issue discussed here is bypassed for the class of methods that consider “all measures”. The penalty and loss function methods build upon the assumption that it is possible to rank models. The literature review in Section 1.2 discusses these methods in more detail. Moreover, Chapter 4 contributes to the literature of quantifying *model ambiguity*.

Since we discussed the overview of the different pricing methods in the literature, we briefly discuss the interpretation and linkage between pricing and risk measures. Risk measures determine the smallest amount of cash that is needed to add to a random variable such that the risk is acceptable, i.e. such that the random variable falls inside the acceptance set (Artzner et al., 1999, 2007). This is equivalent to what a pricing operator does. For each risk measure a different amount of cash is needed, i.e. for each risk measure there is a pricing operator. Note the plus-minus sign switch in the formulation of the risk measure set-up: the amount of money that should be added to the random variable to make it acceptable corresponds with the price of the liability as the negative equivalent.

In the literature of risk measures, the same division as in the pricing literature can be made. Whereas we divided the pricing methods in those that initiate all measures versus those that initiate a subset, the risk measure literature can be divided in convex and coherent risk measures, respectively. Specifying a set of alternatives is mathematically equivalent to considering all measures with the indirect implied penalty function being equal to zero inside the set (in this thesis

we assume ellipsoid uncertainty) and being equal to plus infinity outside. The set of convex risk measures with this specific penalty term leads to a set of coherent risk measures. The concept of coherent risk measures was first introduced by Huber (1981). Later this was generalised to a subset of convex risk measures, see Cheridito et al. (2005) and Föllmer and Schied (2002) for work on this. Cont (2006) quantifies model ambiguity by both coherent and convex sets of risk measures. Jaschke and Küchler (2001) establish the relation that coherent risk measures are equivalent to Good-Deal-Bounds.

1.2 Literature Review: Uncertainty Sets

As already mentioned in the paragraph about robust pricing, we would like to highlight the difference between risk and uncertainty. First of all, in this thesis we use ambiguity and uncertainty interchangeably. Risk (A)² stands for the uncertainty that an event might happen while knowing the probability, *known unknowns*, whereas for uncertainty (B) one does not know the probability, *unknown unknowns*. The Ellsberg paradox (Ellsberg, 1961) states that people prefer known risk over unknown risk. This phenomenon is called ambiguity aversion. People differ among their risk attitudes. If one is indifferent between the bet and the certainty equivalent, he is risk-neutral (D). If the risk premium, the difference between the expected payoff and the certainty equivalent, is negative, he is risk-loving (E). Lastly, a risk-averse person (C) accepts a certain payment lower than the expected payoff of the gamble.

Ben-Tal et al. (2013) focus on ϕ -divergences (F) in robust optimisation. General optimisation problems, examples ranging from finance to operations research, are solved robustly over an uncertainty region. The uncertainty region is identified by the confidence set using a specific ϕ -divergence function. At node (D) of the overview of pricing methods in Figure 1.1 the concept is introduced from an incompleteness viewpoint. The described divergence functions are Kullback-Leibler divergence, Burg entropy, J-divergence, (modified) χ^2 -distance, Hellinger distance, variation distance and χ -divergence of order $\theta > 1$, and Cressie-Read divergence.

Breuer and Csiszár (2013b) base stress tests on plausible sets, that do not include scenarios that are too implausible and do include the dangerous ones. These scenarios are obtained by considering mixed scenarios, known as risk factor distributions. The Kullback-Leibler (J) divergence from these distributions has to

²Letter refers to the corresponding box in Figure 1.2.

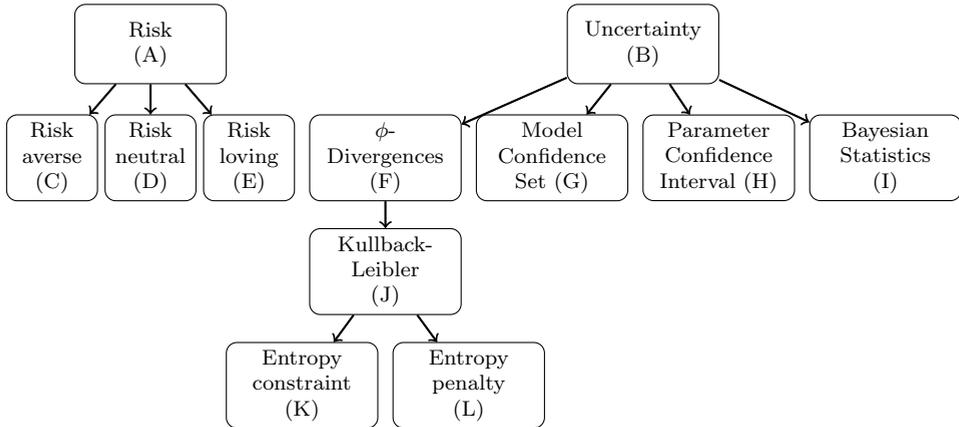


Figure 1.2: Overview of Uncertainty Sets

fall within an uncertainty ball, where the radius is assumed to coincide with the 1% Tail-Value-at-Risk or the bound can be calculated based on historical data. However, for all divergences the critical value that distinguishes a plausible from an implausible model is not quantified. Merely the optimisation problem that incorporates an ambiguity measure is the focus of these papers.

Hansen and Sargent (2008) (hereafter H&S) motivate their approach of uncertainty to robust optimisation by choosing models with bounded entropy (i.e. with bounded Kullback-Leibler divergence). However, when implementing their method they make a subtle switch: they replace the endogenous Lagrange multiplier of the entropy constraint (K), with a fixed entropy penalty (L). On the one hand this leads to a fundamentally different class of optimisation problems on finite horizons. On the other hand, by “giving up” the explicit entropy constraint H&S obtain a time-consistent operator, which is a desirable property for dynamic optimisation problems. Still the strength of preference towards robustness has to be determined. H&S start with an optimisation problem, then they pick a Lagrange multiplier and calculate the worst-case path which depends on the multiplier and the specific optimisation problem. Next they calculate the Type I and II errors for the specific multiplier and the associated worst-case path. If the probability of the average of the two incorrect rejections is too high, then the worst-case choice from mother nature is too extreme, with other words too far from the approximating model. Therefore it is deemed unlikely that these two models cannot be distinguished from each other. Hence this multiplier is rejected and will not be-

long to the set of alternatives. By this procedure one plausible worst-case based multiplier is selected rather than a set. The main difference with our approach is the order of the procedure. H&S start with an explicit optimisation problem whereas we focus on the creation of the set of plausible alternative models that can be applied to and is independent of the particular choice of the optimisation problem. Moreover, still the size of the uncertainty set which coincides with the preference towards robustness has to be identified.

Hansen et al. (2011) introduce the concept of Model Confidence Set (MCS) (G). This method, or actually algorithm, is a sequential method that starts with a collection of possible models and ends with a subset of these that contain the best models with a given level of confidence. Best is in terms of a *chosen* test statistic. In comparison with the indistinguishable method discussed in Chapter 4, the MCS method starts with a collection of competing models. Rather than introducing a discrete number of both plausible and implausible models, the indistinguishable method is exempted from this initial input.

The MCS method is based on an equivalence test and an elimination rule. If the equivalence test indicates that the set of models at hand are not equivalent, in our terms are not indistinguishable, then at least one of the models performs worse than the others and should be removed. The elimination rule performs the execution step repeatedly until the set consists of only those models that are equivalently good for a given confidence level. Hence the MCS method uses the data and evaluates the test sequentially at time T . The method in Chapter 4 is based on a hypothetical test.

For all alternative models identifiable by different constant drifts, a comparison with the confidence interval (H) of the parameter estimate can be made. For historical data over the past T years, the estimated drift is $\hat{\lambda} = W(T)/T$, where $W(T)$ is the realisation of the Brownian motion path at time T . The standard error for this estimator is given by $1/\sqrt{T}$, and the 95% confidence interval for $\hat{\lambda}$ is given by $[\hat{\lambda} - 1.96/\sqrt{T}, \hat{\lambda} + 1.96/\sqrt{T}]$. Hence, for the baseline model $\lambda = 0$ one would consider all models with a constant λ that satisfy $|\lambda| < 1.96/\sqrt{T}$. The confidence interval approach is limited specifically to parameter uncertainty for the constant drift λ . In detail, the parameter approach starts with a path of observed $W(t)$ for $t \in \{1, \dots, T\}$. Based on this the $\hat{\lambda}$ is estimated. This implies that the test is performed ex post.

The confidence interval $\lambda \in CI_{95\%}(\hat{\lambda})$ shows the region of acceptance. If 0 is included, the hypothesis can be accepted. The determination which alternatives

are all accepted is only composed afterwards. To prevent misunderstandings, T is here the amount of data whereas in the indistinguishable approach of Chapter 4 it represents the point in the future at which one would have T years of additional information.

In Bayesian statistics (I) parameters are assumed to be realisations of probability distribution. One can indicate his prior belief about the variable by a prior distribution. The observed data is summarised by the likelihood function. The posterior distribution combines both the prior distribution with the likelihood function. Contrary to the frequentist approach that treats the parameters as constants, the posterior distribution indicates the uncertainty of the parameter conditionally on the data. Strictly speaking, if the prior distribution is known, the Bayesian concept belongs to risk. Whereas if there are multiple priors, it belongs to uncertainty. However, the strict distinction based on the definition of risk and uncertainty is often relaxed and both terms are used interchangeably. In Chapter 2 we apply the Bayesian updating procedure to a term structure model. The max-min expected utility from Gilboa and Schmeidler (1989) is another Bayesian method that deals with model ambiguity by allocating multiple priors to a set of outcomes. Palm and Zellner (1992) investigated which information on the priors of the models is required to solve the decision problem. We described the multiple prior method in the context of market incompleteness at node (I) of the literature review of pricing methods.

1.3 Outline

In Chapter 2 the uncertainty is quantified by Bayesian methods. The incomplete market that underlies the term structure of interest rates is investigated by deriving the posterior distributions of the uncertain parameters. For the mean-reverting model we show the extrapolation based on American and European data.

The goal of Chapter 3 is to price a non-traded item robustly. The robustness is incorporated by the max-min concept, where the malevolent mother nature selects her model from an ellipsoid that indicates the uncertainty around the baseline model. We prove that this price always exists and is unique.

Chapter 4 deals with the quantification of uncertainty. *Ex ante*, we determine the set of models that cannot be distinguished from each other with enough power for a given probability on the Type I error. The critical value can be used to quantify the bounds on ϕ -divergences. Chapter 4 considers a fundamentally large

class of models by allowing for stochastic alternatives.

Chapter 5 is a general conclusion of this thesis. The detailed conclusions can be found at the end of each separate chapter. The valorisation in Chapter 6 states the societal relevance of this research. This is followed by the references, a Dutch summary and *Model Uncertainty: The Effect on Robustness, Estimation and Stochastic Optimisation* ends with a short curriculum vitae.

Chapter 2

What does a Term Structure Model imply about very Long-Term Discount Rates?

We estimate a term structure model on interest rate data with maturities up to 20 years and then extrapolate the yield curve to maturities up to 100 years. Such model based extrapolations are motivated by limited liquidity of very long-dated fixed income instruments. The extrapolation appears mainly driven by the near unit root of the level factor under the risk-neutral measure. In a no-arbitrage term structure model this leads to a strong convexity effect, which implies that extrapolated yield curves are generally upward sloping for maturities longer than 20 years before eventually bending steeply downwards. Our estimates use Bayesian methods. Based on Euro swap rates from 2002 through 2013 and US swap rates from 1998 through 2014, we identify the uncertainty at the long end of the yield curve. The prior is informative on mean-reversion parameters and imposes a zero lower bound on the unconditional means.¹

¹This chapter is based on the paper Balter et al. (2015).

2.1 Introduction

Long maturity discount rates are an essential input for valuing the liabilities of pension funds and insurance companies. Life insurance or pension fund liabilities can be as long as 100 years, whereas the available liquid instruments in the market have much shorter maturities. In most countries market rates can only be observed for maturities up to 20 or 30 years for government debt. Swap rates are available for maturities up to 50 years, but there are doubts about the liquidity of the longest maturities. Fair value, or market-consistent valuation, requires discount rates that are close to market rates, but free of liquidity effects.²

For this purpose various methods have been proposed to extend an observed yield curve. Assuming that market rates are liquid up to a “last liquid point” with maturity of 20 years (say), how should such a yield curve be extended to maturities up to 100 years? One option is the use of numerical extrapolation techniques. A prominent example is the Smith-Wilson methodology adopted by EIOPA³, which extrapolates the forward rate curve using exponential functions.⁴ The extrapolation method provides a smooth extension from the yield at the last liquid point to an externally specified ultimate forward rate (UFR) and a chosen convergence speed parameter. Other methods, such as the Nelson-Siegel methodology, would first fit level, slope and curvature factors using data on the liquid part of the yield curve and then extend the yield curve with the parameters of the fitted model.⁵ In the Smith-Wilson methodology the yield curve always converges to the same constant, whereas in the Nelson-Siegel model long rates converge to a time-varying level factor estimated from the current term structure.

A problem with these methods is that the extension is based on curve fitting, and not on a formal term structure model. The Nelson-Siegel model can be made arbitrage free by adding a yield adjustment term, like in Christensen, Diebold, and Rudebusch (2011), but this adjustment will push very long-term yields to minus infinity.⁶ The strong downward pressure on long-term yields is caused by the unit

²Quoting from From Moody’s Analytics (May 2013): “Fair value is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction (that is, not a forced liquidation or distressed sale) between market participants at the measurement date under current market conditions.”

³European Insurance and Occupational Pensions Authority

⁴See the EIOPA notes at eiopa.europa.eu/fileadmin/tx_dam/files/consultations/QIS/QIS5/ceiops-paper-extrapolation-risk-free-rates-en-20100802.pdf for references and details of the method.

⁵See Diebold and Rudebusch (2013) for a textbook treatment of the Nelson-Siegel model for fitting term structure data.

⁶See the leading adjustment term I_1 in Appendix B in Christensen et al. (2011)

root of the level factor under the risk-neutral measure that drives the convexity adjustment. The arbitrage-free Nelson-Siegel model is a member of the class of essentially affine Gaussian term structure models (Duffee, 2002). Other models in that class do not necessarily have a unit root in the risk-neutral dynamics, and will thus have a smaller convexity adjustment at the very long end of the yield curve.

The existence of a convexity effect is the main argument for using a formal term structure model for the extrapolation. It implies that yields do not always converge monotonically from the last liquid point towards the ultimate yield. Starting from low interest rate levels, the convergence will follow a hump shape, in which yields first overshoot the ultimate yield before finally decreasing slowly towards the limit. With current low interest rates such an extrapolation will result in a much higher level for long-term yields than either the Nelson-Siegel or the Smith-Wilson extrapolation.

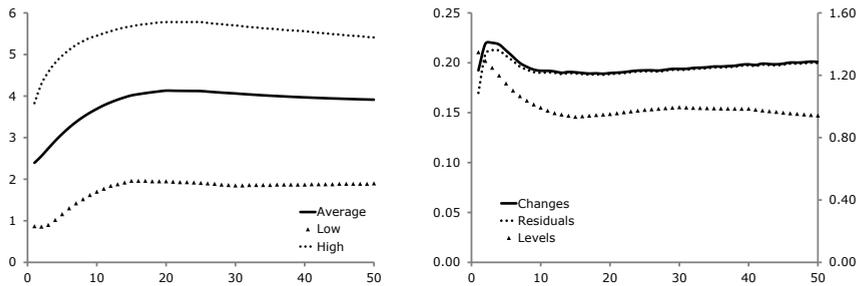
A secondary aim of the chapter is to quantify the uncertainty around extrapolated yields given a formal term structure model. How much can we learn from time series of observed yields with medium to long-term maturities (5 to 20 years) about the shape of the yield curve at very long maturities (between 20 and 100 years). More explicitly, what can we infer about the three crucial elements at the long end of the yield curve: the ultimate forward rate, the convergence speed towards the UFR, and the convexity?

The simplest model to estimate these quantities is a single factor Vasicek model. The model can be parametrised by three key parameters: the ultimate yield, mean-reversion (or convergence speed) and volatility. It therefore directly addresses the main challenges for extrapolating a yield curve. The single factor model cannot fit the complex curvatures at the short end (1 month to 2 years) of the yield curve, but performs reasonably for long maturities. We therefore estimate the parameters using data on 5- and 20-year maturities.

In a Bayesian analysis using Euro swap rate data we find that the mean-reversion of the level factor is non-zero, but with considerable probability mass very close to the unit root (under the risk-neutral measure). That means that convexity effects are important, but not so large that they drive the limiting yield to minus infinity. In our extrapolations the yield curve remains upward sloping for maturities up to 100 years. As expected the Vasicek extrapolation leads to a higher level of very long-term yields than the Nelson-Siegel or Smith-Wilson methods. The uncertainty around the extrapolated yields increases with the maturity. For

Figure 2.1: Euro swap rates

In the left panel the solid line shows the average yield curve over the period from January 2002 to September 2013. The dotted line and the triangles show the yield curves with the minimum and maximum rate at the 20 years maturity. The right panel shows the volatility of yield level, yield changes and the one month prediction errors from an AR(1) model. The horizontal axis in both panels is the maturity in years. The vertical axis unit is percent per year. The right vertical axis is for levels; the left vertical axis is for changes and prediction errors.



market conditions prevailing in the fall of 2013 the Nelson-Siegel and Smith-Wilson extrapolations are at the lower end of the 95% highest posterior density region. The same analysis for the US market prevails the better fit to the one factor model. Because the model is closer to catch the US data the errors are smaller, therefore the mass at the unit root is much smaller than in the EU setting. However the total uncertainty around the extrapolated yields shows to be robust.

2.2 Data

Our yield curve data consists of a monthly panel of discount rates from the website of the Bundesbank.⁷ These yield curve data are constructed from Euro swap rates with maturities ranging from 1 to 50 years. The sample period is from January 2002 to September 2013 resulting in 141 data-points per maturity.

Figure 2.1 provides an overview of the data. The average term structure is increasing until the 20-year maturity, after which it becomes slightly downward sloping for longer maturities. The yield curve has fluctuated substantially over the twelve year sample period. The figure shows a curve from the beginning of the sample (March 2002) when the curve has a similar shape as the average, but at a

⁷http://www.bundesbank.de/Navigation/EN/Statistics/Time_series_databases

1.5% higher level. The lowest long-term yields are from May 2012, where the 50 years maturity yield is just below 2%. With such dominant parallel shifts of yield curves, it will be difficult to fit a long-term common ultimate forward rate to these data. It is definitely not reached before a 50 years maturity and convergence to a common ultimate yield requires a very low level of implied mean-reversion.

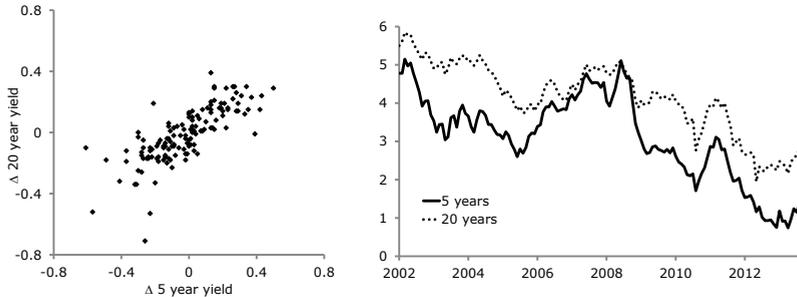
The right-hand panel of Figure 2.1 shows the volatility of yields. When measuring volatility as the standard deviation of yield levels we see a quickly decreasing volatility structure up to maturities of 12 years, after which the volatility stabilizes. Within an affine term structure model this points to at least two factors. One factor would be a level factor that is close to a random walk under the risk-neutral measure, while the second factor is a stationary factor with strong mean-reversion. Due to the strong mean-reversion the second factor has a negligible influence on yields with maturities longer than 10 years. Since the shortest maturity in this data set is one year, it is difficult to identify a third factor.

The same figure also shows the volatility of yield changes. In a one-factor model it should have the same shape as the level volatility, and only the scaling should be different. In a multi-factor model the shape for short and medium-term maturities can be very different from the shape of the level volatilities. The figure shows the familiar hump-shaped volatility, where the volatility peaks at the three year maturity, and then starts a gradual decrease. The initial hump shape for shorter to intermediate maturities can be explained by a two-factor model. The gradual downward sloping volatility is consistent with a slowly mean-reverting level factor. Most puzzling is the upward sloping pattern from a maturity of 15 years onwards, which cannot be explained by standard term structure models. The affine model, for example, implies that the volatility curve is downward sloping for longer maturities. Very long-dated swap prices may contain more noise because the market at these long-term maturities is less liquid. This suggests that the 20- years rate may be a good reference for the last liquid point to start the extrapolation.

In our econometric model we will work with time series data and the discount yields for maturities 5 and 20 years. The time series are shown in Figure 2.2. The scatter diagram suggest that the single factor assumption is not too far off. From the time series plot it appears that both rates have a slight negative trend over the last decade, so that it may be hard to estimate an unconditional mean from these time series.

Figure 2.2: Time series data

The figure shows the monthly time series data for the 5- and 20-year maturity discount yields, both as a time series graph and as a scatter diagram for the first differences of the yields.



2.3 Term Structure Models

Our basic model is the essentially affine term structure model introduced by Duffee (2002) as an extension of the Duffee and Kan (1996) class of affine term structure models. Dai and Singleton (2000) show that the canonical structure of a Gaussian affine model can be written as

$$y_t(\tau) = a(\tau) + \sum_{j=1}^K b_j(\tau)x_{jt} \quad (2.1)$$

where $y_t(\tau)$ is the yield of a discount bond at time t with time to maturity τ ; $a(\tau)$ and $b_j(\tau)$ are function of the time to maturity and the underlying parameters of the model; and x_{jt} are time-varying factors that follow the Ornstein-Uhlenbeck processes

$$dx_j = \tilde{\kappa}_j(\tilde{\mu}_j - x_j)dt + \sigma_j' d\tilde{W} \quad (2.2)$$

with \tilde{W} a K -dimensional Brownian motion under the risk-neutral density (\mathbb{Q} measure) that is normal distributed with mean 0 and the covariance matrix equals the identity matrix and σ_j a vector of volatilities. Multiple factor term structure models are characterized by different mean-reversion parameters $\tilde{\kappa}_j$ that determine the

$b_j(\tau)$ functions in (2.1)

$$b_j(\tau) = \frac{1 - e^{-\tilde{\kappa}_j \tau}}{\tilde{\kappa}_j \tau} \quad (2.3)$$

The larger the $\tilde{\kappa}_j$, the less the impact of the factor on long-term yields. For typical estimates of a three factor model, more than 95% of the variation at maturities longer than 5 years is explained by the first factor, usually referred to as the level factor.

2.3.1 Vasicek Model

Since our aim is to extrapolate the yield curve beyond maturities of 20 years, using data in the segment between 5 and 20 years, we specialize our model to a single factor. For the single factor ‘‘Vasicek’’ model we drop the subscript j . Using the explicit solution for $a(\tau)$ the Vasicek yield curve takes the form

$$y_t(\tau) = \theta + b(\tau) (x_t - \theta) + \frac{1}{2} \omega^2 \tau b(\tau)^2 \quad (2.4)$$

where

$$\begin{aligned} b(\tau) &= \frac{1 - e^{-\tilde{\kappa} \tau}}{\tilde{\kappa} \tau} \\ \omega^2 &= \frac{\sigma^2}{2\tilde{\kappa}} \\ \theta &= \tilde{\mu} - \frac{\omega^2}{\tilde{\kappa}} \end{aligned} \quad (2.5)$$

See Appendix 2.A for the full derivation of the expression for the yields. The function $b(\tau)$ defines the volatility of long-term yields relative to the level factor x_t , ω^2 is the unconditional variance of the factor, and θ is the limiting yield $y_t(\tau)$ when $\tau \rightarrow \infty$. In the Vasicek model the constant θ is both the ultimate yield as well as the ultimate forward rate. It is equal to the unconditional mean of the risk-neutral distribution of the factor minus the infinite horizon convexity adjustment. All zero rates are a weighted average of the factor and the ultimate long-term yield plus a convexity adjustment.

Equation (2.4) provides the cross-sectional relation among yields with different maturities. For the time series dynamics of the yields we need to transform from the \mathbb{Q} dynamics to the physical measure \mathbb{P} . This requires an assumption on the price of risk and a stochastic discount factor. Following Duffee (2002) we make

the essentially affine assumption and specify the stochastic discount factor as

$$\frac{d\Lambda_t}{\Lambda_t} = -x_t dt - \lambda_t dW_t \quad (2.6)$$

with

$$\lambda_t = \Lambda_0 + \Lambda_1 x_t \quad (2.7)$$

With this assumption the time series process for the factor becomes

$$dx_t = \kappa(\mu - x_t)dt + \sigma dW_t \quad (2.8)$$

and the parameters under the \mathbb{P} and \mathbb{Q} measures are related by

$$\tilde{\kappa} = \kappa + \sigma \Lambda_1 \quad (2.9)$$

$$\tilde{\mu} \tilde{\kappa} = \mu \kappa - \sigma \Lambda_0 \quad (2.10)$$

In Appendix 2.B this relation is derived. Combining (2.4) and (2.8) provides an expression for the time series behaviour of different yields $y_t(\tau)$, which are the starting point for the econometric analysis.

2.3.2 Extrapolation

In general extrapolation uses forward rates and the identity

$$y_t(s) = \frac{1}{s} \int_0^s f_t(u) du \quad (2.11)$$

with $f_t(\tau)$ the instantaneous forward rate at time t for time $t + \tau$. If we have reliable data for the term structure up to the reference maturity τ^* (“last liquid point”) the extension to maturities $s > \tau^*$ follows as

$$y_t(s) = \frac{1}{s} \left(\tau^* y_t^* + \int_{\tau^*}^s f(u) du \right) \quad (2.12)$$

where $y_t^* = y_t(\tau^*)$ is the observed yield at maturity τ^* .⁸ The extrapolation ensures continuity of the extended yield curve at the last liquid point. For the one factor

⁸In practice most methods replace the instantaneous forward rate by one year forward rates $F_t(\tau, \tau + 1)$ in which case the integral in (2.12) becomes the sum $\sum_{i=\tau^*}^{s-1} F_t(i, i + 1)$. For expositional purposes we stick with the instantaneous representation.

Vasicek model the forward rate curve itself is a function of the single state variable x_t and we can use y_t^* to solve for the state variable using (2.4). Assuming that y_t^* is exactly on the Vasicek curve, we can simply use (2.4) twice to first express $y_t(s)$ as a function of x_t and then again to replace x_t by the reference yield y_t^* . The result is

$$y_t(s) = \frac{b(s)}{b^*} y_t^* + \left(1 - \frac{b(s)}{b^*}\right) \theta + C^*(s) \quad (2.13)$$

with b^* a shorthand notation for $b(\tau^*)$ and where

$$C^*(s) = \frac{1}{2} \omega^2 b(s) (s b(s) - \tau^* b^*) \quad (2.14)$$

First solving for the forward rate curve and computing (2.12) leads to the same result as long as we use y_t^* as the only input for the extrapolation. More sophisticated methods allow for a slight measurement error in the observed rate y_t^* and extract the state variables from multiple maturities using a Kalman filter. Assuming that the measurement error is small this will only lead to minor changes in the estimate for the state variable x_t and therefore not cause major changes in the extrapolation.

The extrapolated yields are a weighted average of the last liquid point y_t^* and the ultimate yield θ plus a convexity adjustment. The convergence speed is measured by the relative volatility $b(s)/b^*$. This is a decreasing function of s , starting at one for $s = \tau^*$ and gradually moving towards zero as s increases. The convexity adjustment $C^*(s)$ is always positive. Hence, if $y_t^* = \theta$, the extrapolation first moves $y_t(s)$ above the ultimate yield, before slowly converging downwards to the ultimate yield θ again. The Vasicek extrapolation will thus be markedly different from a simple weighted average of the last liquid point and an ultimate forward rate. It will often imply a steeper upward sloping yield curve before eventually flattening (or decreasing) towards the ultimate yield.

In general all long-term yields have a negative convexity exposure. We obtain the positive term in (2.14), because both the reference yield y_t^* and ultimate yield θ themselves are negatively affected by convexity and the convexity in (2.13) is measured relative to these yields. An equivalent expression for the extrapolated yield that explicitly shows the convexity effect in the ultimate yield is

$$y_t(s) = \frac{b(s)}{b^*} y_t^* + \left(1 - \frac{b(s)}{b^*}\right) \left(\tilde{\mu} - \frac{\omega^2}{\tilde{\kappa}}\right) + C^*(s) \quad (2.15)$$

which uses (2.5) to replace θ by the risk-neutral mean $\tilde{\mu}$ of the spot rate. The convexity in θ increases with maturity and reaches a minimum of $-\omega^2/\tilde{\kappa}$ at infinite maturity. The convexity term scales with the unconditional variance ω^2 of the risk-neutral factor dynamics. If the mean-reversion $\tilde{\kappa}$ goes to zero, meaning a true level factor, the variance will tend to infinity. But even if $\tilde{\kappa}$ does not move all the way to its limit, the convexity in the ultimate yield can be substantial due to the additional $\tilde{\kappa}$ in the numerator. For small $\tilde{\kappa}$ it can become so big that yields will be negative for large s and converge to a negative θ (with fixed $\tilde{\mu}$). This will occur, for example, in the arbitrage-free version of the Nelson-Siegel model of Christensen, Diebold and Rudebusch (2011).

The convergence towards the ultimate forward rate is measured by how quickly the forward rate converges to θ . For the Vasicek model the forward rates are given by

$$f_t(s) = \theta + e^{-\tilde{\kappa}s}(x_t - \theta) + \omega^2 e^{-\tilde{\kappa}s} sb(s) \quad (2.16)$$

and thus θ is the ultimate forward rate. Using (2.16) to express the forward rate $f_t(s)$ relative to the forward rate at maturity τ^* we have

$$f_t(s) = \theta + e^{\tilde{\kappa}(\tau^* - s)}(f_t^* - \theta) + \omega^2 e^{-\tilde{\kappa}s}(sb(s) - \tau^* b^*) \quad (2.17)$$

The mean-reversion parameter $\tilde{\kappa}$ can thus alternatively be labelled as the convergence rate of the forward curve. Using the forward rate f_t^* to solve for the state variable x_t is equivalent to using y_t^* , if the Vasicek model would fit perfectly. With measurement errors the yield y_t^* will most likely provide a more accurate estimate of the factor than the forward rate f_t^* . Since the forward rate is a function of the derivative of the yield with respect to maturity, it is more sensitive to measurement error than the yield itself.

2.4 Econometric Model

To extrapolate the yield curve we would only need the parameters $(\tilde{\kappa}, \tilde{\mu}, \sigma^2)$ of the risk-neutral distribution \mathbb{Q} . These parameters can be identified from a single cross-section, but this would be very inefficient. If these parameters are time invariant, a panel estimate from multiple cross-sections increases efficiency.

Further information on the parameters σ^2 and $\tilde{\kappa}$ can be obtained from time series data, since these two parameters determine the volatility of long-term yields.

Since volatility parameters are typically estimated with more precision than location parameters, time series data may be best for estimating σ^2 and $\tilde{\kappa}$. Since $\tilde{\kappa}$ determines the relative volatility of different yields through the function $b(\tau)$, we need time series data on at least two maturities to identify $\tilde{\kappa}$ from volatility moments. For the estimation we use the time series data for both the 5-year and 20-year maturity discount yields. The 5-year yield is the shortest maturity that seems uncontaminated by additional factors, while the 20-years rate is the longest one that still is on the downward sloping part of the volatility curve in the empirical data. By choosing the two maturities relatively far apart we also include as much of the cross sectional information as possible.⁹

The parameters on the physical measure \mathbb{P} are fully identified from time series data. For a one-factor model a single time series would be enough to identify the model parameters (κ, μ, σ) . The parameters under the \mathbb{P} and \mathbb{Q} measures are connected through the price of risk function λ_t , which contains two free parameters. Hence there is an overlap between the two sets of parameters, which we indicate by including the common parameter σ^2 in both parameters sets. Since the relation between the two measures depends on the market price of risk, knowing Λ_0 and Λ_1 is equivalent to knowing $\tilde{\kappa}$ and $\tilde{\mu}$. With a single interest rate time series it is impossible to identify the cross-sectional parameters. With multiple maturities the parameters are over-identified.

Henceforth, we consider a single factor model for two maturities as following the restricted VAR(1) process

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \end{pmatrix} = \begin{pmatrix} y_{t-h}(\tau_1) \\ y_{t-h}(\tau_2) \end{pmatrix} - \alpha \begin{pmatrix} y_{t-h}(\tau_1) - m(\tau_1) \\ y_{t-h}(\tau_2) - m(\tau_2) \end{pmatrix} + \begin{pmatrix} e_t(\tau_1) \\ e_t(\tau_2) \end{pmatrix} \quad (2.18)$$

where h is the length of the time interval between two observations (one month, $h = 1/12$), $m(\tau)$ is the unconditional mean of a discount rate with maturity τ , and the shocks $e_t(\tau)$ are normally distributed with mean zero and covariance matrix Σ . The mean-reversion parameter α is the discrete time equivalent of the continuous time mean-reversion parameter κ ,

$$\alpha = 1 - e^{-\kappa h} \quad (2.19)$$

In this bivariate process the mean-reversion parameter α should be the same for the two different interest rates.

⁹In the robustness analysis we analyse the effect of using different maturities.

According to (2.4) and (2.8) the error covariance matrix takes the form

$$\boldsymbol{\Sigma}^* = s_h^2 \sigma^2 \begin{pmatrix} b_1^2 & b_1 b_2 \\ b_1 b_2 & b_2^2 \end{pmatrix} \quad (2.20)$$

where $b_i = b(\tau_i)$ and where s_h^2 is a scaling constant that links the discrete time model to the continuous time parametrisation through

$$s_h^2 = \frac{1 - e^{-2\kappa h}}{2\kappa} \quad (2.21)$$

In an Euler discretisation we would have $\alpha = \kappa h$ and $s_h^2 = h$. Since we are using a single factor model, the matrix $\boldsymbol{\Sigma}^*$ has rank one. To avoid the stochastic singularity in the estimation it is common to assume a small measurement (or model) error. This can be done through a formal measurement equation and a Kalman filter model as in De Jong (2000). Since here we only have two time series, we take the simpler approach by adding a small positive variance to the diagonal elements of $\boldsymbol{\Sigma}^*$,

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^* + s_h^2 \eta^2 \mathbf{I} \quad (2.22)$$

With this specification the covariance matrix is a function of three parameters: σ^2 , $\tilde{\kappa}$, and η^2 . The parameter $\tilde{\kappa}$ enters through the function $b(\tau)$. The mean-reversion under the risk-neutral measure is identified through the covariance matrix, since $b(\tau)$ defines the volatility of a bond with maturity τ and $\tilde{\kappa}$ is thus primarily a volatility parameter. The parametrisation with η^2 interpreted as a measurement error variance will only be credible if η^2 is small relative to the overall volatility of the shocks \mathbf{e}_t . Large measurement error would not only cast doubt on the model specification, but would also imply that the regressors $y_{t-h}(\tau_i)$ are subject to errors-in-variables.

When we estimate the model we obtain estimates of the covariance matrix $\boldsymbol{\Sigma}$. From the three elements in $\boldsymbol{\Sigma}$ we solve for the three model parameters σ^2 , η^2 and $\tilde{\kappa}$. Estimates of $\tilde{\kappa}$ are only admissible if the implied $\tilde{\kappa} \geq 0$. Non-negativeness of $\tilde{\kappa}$ requires the following two conditions to hold

$$\begin{aligned} \sigma_{11} &\geq \sigma_{22} \\ \sigma_{21} &\geq 0 \end{aligned} \quad (2.23)$$

The first condition imposes that volatility decreases monotonically with maturity. The second condition states that shocks to long-term interest rates are positively correlated. In Appendix 2.C we derive the further admissibility condition

$$\sigma_{11} - \sigma_{22} \leq \frac{15}{4}\sigma_{21} \quad (2.24)$$

which is specific to the maturities $\tau_1 = 5$ and $\tau_2 = 20$.

The intercepts $m(\tau_i)$ in (2.18) are related to the parameters μ and $\tilde{\mu}$ in the Vasicek model, or equivalently μ and the ultimate forward rate θ ,

$$\begin{pmatrix} m(\tau_1) \\ m(\tau_2) \end{pmatrix} = \begin{pmatrix} b_1 & 1 - b_1 \\ b_2 & 1 - b_2 \end{pmatrix} \begin{pmatrix} \mu \\ \theta \end{pmatrix} + \frac{\sigma^2}{4\tilde{\kappa}} \begin{pmatrix} \tau_1 b_1^2 \\ \tau_2 b_2^2 \end{pmatrix} \quad (2.25)$$

Since $\tilde{\kappa}$ and σ^2 are already identified from the covariance matrix Σ , these two equations uniquely identify μ and θ . Inverting (2.25) is problematic when $\tilde{\kappa} \rightarrow 0$. For small values of $\tilde{\kappa}$ the system becomes almost singular because $b_1 \rightarrow b_2$, while at the same time the intercepts go to infinity (unless $\sigma^2 \rightarrow 0$). The ultimate yield may therefore be very difficult to identify from the data if the risk-neutral mean-reversion is small.

Assuming normality and time series independence for the error terms, we obtain a normal likelihood function from which we can estimate the six unknown parameters κ , $\tilde{\kappa}$, μ , $\tilde{\mu}$, σ^2 and η^2 . The parameters are also exactly identified from the reduced form parameters α , \mathbf{m} and Σ .

2.5 Maximum Likelihood Estimates

For a first formal evaluation of the model we estimate the parameters by conditional maximum likelihood, where we condition on the initial observation. The results are in Table 2.1.

In contrast to many other studies the model is estimated with data on medium- to long-term maturities. Results are similar, however, to what has been found for other sample periods and countries. The time series mean-reversion κ , corresponding to a monthly first order autocorrelation of 0.975, is not significantly different from zero. A unit root cannot be rejected. We also find, like *e.g.* De Jong (2000), that the risk-neutral mean-reversion parameter is much smaller than the time series mean-reversion. It is also not significantly different from zero, even though the asymptotic standard error is very small. Due to the near unit root for the time

Table 2.1: ML parameter estimates

The table reports conditional Maximum Likelihood estimates for the bivariate model using interest rates with maturities of 5 and 20 years. The asymptotic standard errors (“se”) are from the Hessian of the log-likelihood function. The columns on the right side of the table are estimated under the restriction $\tilde{\kappa} = \kappa$. All parameters are reported in their natural units with time measured in years. The last line is the value of the log-likelihood function.

	Estimate	se	Estimate	se
$\tilde{\kappa}$	0.0202	0.0101	0.0210	0.0103
κ	0.3023	0.1685	= $\tilde{\kappa}$	
$\tilde{\mu}$	0.1338	0.2677	0.2034	0.4720
μ	0.0155	0.0102	-0.1317	0.1392
θ	0.0717	0.0324	-0.1090	0.4541
Λ_0	0.2940	0.7668	-1.0199	1.5508
Λ_1	-40.556	24.157	0	
σ^2	4.710×10^{-5}	1.313×10^{-5}	4.744×10^{-5}	8.374×10^{-6}
ω^2	1.190×10^{-3}	4.928×10^{-4}	1.113×10^{-3}	4.568×10^{-4}
η^2	1.099×10^{-5}	1.315×10^{-6}	1.093×10^{-5}	1.306×10^{-6}
$\ln L$	1665.10		1663.67	

series, the estimate of the unconditional mean μ at the short end of the yield curve is very imprecise. With a standard error of around 1% it is difficult to anchor an average interest rate. The long-term yield θ is even more imprecise, partly because of the same unit root problem, and partly because of the additional uncertainty on the long-term convexity in the term structure. The implied parameters for the price of risk are both insignificant. The imprecision in Λ_0 and Λ_1 can be reduced by setting $\Lambda_1 = 0$, which would also imply that $\tilde{\kappa} = \kappa$. Since $\Lambda_1 = (\tilde{\kappa} - \kappa)/\sigma$ the insignificance of Λ_1 suggests that this restriction cannot be rejected.

Re-estimating under the restriction $\tilde{\kappa} = \kappa$ results in the same estimate for mean-reversion for the \mathbb{Q} dynamics. Therefore the time series mean-reversion κ is much closer to the unit root than in the unrestricted case. Consistent with the estimates of the unrestricted model, the likelihood ratio statistic shows that the restriction cannot be rejected. A consequence of moving much closer to the unit root under the time series measure \mathbb{P} is that the unconditional means are much harder to estimate and in fact almost unidentified. The ultimate yield θ becomes negative with a much larger asymptotic standard error than in the unrestricted

model where κ is much larger.¹⁰ The restriction does not help in learning about the shape of the yield curve at very long maturities.

The model error variance is so small that results cannot be affected by errors-in-variables problems. Assuming the measurement error to be uncorrelated over time, the measurement error in y_{t-h} is of the order $\frac{1}{2}s_h^2\eta^2 \approx 5 \times 10^{-7}$, which is negligible relative to its true variance $b_t^2\omega^2 \approx 10^{-3}$.

Statistically it is impossible to distinguish between the two sets of parameter estimates. Implications of the parameters are, however, very different due to the differences in the estimate of the ultimate yield. One element on which both estimates agree is the convexity of the yield curve. The parameters $\tilde{\kappa}$ and σ^2 are both estimated with large precision and are independent of the estimates of the unconditional means.

Our estimated cross-sectional mean-reversion $\tilde{\kappa} = 0.02$ implies a convergence of forward rates towards the ultimate forward rate θ that is much slower than the convergence assumed in the Smith-Wilson methodology adopted by EIOPA, which for most periods will be based on a convergence rate of 0.10.

2.6 Bayesian Analysis

Since it is hard to decide which parameters to use, and since the unconditional mean parameters are also highly uncertain, we turn to a Bayesian analysis of the model. This provides a way to account for the parameter uncertainty by computing the extrapolation as a weighted average of different sets of parameters with weights given by the posterior density of the parameters. The Bayesian analysis also allows for informative priors, by which we can impose stationarity of the dynamics under both \mathbb{P} and \mathbb{Q} and add a prior view on the unconditional mean of the interest rates that we analyse.

2.6.1 Priors

We will use mildly informative priors on the long-term means of the interest rate data and on the mean-reversion parameters under \mathbb{P} and \mathbb{Q} . We require all these parameters to be non-negative. For the time series mean-reversion we use

¹⁰Standard errors reported in the table are from the Hessian of the log-likelihood function. Robust standard errors allowing for heteroscedasticity and non-normality do not make a difference to the conclusions in this case.

a truncated normal prior

$$p(\alpha) \sim \text{TN}(h\mu_a, h^2\psi_a^2) \quad (2.26)$$

with $\mu_a = 0$, $\psi_a = 0.2$, such that the truncation implies a prior mean and standard deviation of $E[\alpha] = 0.16h$ and $s[\alpha] = 0.12h$, respectively. With monthly data the prior is centred around a first order autocorrelation of 0.987. The prior is centred close to a unit root, but the relatively tight precision also ensures that the posterior will be away from the unit unless the data are very informative on the dynamics.

To impose that the long-term means $m_i = m(\tau_i)$ of the interest rates are positive we also use a truncated normal. We assume independent priors for m_1 and m_2 that are both specified as

$$p(m_i) \sim \text{TN}(\mu_m, h_m^2) \quad (2.27)$$

with $\mu_m = -0.923$ and $h_m^2 = 0.2$ implying $E[m_i] = 0.04$ and $s[m_i] = 0.039$. The prior ensures that the unconditional means are positive at maturities τ_1 and τ_2 , but it does not guarantee that the unconditional mean is positive for all maturities. Most problematic could be the ultimate yield θ , since it is extremely sensitive to a near unit root in the risk-neutral process.¹¹

For the covariance matrix Σ we assume a truncated inverted Wishart distribution.

$$p(\Sigma^{-1}) \sim \text{TW}(\Psi, \nu) \quad (2.28)$$

where

$$\Psi = 0.01^2 \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix} \quad (2.29)$$

and the degrees of freedom parameter is set to $\nu = 3$, which is slightly above the minimum value of 2. The prior is truncated to the region that satisfies the inequalities (2.23) and (2.24). The prior for $\tilde{\kappa}$ is implicit in the prior for Σ . Even though the prior is almost non-informative for Σ , the prior for $\tilde{\kappa}$ is mildly informative, since it only depends on the ratio $S = \frac{\sigma_{11} - \sigma_{22}}{\sigma_{21}}$. Accounting for the truncation by the inequality constraints, the prior for $\tilde{\kappa}$ has a mean of 0.033 and standard deviation of 0.049 (based on a numerical evaluation).

¹¹Imposing $\theta > 0$ introduces a highly non-linear dependence of \mathbf{m} on Σ .

Since all priors are proper and have well-defined means and variances for the reduced form parameters, the posterior moments for the reduced form parameters also exist. The posterior is not available in closed form due to the truncation and the non-linear parametrisation involving the product $\alpha \mathbf{m}$. Numerically the posterior can be easily obtained through Gibbs sampling, since all conditional posteriors are straightforward except for the truncation. When we sample from the conditional posteriors we reject a draw if it is outside the admissible region. In some cases the probability of accepting a draw can be extremely low. This happens when we need to draw the unconditional means $\mathbf{m} = (m_1 \ m_2)'$ at a point where the mean-reversion parameter α is close to zero. In this case the data are uninformative about the unconditional mean, meaning that we need to draw \mathbf{m} from a distribution that is approximately equal to the prior. Since this is a truncated normal with a negative mean, the probability of obtaining a positive number by drawing from a normal distribution becomes very small. For small α we therefore use the exponential rejection sampling algorithm suggested by Geweke (1991).

2.6.2 Posterior Densities

Results of the Bayesian analysis are reported in Table 2.2. Posterior moments are based on 1 million draws¹² from the MCMC sampler. See Appendix 2.D for the relation between the simulated draws of α , \mathbf{m} and Σ and the parameters $\tilde{\kappa}$, κ , $\tilde{\mu}$, μ , θ , Λ_0 , Λ_1 , σ^2 and η^2 .

The posterior moments for κ and $\tilde{\kappa}$ are close to the maximum likelihood estimates. Due to the prior specification the posterior mean for the time series mean-reversion is a bit closer to the unit root and also a bit more precise. Similar to the ML estimates the mean-reversion under \mathbb{P} is still substantially larger than under \mathbb{Q} . The risk parameter Λ_1 provides a direct comparison on the equality of the two parameters and, as for the ML estimates, the 95% credible interval for Λ_1 contains zero.

Most of the differences with the ML results are in the unconditional means. Although the prior imposes that the unconditional means of the 5- and 20-year yields exist and are positive, this does not guarantee that the unconditional means at other maturities are also positive. The ultimate yield θ is hard to identify from

¹²We checked the convergence based on the ACF, the CUSUM statistic and Geweke's test. The thinning procedure and the burn-in period do not play a role due the extreme choice of the number of draws.

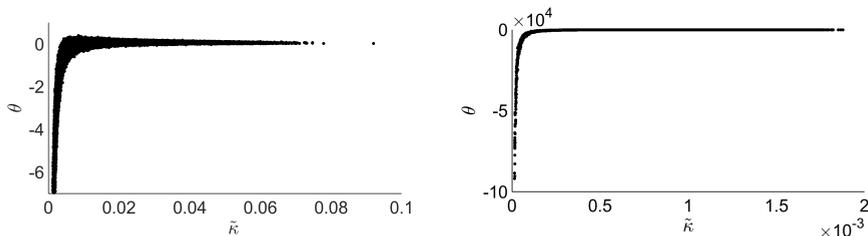
Table 2.2: Posterior moments of parameters

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0203	0.0096	1.647×10^{-3}	0.0382
κ	0.1687	0.0915	7.534×10^{-7}	0.3332
$\tilde{\mu}$	0.2017	1.2164	3.149×10^{-2}	0.4101
μ	0.0139	0.0044	7.232×10^{-3}	0.0212
θ	-7.313	473.8	-0.4038	0.2106
Λ_0	-0.0007	0.2388	-0.4957	0.4881
Λ_1	-21.6018	13.5835	-47.0256	2.7775
σ^2	4.850×10^{-5}	8.395×10^{-6}	3.301×10^{-5}	6.520×10^{-5}
ω^2	0.0021	0.0119	4.9004×10^{-4}	0.0039
η^2	1.086×10^{-5}	1.300×10^{-6}	8.449×10^{-6}	1.347×10^{-5}

The table shows the posterior means, standard deviations and 95% highest posterior density intervals for the parameters of the Vasicek term structure model estimated on interest rates with maturities $\tau_1 = 5$ and $\tau_2 = 20$. Results are based on one million draws from the Gibbs sampler for the parameters α , \mathbf{m} and Σ , which are solved for the structural parameters in the table.

Figure 2.3: Conditional posterior draws of θ given $\tilde{\kappa}$

The figure shows a scatter plot of the draws $\theta^{(i)}$ conditional on $\tilde{\kappa}^{(i)}$. The left-hand panel shows all draws except the 1% smallest values of $\theta^{(i)}$. The right-hand panel shows the 1% smallest values of $\theta^{(i)}$. Note the different scales for the two panels.



the data. Its posterior mean is negative with a huge standard deviation that is due to a few extremely negative outliers when $\tilde{\kappa}$ is close to zero. Figure 2.3 shows a scatter plot of the posterior draws for θ conditional on $\tilde{\kappa}$. The left-side of the figures zooms in on the 1% smallest draws for θ . All of these occur conditional on very small values for $\tilde{\kappa} < 10^{-4}$. Given the large uncertainty on the ultimate yield, it is doubtful if the posterior mean of θ exists with our prior specification. A clear sign that the posterior mean may not exist is the fact the average of the

simulated $\theta^{(i)}$ is far below the lower bound of the 95% HPD. The posterior density is extremely skewed with a very long left tail.¹³

In contrast to the ultimate yield, the posterior on the unconditional variance under \mathbb{Q} is well-behaved. It does have a fat right tail, but the posterior simulation does not produce any of the severe outliers that we encountered for θ . The unconditional variance depends on $\tilde{\kappa}^{-1}$, whereas the ultimate yield depends on $\tilde{\kappa}^{-2}$.

2.7 Results

For the extrapolation we set the reference maturity as $\tau^* = 20$ years. This is the longest maturity in our model and is also the choice of the “last liquid point” made by EIOPA. Below we discuss our Vasicek extrapolation results and the different components that determine the extrapolation, i.e. the convergence speed, the ultimate yield and the convexity. Subsequently we compare the results to the available market data and we compare them with the Smith-Wilson methodology adopted by EIOPA and an extrapolation based on the Nelson-Siegel model.

2.7.1 Vasicek Extrapolation

The extrapolation formula (2.13) consists of three terms. The ratio $b(s)/b^*$ in the first term defines the dependence of the extrapolated yield on the last liquid yield y_t^* and is solely a function of $\tilde{\kappa}$. Figure 2.4 shows the posterior of this ratio as a function of s . At $s = \tau^*$ the ratio is equal to one by construction and therefore not subject to any uncertainty. The posterior mean of $b(s)/b^*$ is decreasing in s . The uncertainty in the ratio increases with s , however. The distribution is right skewed for large s , meaning that there is considerable probability mass for $b(s)/b^*$ remaining very close to one at long maturities.

The second term is the effect of the ultimate yield. It is the product of θ and the weight of the ultimate yield. The parameter θ is very poorly determined by the data, but the large negative outliers for θ occur when $\tilde{\kappa}$ is very close to zero. For small $\tilde{\kappa}$ the weight $(1 - b(s)/b^*)$ will also go to zero, so that the overall effect is unclear. Figure 2.5 shows the posterior mean of the second term. Despite the

¹³The parameter $\tilde{\kappa}$ depends on the error covariance matrix Σ , which follows a Wishart distribution. Since we have sufficiently many time series observations, low order moments of Σ clearly exist. But $\tilde{\kappa}$ is an implicit non-linear function of Σ and therefore its properties cannot be determined analytically. What matters for the existence of the mean of θ are the properties of the ratio $\sigma^2/\tilde{\kappa}^2$, where σ^2 and $\tilde{\kappa}$ are dependent functions of the same matrix Σ .

Figure 2.4: Convergence speed

The figure shows the posterior mean of $b(s)/b^*$ for different values of $s > \tau^* = 20$. The dashed lines define the 95% HPD interval.

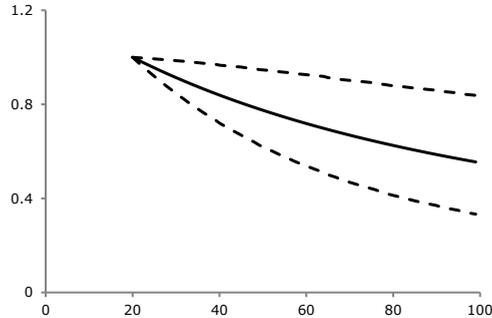
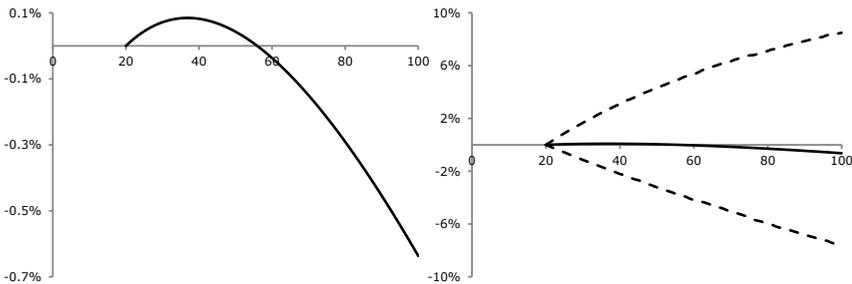


Figure 2.5: Extrapolation towards the ultimate yield

The left panel shows the posterior mean of $(1 - b(s)/b^*)\theta$ for $s > \tau^* = 20$. The right panel adds the 95% highest posterior density region.

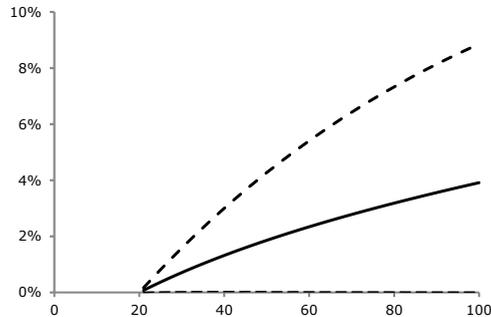


outliers in θ the product $(1 - b(s)/b^*)\theta$ has a very small posterior mean. The total effect of θ on yields up to maturity of 60 years is less than 10 basis points on the overall extrapolated yields $y_t(s)$. The effect becomes negative for longer maturities. Combining the results in Figures 2.4 and Figure 2.5 implies that the weighted average of the reference yield y_t^* and the ultimate yield will thus generally be below y_t^* .

The right-hand panel in Figure 2.5 adds the 95% HPD bounds to the posterior mean. The scaling is very different. Because the bounds are wide, the posterior mean looks like a flat line around zero. Still the range of uncertainty in $(1 - b(s)/b^*)\theta$ is much smaller than the 95% HPD region for θ itself (see Table 2.2).

Figure 2.6: Convexity

The figure shows the posterior mean and 95% HPD region for the convexity term $C^*(s)$ in (2.14).



The interval is also much more symmetric, showing once more that θ and $\tilde{\kappa}$ are highly correlated in the tails.

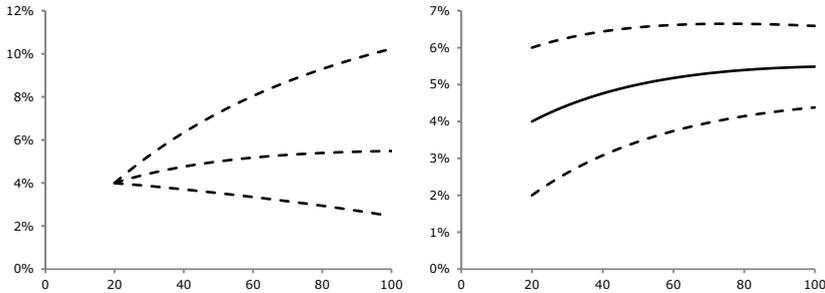
The convexity term $C^*(\tau)$ adds positively to the extrapolated yield. Figure 2.6 shows the posterior mean of $C^*(s)$ for $\tau^* = 20$. Therefore, convexity at $\tau^* = 20$ is zero by construction and positive for all $s > \tau^*$. The posterior mean increases over the entire range to the maturity of 100 years (even though it must decrease to zero by construction as $s \rightarrow \infty$). At the 60 years maturity it will contribute about 2% to the yield curve on top of the weighted average of y_t^* and θ ; at 100 years the effect increases to 4%. As with the ultimate yield, the uncertainty in this term is large. At the lower end of the 95% HPD region the convexity effect is negligible. Small convexity effects coincide with relatively large values of $\tilde{\kappa}$. In these cases the ultimate yield will contribute positively to the extrapolation.

Taking all terms in (2.13) together, Figure 2.7 shows the posterior mean and the 95% HPD interval for the extrapolated yield curve conditional on a 20-years rate equal to $y_t^* = 4\%$. The posterior mean shows a relatively flat yield curve at 4%. In the posterior mean the downward effect of the ultimate yield and the upward convexity effect balance each other. For small $\tilde{\kappa}$ we will often have a very negative θ pushing long-term yields downward, and at the same time a very large positive $C^*(s)$ pulling yields upwards. For large $\tilde{\kappa}$ the opposite happens: negligible convexity and a large θ with substantial weight.

The error bounds show the uncertainty in the extrapolation. The bounds are wide for practical purposes: at 60 years maturity the 95% region covers an

Figure 2.7: Posterior extrapolation

The left panel shows the posterior mean of $y_t(s)$ for $s > \tau^* = 20$ given $y_t^* = 4\%$. The dashed lines define the 95% highest posterior density region. The right panel shows the posterior mean of the extrapolated term structure for different last liquid points y_t^* , being 2% (dashed), 4% (solid) and 6% (dashed) respectively.



interval from 3.5% to 8%. The uncertainty is the joint effect of the uncertainty in all parameters that enter the extrapolation formula. The basic uncertainty relates to $\tilde{\kappa}$, which has strong effects on both θ and ω^2 and the weights of these terms.

The extrapolation is similar for different values of y_t^* . The curves in Figure 2.7 show the posterior means for three different values of y_t^* . The extrapolated curves are almost parallel consistent with a very small convergence rate. Even at the 100 year maturity the curves are still far apart. All curves are slowly upward sloping because of the strong convexity effect (relative to θ).

2.7.2 Extrapolation Errors

Our methodology generally leads to discount yields that are above the observed yields implied by the swap rates. The distance between the observed and extrapolated yield increases with maturity. The extrapolated yield curve is generally upward sloping, whereas the observed yield curve is mostly flat for maturities larger than 20 years. From a time series perspective the errors are positive but small until mid 2008; thereafter they increase quickly to come down towards the end of the sample. Figure 2.8 shows the difference for the 30 years maturity.

The most likely explanation for the positive extrapolation errors is market liquidity and unbalanced demand and supply at long maturities. Pension funds and insurance companies have a strong demand for long-dated fixed income instruments, but supply for these maturities is limited. As a result yields are very low, at a level below what one would obtain from a no-arbitrage term structure model.

Figure 2.8: Errors

The figure shows time series of the difference between extrapolated and observed yield curves at the 30 years maturity. Extrapolation is based on the Vasicek model calibrated to the yield at 20-years maturity.



2.7.3 Alternative Extrapolations

For pension funds and insurance companies recent developments about pricing of long-term obligations is under debate. In some countries the UFR is applied by central banks as explained in Solvency II. We apply the Smith-Wilson smoothing technique using Thomas and Maré (2007) and implementation notes from Norway (2010)¹⁴ to the swap curve input data with a UFR of 4.2%, a last liquid point of 20 years and the aim of reaching the UFR in 60 years from now by approaching it by a deviation of at most 3 basis points.

To implement the Nelson-Siegel model we run a cross-sectional regression for each month to estimate the factors x_{it} in the model

$$y_t(\tau) = x_{1t} + x_{2t} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + x_{3t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_t(\tau) \quad (2.30)$$

The model is estimated using data with maturities $\tau = 1, \dots, 20$ years. We estimate the convergence parameter λ once as a constant for all months in the sample. The value $\lambda = 0.51$ minimizes the overall sum of squared errors over all maturities and months. Conditional on the estimated factors \hat{x}_{it} we construct the Nelson-

¹⁴Financial Supervisory Authority of Norway: www.finanstilsynet.no/Global/Forsikring%20og%20pensjon/Skadeforsikring/Tilsyn%20og%20overv%C3%A5king/Rapportering/A-Technical_Note_on_the_Smith-Wilson_Method.100701.pdf.

Siegel forward curve as

$$f_t(s) = \hat{x}_{1t} + \hat{x}_{2t}e^{-\lambda s} + \hat{x}_{3t}\lambda se^{-\lambda s} \quad (2.31)$$

for maturities $s > \tau^*$ and use these to extrapolate the yield curve using (2.12).

Figure 2.9 compares the different extrapolations for September 2013, the last month in our sample. Both the Smith-Wilson and Nelson-Siegel curve provide a very smooth extrapolation for maturities beyond 20 years. The Nelson-Siegel curve does not move far from the last observed yield, since the effect of the second and third factors are already small at these long maturities. The Smith-Wilson curve is slightly higher in order to converge to the level of 4.2%, which is above the estimated value of the Nelson-Siegel level factor for this month. The two Vasicek extrapolations are above the two alternatives. As discussed before, this is related to the convexity terms in the extrapolation. The very high level of the ML extrapolation is also due to the large, but inaccurate, estimate for the ultimate yield θ . The Vasicek extrapolation shows a kink at the last liquid point, because it does not take into account the local curvature at this point. It just calibrates the single factor to the level of the 20-year rate for continuity. The kink can be smoothed away by using the alternative calibration of the factor by the forward rate at τ^* instead of the yield y_t^* . We show this in the next section.

Figure 2.10 adds 95% HPD bounds to the graph, which puts the differences in perspective. Scaling on this graph is different, because the uncertainty around the extrapolation is large. Both two alternative extrapolations as well as the observed data are within the bounds.

Figure 2.9: Alternative extrapolations

The figure shows extrapolated yield curves for September 2013 using alternative extrapolation methods. The solid blue line shows the actual discount rates extracted from the swap curve. ML refers to extrapolation based on the maximum likelihood parameter estimated in the unrestricted model in Table 2.1.

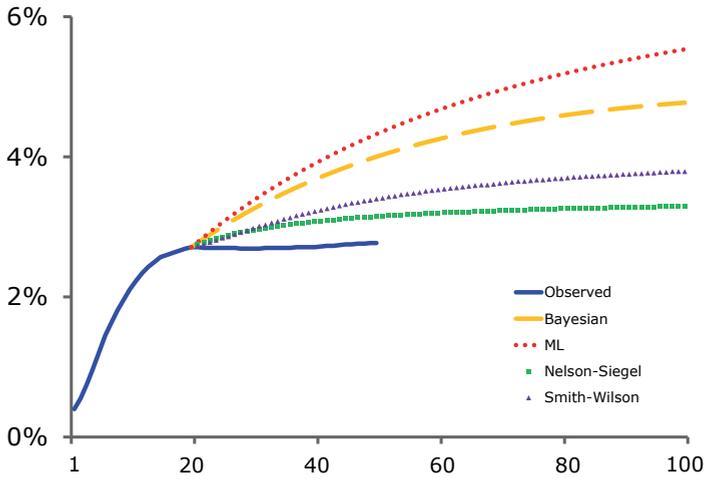
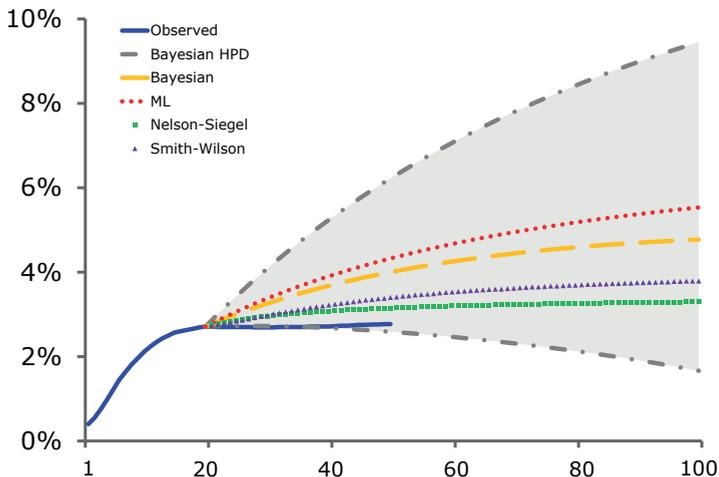


Figure 2.10: Extrapolation uncertainty

The dashed lines are the 95% HPD upper and lower bounds from the Bayesian extrapolation.



2.8 Robustness

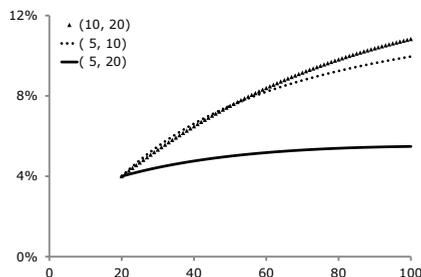
We performed several robustness checks with respect to the Vasicek extrapolations. One set of variations considers the maturities on which the model parameters are estimated. Secondly, it appears that the extrapolation produces a kink at the 20 years maturity. One way to get around the discontinuity in the slope is by using forward rates in the extrapolation. For a third robustness check we evaluate if the same features, a strong convexity and extrapolations above the observed yield curve, are present in US swap rate data.

2.8.1 Maturities

For the basic setting we used maturities of 5 and 20 years to estimate the parameters. The 5-year maturity could be too low and could be too much affected by other less persistent factors. For this reason we also consider the combination $(\tau_1, \tau_2) = (10, 20)$ years. On the other hand, since most term structure models are estimated on maturities up to 10 years we also estimate the model on the pair $(\tau_1, \tau_2) = (5, 10)$ years. In both cases the distance between the maturities is less than in the basic model. As a result it turns out that the mean-reversion parameter $\tilde{\kappa}$ is estimated with less precision and has much more probability mass around the unit root. This further emphasises the problem of estimating the ultimate forward rate θ . Figure 2.11 displays the resulting extrapolations. For both alternative pairs of (τ_1, τ_2) the convexity effect is more important resulting in extrapolations above that of the baseline model. The posterior moments of the parameters for several other maturities are attached in Appendix 2.F.

Figure 2.11: Extrapolation based on alternative maturities

The figure shows yield curve extrapolations from alternative parameter estimates. The parameters are estimated for alternative pairs of maturities (τ_1, τ_2) .



2.8.2 Smoothness

From the figures it appears that the extrapolation produces a kink at the 20 years maturity. The slope of the yield curve increases abruptly at the start of the extrapolation. The curve itself is continuous, but the slope is not. This is not an artefact of the yield curve of the particular date of September 2013, but this holds more generally. The observed yield curve is already almost flat at $\tau^* = 20$ years, but the extrapolation implies a steep upward slope.

One way to get around the discontinuity in the slope is by using forward rates in the extrapolation. Let f_t^* be the observed instantaneous forward rate at maturity τ^* . We can recover the unknown state variables x_t using the Vasicek relation for the forward rate

$$f_t(\tau) = \theta + e^{-\tilde{\kappa}\tau}(x_t - \theta) + e^{-\tilde{\kappa}\tau} \frac{1 - e^{-\tilde{\kappa}\tau}}{\tilde{\kappa}} \omega^2 \quad (2.32)$$

In Appendix 2.E we derive the yield extrapolation

$$y_t(s) = \theta + \frac{\tau^*}{s}(y_t^* - \theta) + \left(1 - \frac{\tau^*}{s}\right)b(s - \tau^*)(f_t^* - \theta) + c^*(s) \quad (2.33)$$

where

$$c^*(s) = \frac{1}{2} \left(\frac{e^{-\tilde{\kappa}s} - e^{-\tilde{\kappa}\tau^*}}{\tilde{\kappa}} \right)^2 \omega^2 \frac{1}{s} \quad (2.34)$$

As in the original yield extrapolation the convexity term $c^*(s)$ is positive for all $s > \tau^*$ and reaches zero at $s = \tau^*$ and as $s \rightarrow \infty$. The extrapolation will be much smoother than the original extrapolation, since it takes both the local level and slope to start the extrapolation. This will avoid the kink at τ^* . Apparently forward rates at τ^* are on average lower than they should be according to the Vasicek model. It is also evidence against a single factor model.

The instantaneous forward rate f_t^* is not directly observable from the yield curve data. Instead we can use the one-year forward

$$f_t^* \equiv f_t(\tau^* - 1, \tau^*) = \tau^* y_t(\tau^*) - (\tau^* - 1) y_t(\tau^* - 1) \quad (2.35)$$

as a finite difference approximation. Figure 2.12 shows the separate components of the smooth extrapolation with the highest posterior densities. Figure 2.13 shows the extrapolation calibrated with the forward rate and the yield.

Figure 2.12: Smooth components

The upper left panel shows the convexity $c^*(s)$, the middle left panel shows the theta term $\theta \left(1 - \frac{\tau^*}{s} - (1 - \frac{\tau^*}{s})b(s - \tau^*)\right)$, the lower left panel shows the forward term $(1 - \frac{\tau^*}{s})b(s - \tau^*)f_t^*$ and the right panels add the 95% highest posterior densities.

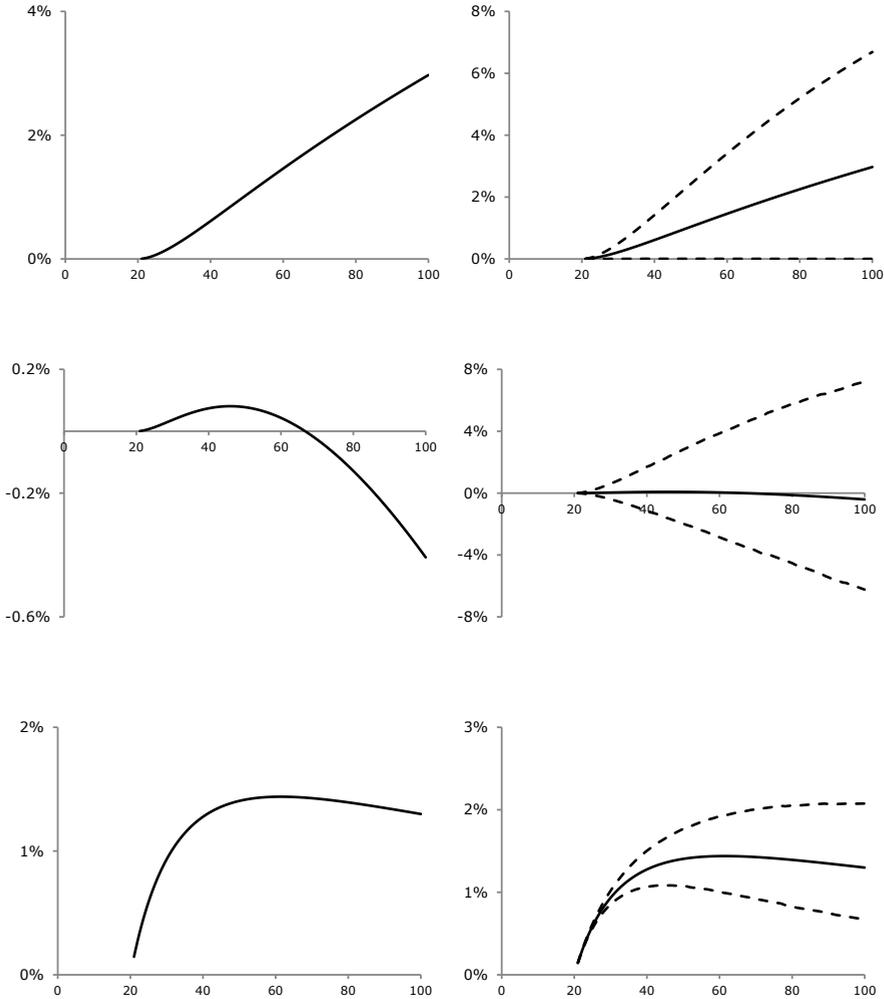
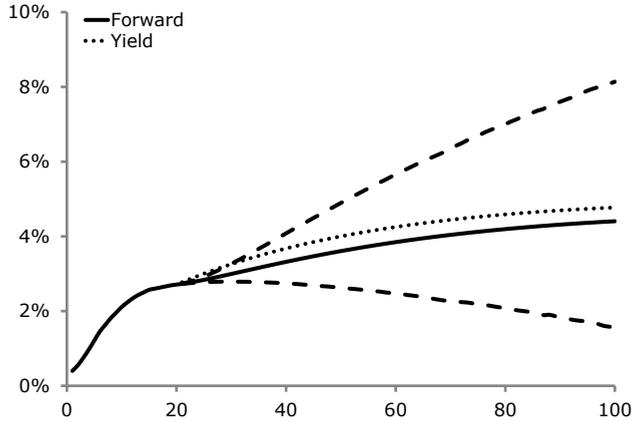


Figure 2.13: Smooth posterior extrapolation

The figure shows the extrapolated yield curve for September 2013 based on the calibration with the forward and the yield. “Yield” refers to extrapolation based on the observed 20-year discount yield; “Forward” uses the 20-year forward rate for the extrapolation.



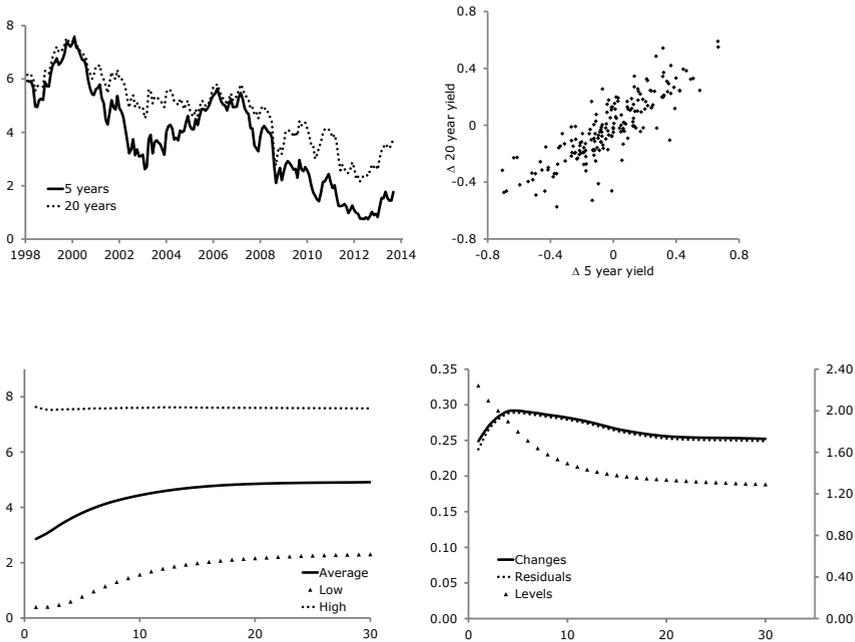
2.8.3 Analysis United States

In this subsection we apply the Bayesian model to US swap rates. So far we have shown all resulting components for the Euro swap rates. Our yield curve data consists of a monthly panel of discount rates from the website of the FED and LIBOR rates from Datastream. These yield curve data are constructed from US swap rates with maturities ranging from 1 to 30 years. The sample period is from April 1998 to May 2014 resulting in 193 data-points per maturity.

Figure 2.14 shows the overview of the US statistics. In general there are no remarkable differences compared with the EU analysis, though the scatterplot of the first differences shows a higher R^2 than for the European data.

Figure 2.14: Data analysis US

In the upper left panel the time series data. The upper right panel is a scatterplot of the first differences. In the lower left panel the solid line shows the average yield curve over the period from April 1998 to May 2014. The dotted line and the line with triangles show the yield curves with the minimum and maximum rate at the 20 years maturity. The lower right panel shows the volatility of yield level, yield changes and the one month prediction errors from an AR(1) model. The horizontal axis in both panels is the maturity in years. The vertical axis unit is percent per year. The right vertical axis is for levels; the left vertical axis is for changes and prediction errors.



Also here we estimate the parameters by conditional maximum likelihood first. Table 2.3 shows the unrestricted model in the first two columns and the results of re-estimating under the restriction that $\Lambda_1 = 0$. The mean-reversion under both measures is close to the point estimate from the European data, however note that the standard deviation under \mathbb{Q} is a factor 10 smaller. The parameters are to be estimated with more precision due to the fact that a one-factor model is a better model for the US market than for the EU market. The smaller relative size of the error term η^2 to the standard deviation σ^2 reveals this as well. As a consequence $\tilde{\kappa}$ can be better estimated and is therefore going to the unit root with a smaller probability, which explains θ to be much less extreme.

Table 2.3: ML parameter estimates US

The table reports conditional Maximum Likelihood estimates for the bivariate model using interest rates with maturities of 5 and 20 years. The asymptotic standard errors (“se”) are from the Hessian of the log-likelihood function. The columns on the right side of the table are estimated under the restriction $\tilde{\kappa} = \kappa$. All parameters are reported in their natural units with time measured in years. The last line is the value of the log-likelihood function.

	Estimate	se	Estimate	se
$\tilde{\kappa}$	0.0213	6.084×10^{-3}	0.0217	6.139×10^{-3}
κ	0.2745	0.1301	= $\tilde{\kappa}$	
$\tilde{\mu}$	1.182×10^{-3}	3.71×10^{-3}	1.761×10^{-3}	3.704×10^{-3}
μ	2.45×10^{-4}	1.12×10^{-4}	-8.61×10^{-4}	1.326×10^{-3}
θ	1.171×10^{-3}	3.68×10^{-4}	1.750×10^{-3}	3.703×10^{-3}
Λ_0	0.4170	0.5374	-0.56935	0.9315
Λ_1	-2514.61	1279.43	0	
σ^2	1.01×10^{-4}	1.3×10^{-5}	1.00×10^{-4}	1.3×10^{-5}
ω^2	2.387×10^{-3}	5.89×10^{-4}	2.303×10^{-3}	5.61×10^{-4}
η^2	1.2×10^{-5}	$1. \times 10^{-6}$	1.2×10^{-5}	$1. \times 10^{-6}$
$\ln L$	3978.84		3976.90	

The Bayesian analysis on the US data is reported in Table 2.4. Contrary to the European case where θ was not well-behaved, here it is. Although the 95% HPD range is still very wide, it is still economically sound. Moreover, we can see that the distribution has a fat right tail. Hence although the standard deviation of $\tilde{\kappa}$ reduced by a factor 2 compared to a factor 10 in the ML setting, the impact is still evident. The spread of $\tilde{\kappa}$ going to the unit root and therefore θ going to infinity is mitigated by the match of the one-factor model (Figure 2.15).

Figure 2.15: Conditional posterior draws of θ given $\tilde{\kappa}$ for US

The figure shows a scatter plot of the draws $\theta^{(i)}$ conditional on $\tilde{\kappa}^{(i)}$. The vertical axis is in natural units: -1 therefore means -100% .

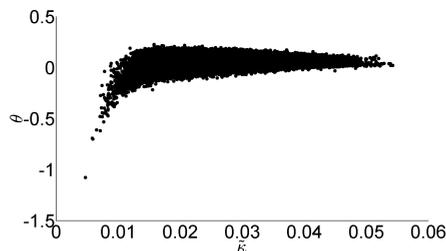


Table 2.4: Posterior moments of parameters US

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0278	0.0050	1.795×10^{-2}	0.0377
κ	0.1267	0.0758	6.012×10^{-6}	0.2656
$\tilde{\mu}$	0.1076	0.0399	4.611×10^{-2}	0.1873
μ	0.0148	0.0040	8.660×10^{-3}	0.0208
θ	4.052×10^{-2}	3.921×10^{-2}	-0.0246	0.1229
Λ_0	-0.1009	0.1537	-0.4317	0.1847
Λ_1	-10.2142	7.8614	-24.9190	2.9799
σ^2	9.490×10^{-5}	1.134×10^{-5}	7.374×10^{-5}	1.177×10^{-4}
ω^2	1.751×10^{-3}	3.1757×10^{-4}	1.1971×10^{-3}	2.3807×10^{-3}
η^2	6.760×10^{-6}	6.924×10^{-7}	5.440×10^{-6}	8.124×10^{-6}

The table shows the posterior means, standard deviations and 95% highest posterior density intervals for the parameters of the Vasicek term structure model estimated on interest rates with maturities $\tau_1 = 5$ and $\tau_2 = 20$. Results are based on one million draws from the Gibbs sampler for the parameters α , \mathbf{m} and Σ , which are solved for the structural parameters in the table.

The separate components of the extrapolation are displayed in Figure 2.16. Remarkable differences are the average estimate of the posterior mean that is positive and upward sloping whereas in the EU market this component was almost negligible in its size. This is due to the fact that this term goes less often to zero since also $\tilde{\kappa}$ does so less. On the other hand the convexity term is also smaller. Hence in the EU setting the larger negative posterior mean plus the larger positive convexity appear to result in the same size uncertainty in the US setting. Therefore the ultimate extrapolations appear to be robust.

The extrapolation is shown in Figure 2.17 and the 95% HPD is added to Figure 2.18. Beyond the last liquid point of 20 years, we only observe swap rates until 30 years in the American data set. Similar as in the Euro market are the extrapolations overestimating the data.

Figure 2.16: US components

In the upper left panel the solid line shows the extrapolation with a last liquid yield of 4% and the 95% HPD dashed. The upper right panel is the extrapolation with a last liquid yield of 2%, 4% and 6%. The middle left panel shows the convexity and the HPD and on the right the speed of convergence is shown $(b(s)/b^*)$. The lower left panel shows the posterior mean of $(1 - b(s)/b^*)\theta$ for $s > \tau^* = 20$, the lower right panel adds the 95% highest posterior density region.

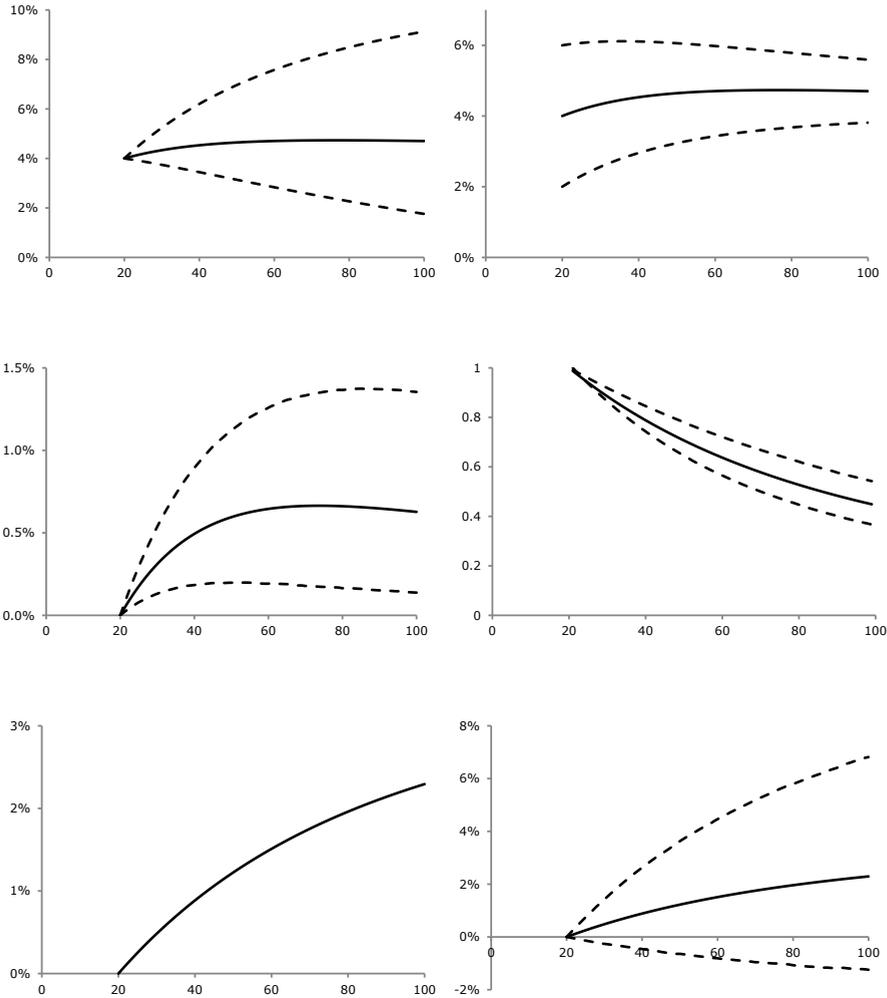


Figure 2.17: Alternative extrapolations US

Observed rates from May 2014, and the described alternative methods.

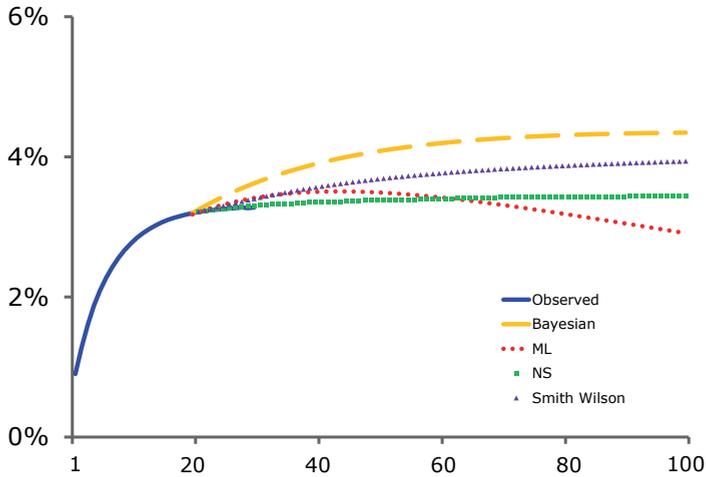
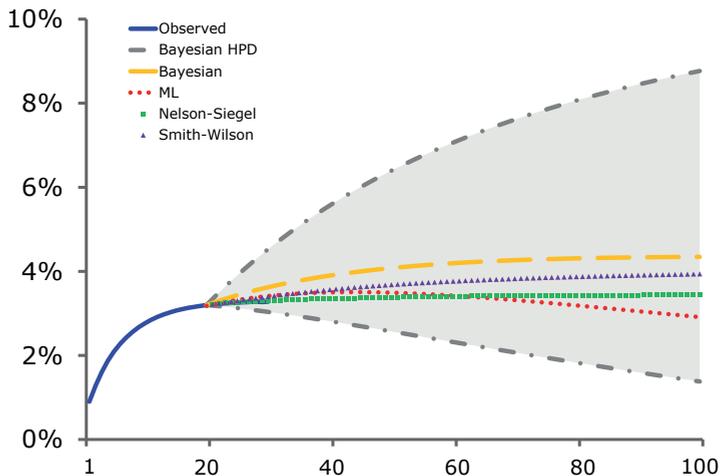


Figure 2.18: Extrapolation uncertainty US

The figure shows extrapolated yield curves for May 2014. The solid blue line shows the actual discount rates extracted from the swap curve. The dashed lines are the 95% HPD upper and lower bounds from the Bayesian extrapolation. The purple triangles is the Smith-Wilson method applied to the rates from May 2014, the green squares is the Nelson-Siegel method applied to the rates from May 2014 and the yellow and red lines are the Vasicek method based on the Bayesian method and maximum likelihood estimation respectively.



2.9 Conclusion

We extrapolated the yield curve using the Vasicek model. The Vasicek model produces extrapolated yield curves that are in most cases slightly upward sloping at very long maturities. The main difference with alternative extrapolation techniques is the convexity effect in very long-term yields. The convexity effect is an important element in no-arbitrage term structure models and can be a large component due to the slow mean-reversion of the dominant interest rate level factor under the risk-neutral measure.

The extrapolation based on a no-arbitrage term structure model is always above the observed yield curve. This both underscores the importance of using a model based yield curve for valuations of very long-dated maturities, but also raises the issue of why observations at the very long end are so low and why they do not fit a standard term structure model.

Moreover, by use of the Bayesian methodology we are able to construct the credible interval to quantify the uncertainty of the extrapolations. Note that the combination of an affine term structure model with the acknowledgement of uncertainty could be applied to different interest rate models as well, such as the Cox-Ingersoll-Ross model. The range of the 95% interval of the extrapolations includes several alternatives used in the industry. Hence, the appropriateness to rely on a single point estimate should be questioned based on this quantification of uncertainty. Therefore in the next chapter we investigate the optimal hedging strategy that minimises the mismatch between the assets and the liabilities.

2.A Derivation of Yields

The Vasicek process of the factor (short rate¹⁵) under \mathbb{P} is

$$dx_t = \kappa(\mu - x_t)dt + \sigma dW_t \quad (2.36)$$

By Itô's Lemma, the explicit expression under \mathbb{Q} is

$$x_{t+s} = x_t e^{-\tilde{\kappa}s} + \tilde{\mu}(1 - e^{-\tilde{\kappa}s}) + \sigma \int_t^{t+s} e^{-\tilde{\kappa}(t+s-u)} dW_u \quad (2.37)$$

The integral is

$$\begin{aligned} \int_0^\tau x_{t+s} ds &= (x_t - \tilde{\mu}) \frac{(1 - e^{-\tilde{\kappa}\tau})}{\tilde{\kappa}} + \tilde{\mu}\tau + \\ &\int_0^\tau \left(\int_t^{t+s} \sigma e^{-\tilde{\kappa}(t+s-u)} dW_u \right) ds \end{aligned} \quad (2.38)$$

Changing the order of integrals gives

$$\int_0^\tau \left(\int_t^{t+s} \sigma e^{-\tilde{\kappa}(t+s-u)} dW_u \right) ds = \int_t^{t+\tau} \left(-\frac{\sigma}{\tilde{\kappa}} (1 - e^{-\tilde{\kappa}(t+\tau-u)}) \right) dW_u$$

Let

$$\begin{aligned} M &= \mathbb{E} \left[\int_0^\tau x_{t+s} ds \middle| x_t \right] \\ &= (x_t - \tilde{\mu}) \frac{(1 - e^{-\tilde{\kappa}\tau})}{\tilde{\kappa}} + \tilde{\mu}\tau \end{aligned} \quad (2.39)$$

and let

$$\begin{aligned} V &= \text{var} \left[\int_0^\tau x_{t+s} ds \middle| x_t \right] \\ &= \frac{\sigma^2}{\tilde{\kappa}^2} \left(\tau - \frac{1 - e^{-\tilde{\kappa}\tau}}{\tilde{\kappa}} - \frac{(1 - e^{-\tilde{\kappa}\tau})^2}{2\tilde{\kappa}} \right) \end{aligned} \quad (2.40)$$

If $X \sim N(M, V)$, then $\mathbb{E}[e^X] = e^{M + \frac{1}{2}V}$. Therefore

$$\mathbb{E}[e^{-\int_0^\tau x_s ds}] = e^{-M + \frac{1}{2}V} = e^{-\tau y(\tau)} \quad (2.41)$$

¹⁵For the one-factor model, the short rate r_t coincides with the factor x_t .

and

$$y(\tau) = (x_t - \tilde{\mu}) \frac{(1 - e^{-\tilde{\kappa}\tau})}{\tilde{\kappa}\tau} + \tilde{\mu} - \frac{\sigma^2}{2\tilde{\kappa}^2} \left(1 - \frac{1 - e^{-\tilde{\kappa}\tau}}{\tilde{\kappa}\tau} - \frac{(1 - e^{-\tilde{\kappa}\tau})^2}{2\tilde{\kappa}\tau} \right)$$

Then it follows that

$$\begin{aligned} y(\tau) &= b(\tau) \left[x_t - \theta \right] + \theta + \frac{1}{2} \tau \omega^2 b(\tau)^2 \\ y(s) &= \tilde{b}(s) \left[x_t - \theta \right] + \theta + \frac{1}{2} s \omega^2 \tilde{b}(s)^2 \\ y(s) &= \frac{\tilde{b}(s)}{b(\tau)} \left[y(\tau) - \theta \right] + \theta + \frac{1}{2} \omega^2 \tilde{b}(s) \left(s \tilde{b}(s) - \tau b(\tau) \right) \end{aligned} \quad (2.42)$$

where

$$b(\tau) = \frac{1 - e^{-\tilde{\kappa}\tau}}{\tilde{\kappa}\tau} \quad (2.43)$$

$$\theta = \tilde{\mu} - \frac{\sigma^2}{2\tilde{\kappa}^2} \quad (2.44)$$

$$\omega^2 = \frac{\sigma^2}{2\tilde{\kappa}} \quad (2.45)$$

2.B Affine Relation Between \mathbb{P} and \mathbb{Q}

Under the assumption that two short rate models exist under two different probability measures, risk-neutral \mathbb{Q} and risky \mathbb{P} , the relation between the parameters of the different measures can be derived by use of a stochastic discount factor (SDF).

Let the SDF be

$$\frac{d\Lambda_t}{\Lambda_t} = -x_t dt - \lambda_t dW_t \quad (2.46)$$

where

$$\lambda_t = \Lambda_0 + \Lambda_1 x_t \quad (2.47)$$

Since under the risk-neutral measure $\lambda = 0$ we can derive the relation between the two probability measures. The log of the price is affine with respect to the short

rate x . The price at t (Cochrane (2001)) is

$$p(\tau, t, r) = -A(\tau) - B(\tau)x_t \quad (2.48)$$

where $T = t + \tau$ is the maturity date of the bond and τ is the remaining time to maturity. Thus the (antilog) of the price (Duffie and Kan (1996)) is

$$P(\tau, t, x_t) = \exp(-A(\tau) - B(\tau)x_t) \quad (2.49)$$

By Itô's Lemma

$$dP(\tau, t, x_t) = -B(\tau)P_t dx_t + \left(\frac{\partial A(\tau)}{\partial t} + \frac{\partial B(\tau)}{\partial t} x_t \right) P_t dt + \frac{1}{2} B^2(\tau) \sigma^2 P_t dt$$

The Fundamental Pricing Equation states

$$\mathbb{E}_t \left[\frac{dP_t}{P_t} \right] - x_t dt = -\mathbb{E} \left[\frac{dP_t}{P_t} \frac{d\Lambda_t}{\Lambda_t} \right] \quad (2.50)$$

Selecting only those term excluding x_t , leads to

$$\frac{\partial A(\tau)}{\partial t} = B(\tau) [\kappa\mu - \sigma\Lambda_0] - \frac{1}{2} B^2(\tau) \sigma^2 \quad (2.51)$$

And the terms including x_t

$$\frac{\partial B(\tau)}{\partial t} = 1 - B(\tau) [\sigma\Lambda_1 + \kappa] \quad (2.52)$$

For completeness, the above formulas are all in terms of probability measure \mathbb{P} . The derivatives of component A and B do not depend on the probability measure. Since under the risk-neutral measure $\lambda_t = 0$, we put the terms in brackets equal to $\tilde{\kappa}\tilde{\mu}$ and $\tilde{\kappa}$ respectively. Hence, we obtain the relation

$$\begin{aligned} \tilde{\kappa} &= \kappa + \sigma\Lambda_1 \\ \tilde{\mu}\tilde{\kappa} &= \mu\kappa - \sigma\Lambda_0 \end{aligned} \quad (2.53)$$

2.C Admissability Conditions for Σ

The bivariate restricted VAR has error covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = s_h^2 \begin{pmatrix} \sigma^2 b_1^2 + \eta^2 & \sigma^2 b_1 b_2 \\ \sigma^2 b_1 b_2 & \sigma^2 b_2^2 + \eta^2 \end{pmatrix} \quad (2.54)$$

Assuming $\tau_1 < \tau_2$ we impose $\sigma_{11} > \sigma_{22}$ to ensure $b_1 > b_2$. Similarly, since b_1 and b_2 are both positive, we also impose $\sigma_{21} > 0$. To solve for $\tilde{\kappa}$ we construct

$$S \equiv \frac{\sigma_{11} - \sigma_{22}}{\sigma_{21}} = \frac{b_1^2 - b_2^2}{b_1 b_2} \quad (2.55)$$

where the scalar S is a function of the elements of Σ and where the last expression only depends on $\tilde{\kappa}$. The conditions on Σ imply that $S > 0$. The condition can be rewritten as

$$\frac{b_2}{b_1} = \frac{1}{2} \left(\sqrt{S^2 + 4} - S \right) \quad (2.56)$$

Since in our application $\tau_2 = 4\tau_1$, the left-hand side can be rewritten as

$$\frac{b_2}{b_1} = \frac{1}{4} (1 + e^{-\tilde{\kappa}\tau_1}) (1 + e^{-2\tilde{\kappa}\tau_1}) \quad (2.57)$$

For positive $\tilde{\kappa}$ this is a monotone decreasing function in $\tilde{\kappa}$, and hence the equation has a unique solution for $\tilde{\kappa}$, if it exists. For $S > 0$ the right-hand side varies between 0 and 1, meaning that $b_2 < b_1$ as required by positive mean-reversion $\tilde{\kappa} > 0$. Since the left-hand side has a lower bound of $\frac{1}{4}$, we must also restrict the right-hand side to be above $\frac{1}{4}$, which implies the restriction $S < \frac{15}{4}$. This upper bound is a third admissibility condition on Σ .

2.D Covariance Decomposition

The relation between the draws of the MCMC sampler and the parameters of the extrapolation are displayed in Table 2.5.

Table 2.5: Relation parameters

	Relation
$\tilde{\kappa}$	$\frac{(1-e^{-\tilde{\kappa}\tau_1})\tau_2}{(1-e^{-\tilde{\kappa}\tau_2})\tau_1} - \frac{(1-e^{-\tilde{\kappa}\tau_2})\tau_1}{(1-e^{-\tilde{\kappa}\tau_1})\tau_2} = \frac{\sigma_{(11)} - \sigma_{(22)}}{\sigma_{(21)}}$
κ	$-\frac{\ln(1-\alpha)}{h}$
$\tilde{\mu}$	$\theta + \frac{\sigma^2}{2\tilde{\kappa}^2}$
μ	$\frac{(1-b(\tau_2))m(\tau_1) - (1-b(\tau_1))m(\tau_2)}{b(\tau_1) - b(\tau_2)} - \frac{1}{2}\omega^2 \frac{\tau_1 b(\tau_1)^2(1-b(\tau_2)) - \tau_2 b(\tau_2)^2(1-b(\tau_1))}{b(\tau_1) - b(\tau_2)}$
θ	$\frac{b(\tau_2)m(\tau_1) - b(\tau_1)m(\tau_2)}{b(\tau_2) - b(\tau_1)} - \frac{1}{2}\omega^2 b(\tau_1)b(\tau_2) \frac{\tau_1 b(\tau_1) - \tau_2 b(\tau_2)}{b(\tau_2) - b(\tau_1)}$
Λ_0	$\frac{\mu\kappa - \tilde{\mu}\tilde{\kappa}}{\sigma}$
Λ_1	$\frac{\tilde{\kappa} - \kappa}{\sigma_{(21)}}$
σ^2	$\frac{\sigma_{(21)}}{b(\tau_1)b(\tau_2)}$
η^2	$\sigma_{(11)} - \sigma^2 b(\tau_1)^2, \sigma_{(22)} - \sigma^2 b(\tau_2)^2$

2.E Forward Rate Extrapolation

Let f_t^* be the observed instantaneous forward rate at maturity τ^* . We can recover the unknown state variables x_t using the Vasicek relation for the forward rate

$$f_t(\tau) = \theta + e^{-\tilde{\kappa}\tau}(x_t - \theta) + e^{-\tilde{\kappa}\tau} \frac{1 - e^{-\tilde{\kappa}\tau}}{\tilde{\kappa}} \omega^2 \quad (2.58)$$

Solving for the state variable x_t at $\tau = \tau^*$

$$x_t - \theta = e^{\tilde{\kappa}\tau^*} (f_t^* - \theta) - \frac{1 - e^{-\tilde{\kappa}\tau^*}}{\tilde{\kappa}} \omega^2 \quad (2.59)$$

and the extrapolated forward rates

$$f_t(s) = \theta + e^{-\tilde{\kappa}(s-\tau^*)} (f_t^* - \theta) + \frac{e^{-\tilde{\kappa}(s+\tau^*)} - e^{-2\tilde{\kappa}s}}{\tilde{\kappa}} \omega^2 \quad (2.60)$$

For the extrapolation we use the relation

$$sy_t(s) = \tau^* y_t^* + \int_{\tau^*}^s f_t(u) du \quad (2.61)$$

Integrating the different terms in (2.60) gives

$$\int_{\tau^*}^s e^{-\tilde{\kappa}(u-\tau^*)} du = \frac{1 - e^{-\tilde{\kappa}(s-\tau^*)}}{\tilde{\kappa}} \quad (2.62)$$

$$\frac{1}{\tilde{\kappa}} \int_{\tau^*}^s e^{-\tilde{\kappa}(u+\tau^*)} du = \frac{e^{-2\tilde{\kappa}\tau^*} - e^{-\tilde{\kappa}(s+\tau^*)}}{\tilde{\kappa}^2} \quad (2.63)$$

$$\frac{1}{\tilde{\kappa}} \int_{\tau^*}^s e^{-2\tilde{\kappa}u} du = \frac{e^{-2\tilde{\kappa}\tau^*} - e^{-2\tilde{\kappa}s}}{2\tilde{\kappa}^2} \quad (2.64)$$

Subtracting (2.64) from (2.63) leads to the convexity term

$$c^*(s) = \frac{1}{2} \left(\frac{e^{-\tilde{\kappa}s} - e^{-\tilde{\kappa}\tau^*}}{\tilde{\kappa}} \right)^2 \omega^2 \frac{1}{s} \quad (2.65)$$

Substituting these results back in (2.60) provides the cumulative forward rate extrapolation

$$\int_{\tau^*}^s f_t(u) du = (s - \tau^*)\theta + \frac{1 - e^{-\tilde{\kappa}(s-\tau^*)}}{\tilde{\kappa}} (f_t^* - \theta) + sc^*(s) \quad (2.66)$$

and the yield extrapolation

$$y_t(s) = \theta + \frac{\tau^*}{s} (y_t^* - \theta) + \left(1 - \frac{\tau^*}{s}\right) b(s - \tau^*) (f_t^* - \theta) + c^*(s) \quad (2.67)$$

2.F Sensitivity

As a sensitivity analysis we use different maturities as input, in Table 2.6 for European data and in Table 2.7 for data from the United States. The first two subtables in Table 2.7 show the results for different priors on \mathbf{m} .

Table 2.6: EU sensitivity

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0060	0.0044	3.361×10^{-8}	0.0143
κ	0.0940	0.0710	2.994×10^{-8}	0.2304
$\tilde{\mu}$	52.0148	4.2063×10^4	-0.1615	7.1668
μ	0.0031	0.0090	-0.01308	0.0130
θ	-1.123×10^5	4.489×10^7	-75.1360	0.8137
Λ_0	-0.4715	0.3877	-1.0647	0.0343
Λ_1	-13.2024	10.6898	-34.1862	1.4707
σ^2	4.527×10^{-5}	6.373×10^{-6}	3.368×10^{-5}	5.812×10^{-5}
η^2	2.839×10^{-6}	3.413×10^{-7}	2.210×10^{-6}	3.528×10^{-6}

The table shows the posterior means, standard deviations and 95% highest posterior density intervals for the parameters of the Vasicek term structure model based on European swap rates with maturities $\tau_1 = 10$ and $\tau_2 = 20$.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0461	0.0168	0.01356	0.0792
κ	0.1417	0.0936	5.833×10^{-7}	0.3121
$\tilde{\mu}$	0.1734	0.7167	0.02985	0.3812
μ	0.0052	0.0068	-9.516×10^{-3}	0.0133
θ	-0.1265	33.21	0.0190	0.3382
Λ_0	-0.6815	0.4866	-1.6798	-0.0036
Λ_1	-12.2359	12.1459	-35.4197	7.1246
σ^2	6.346×10^{-5}	1.084×10^{-5}	4.366×10^{-5}	8.523×10^{-5}
η^2	4.151×10^{-6}	4.991×10^{-7}	3.225×10^{-6}	5.152×10^{-6}

With maturities $\tau_1 = 5$ and $\tau_2 = 10$.

2 IMPLICATION TSM ON LONG-TERM DISCOUNT RATES

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0109	6.10×10^{-3}	6.676×10^{-6}	0.0219
κ	0.1721	0.0912	3.715×10^{-5}	0.3343
$\tilde{\mu}$	0.3651	2.7169	0.03733	0.7732
μ	0.0155	4.80×10^{-3}	7.789×10^{-3}	0.0242
θ	-71.33	362.0	-4.6994	0.1472
Λ_0	0.1606	0.257	-0.2974	0.6953
Λ_1	-24.6762	14.1726	-50.8244	1.2185
σ^2	4.367×10^{-5}	7.469×10^{-6}	2.994×10^{-5}	5.862×10^{-5}
η^2	1.359×10^{-5}	1.625×10^{-6}	1.056×10^{-5}	1.683×10^{-5}

With maturities $\tau_1 = 5$ and $\tau_2 = 30$.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0033	2.70×10^{-3}	9.537×10^{-7}	0.0086
κ	0.1070	0.0712	1.027×10^{-7}	0.239
$\tilde{\mu}$	4.2149	49.6704	-0.07197	9.0853
μ	0.0104	5.60×10^{-3}	2.481×10^{-3}	0.0212
θ	-1.025×10^4	4.270×10^5	-351.4672	0.2753
Λ_0	-0.1587	0.2564	-0.5534	0.2115
Λ_1	-16.1018	11.1135	-37.0919	0.7036
σ^2	4.211×10^{-5}	5.877×10^{-6}	3.125×10^{-5}	5.388×10^{-5}
η^2	5.802×10^{-6}	6.972×10^{-7}	4.510×10^{-6}	7.204×10^{-6}

With maturities $\tau_1 = 10$ and $\tau_2 = 30$.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0023	2.00×10^{-3}	9.537×10^{-7}	0.0063
κ	0.0339	0.0462	6.757×10^{-8}	0.1184
$\tilde{\mu}$	14.6400	145.1763	-1.137	34.7334
μ	-0.0194	0.0210	-0.05025	0.0305
θ	-1.931×10^4	6.173×10^5	-1.0934×10^3	1.0269
Λ_0	-0.7075	0.4595	-1.3337	0.2315
Λ_1	-4.6955	6.8701	-17.4774	1.0469
σ^2	4.601×10^{-5}	6.102×10^{-6}	3.463×10^{-5}	5.809×10^{-5}
η^2	1.482×10^{-6}	1.787×10^{-7}	1.147×10^{-6}	1.838×10^{-6}

With maturities $\tau_1 = 20$ and $\tau_2 = 30$.

Table 2.7: US sensitivity

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0277	5.00×10^{-3}	0.01784	0.0376
κ	0.1100	0.0636	2.346×10^{-7}	0.2253
$\tilde{\mu}$	0.0989	0.0285	0.05025	0.1549
μ	0.0089	2.80×10^{-3}	4.286×10^{-3}	0.013
θ	0.03133	0.02775	-0.0181	0.0842
Λ_0	-0.1743	0.0861	-0.3500	-0.0158
Λ_1	-8.4931	6.6128	-20.7558	2.9676
σ^2	9.495×10^{-5}	1.136×10^{-5}	7.387×10^{-5}	1.18×10^{-4}
η^2	6.760×10^{-6}	6.923×10^{-7}	5.459×10^{-6}	8.14×10^{-6}

The table shows the posterior means, standard deviations and 95% highest posterior density intervals for the parameters of the Vasicek term structure model based on American swap rates with maturities $\tau_1 = 5$ and $\tau_2 = 20$ and the prior $\mu_m = [-1.5; -1.5]$ which implies that the prior conditional expectation equals 0.0222.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0278	5.00×10^{-3}	0.01807	0.0378
κ	0.1337	0.0907	3.114×10^{-7}	0.3085
$\tilde{\mu}$	0.1212	0.0576	0.04078	0.2377
μ	0.0240	8.20×10^{-3}	9.911×10^{-3}	0.0375
θ	0.05404	0.05655	-0.0320	0.1767
Λ_0	0.0025	0.2870	-0.6201	0.5782
Λ_1	-10.9381	9.4029	-29.3444	3.1095
σ^2	9.496×10^{-5}	1.135×10^{-5}	7.371×10^{-5}	1.178×10^{-4}
η^2	6.764×10^{-6}	6.930×10^{-7}	5.448×10^{-6}	8.137×10^{-6}

With maturities $\tau_1 = 5$ and $\tau_2 = 20$ and prior $\mu_m = [-0.5; -0.5]$, the conditional expectation equals 0.064549.

2 IMPLICATION TSM ON LONG-TERM DISCOUNT RATES

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0888	0.0201	0.05028	0.1289
κ	0.1371	0.0816	1.944×10^{-6}	0.2852
$\tilde{\mu}$	0.1032	0.0439	0.03880	0.1824
μ	0.0038	7.80×10^{-3}	-0.01262	0.0145
θ	0.09003	0.6391	0.0290	0.1664
Λ_0	-0.6355	0.2973	-1.2490	-0.183
Λ_1	-3.7753	6.4767	-16.1996	7.325
σ^2	1.751×10^{-4}	3.126×10^{-5}	1.1179×10^{-4}	2.378×10^{-4}
η^2	9.245×10^{-6}	1.112×10^{-6}	7.168×10^{-6}	1.147×10^{-5}

With the standard priors and maturities $\tau_1 = 5$ and $\tau_2 = 10$.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0165	3.50×10^{-3}	9.826×10^{-3}	0.0234
κ	0.1371	0.0816	1.388×10^{-7}	0.2857
$\tilde{\mu}$	0.1701	0.0503	0.08724	0.2673
μ	0.0163	4.30×10^{-3}	9.344×10^{-3}	0.0239
θ	-0.08742	0.1505	-0.2734	0.0585
Λ_0	-0.0372	0.1479	-0.2983	0.2481
Λ_1	-10.9329	7.4292	-24.6567	1.5356
σ^2	1.232×10^{-4}	1.671×10^{-5}	9.198×10^{-5}	1.564×10^{-4}
η^2	9.245×10^{-6}	1.109×10^{-6}	7.172×10^{-6}	1.146×10^{-5}

With maturities $\tau_1 = 5$ and $\tau_2 = 30$.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0444	0.0100	0.02509	0.0644
κ	0.1370	0.0816	3.893×10^{-7}	0.2854
$\tilde{\mu}$	0.1231	0.0427	0.05453	0.2063
μ	0.0012	7.80×10^{-3}	-0.01533	0.0119
θ	0.07324	0.03836	0.0116	0.1490
Λ_0	-0.3900	0.1844	-0.7784	-0.1170
Λ_1	-7.1169	6.3590	-19.0809	3.5780
σ^2	1.751×10^{-4}	3.129×10^{-5}	1.181×10^{-4}	2.381×10^{-4}
η^2	9.244×10^{-6}	1.108×10^{-6}	7.170×10^{-6}	1.145×10^{-5}

With maturities $\tau_1 = 10$ and $\tau_2 = 20$.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0212	4.60×10^{-3}	0.01230	0.0302
κ	0.1371	0.0817	4.322×10^{-7}	0.2856
$\tilde{\mu}$	0.1653	0.0492	0.08545	0.2612
μ	0.0089	4.90×10^{-3}	2.536×10^{-4}	0.0160
θ	-0.01182	0.0934	-0.1293	0.0953
Λ_0	-0.1800	0.1178	-0.4122	0.0226
Λ_1	-9.8792	7.0003	-22.8645	1.8594
σ^2	1.400×10^{-4}	2.065×10^{-5}	1.016×10^{-4}	1.812×10^{-4}
η^2	9.245×10^{-6}	1.111×10^{-6}	7.153×10^{-6}	1.144×10^{-5}

With maturities $\tau_1 = 10$ and $\tau_2 = 30$.

	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0501	0.0133	0.02517	0.0766
κ	0.1371	0.0817	4.247×10^{-6}	0.2857
$\tilde{\mu}$	0.1440	0.0449	0.07086	0.2312
μ	-0.0309	0.0193	-0.07200	-0.0027
θ	0.08026	0.03820	0.0197	0.1557
Λ_0	-0.6715	0.3039	-1.2987	-0.2369
Λ_1	-5.3420	5.1560	-15.0461	3.1827
σ^2	2.918×10^{-4}	8.669×10^{-5}	1.483×10^{-4}	4.626×10^{-4}
η^2	9.244×10^{-6}	1.110×10^{-6}	7.181×10^{-6}	1.147×10^{-5}

With maturities $\tau_1 = 20$ and $\tau_2 = 30$.

Chapter 3

Pricing and Hedging in Incomplete Markets with Model Ambiguity

We search for pricing methods of assets in incomplete markets. Our set-up is that we postulate an agent who wants to maximise the expected surplus by choosing an optimal hedging strategy. Furthermore, we assume that the agent is concerned about model misspecification. This robust optimal control problem based on the assumption of model ambiguity leads to: risk-neutral pricing for the traded risky assets and adjusting the drift from the non-traded risk drivers in the conservative direction, known as “actuarial” or “prudential” pricing. In a multivariate incomplete market our method yields the existence and uniqueness of the robust price.¹

¹This chapter is based on the paper Balter and Pelsser (2015a).

3.1 Introduction

If markets are complete then there exists a unique pricing measure such that all contracts have a unique price. Due to illiquidity of financial instruments multiple equivalent martingale measures exist in an incomplete market. Pension funds, insurance companies and many other parties need to price liabilities that are not or only partially hedgeable. The existing literature gives many pricing and hedging methods in incomplete markets. However, additional to pricing in an incomplete market, there is the limitation that the underlying model can be wrong. In financial and economic decision making allowing for model uncertainty has an impact on the optimal strategies. Our goal is to find a pricing method for assets in incomplete markets under the acknowledgement of model uncertainty.

Pricing in an incomplete market setting often boils down to stipulating a selection rule to pick the equivalent martingale measure that has certain characteristics or implications. In Chapter 1, the literature on different pricing methods in incomplete markets is discussed.

We build on the objective to minimise the discrepancy in incomplete markets by maximising the difference between the hedge position and the liabilities. Maximising the profit is a frequently observed objective of doing business. This coincides with a linear expectation or equivalently a risk-neutral agent. Profit maximisation in an incomplete market leads to a constrained optimisation problem. Since insurance companies and pension funds want to offset the risk of their liabilities that are unhedgeable, hedging constraints alter the standard delta-hedge position. Examples of pricing in incomplete markets are; extremely long-dated obligations such as pensions that have to be paid on a horizon up to a century (since life expectancy increases the horizon grows even further) or insurance contracts linked to instruments that are not or barely traded, such as the number of survivors (this example is shown in Section 3.3.4), temperature or natural disasters. Moreover, by considering a risk-neutral agent who is ambiguous about the model we focus on the latter uncertainty factor.

Thus, we postulate an agent who wants to find the optimal hedge position, or equivalently the optimal price, for the (un)hedgeable item. This is solved by maximising the surplus between the hedge position and the item to be priced. As he distrusts the baseline model, he intends to search for a robust decision. The robust price is deduced from the robust optimal hedging strategy that is least sensitive to perturbations of the underlying model. Therefore the optimal

control problem that follows is formulated as a max-min optimisation problem. The objective of the agent is to maximise the surplus for the worst-case model that might happen. The worst-case is selected by a so called mother nature, who minimises the surplus subject to the specification of the set of plausible alternative models. Since it is difficult to estimate the return with much precision contrary to the accuracy of the volatility estimation we still take a wide class of models into account as alternative deviations on the baseline model. Wide in the sense that the derivations in this chapter hold for both deterministic and stochastic deviations, implying a change in the distribution or volatility as well. In specific, we consider ellipsoid uncertainty. The quantification of the size of the set is extensively investigated in the next chapter. Moreover, the uncertainty set is time-consistent since the constant bound on the set corresponds with a Lipschitz driver. This is also proven by the connection with the backward stochastic differential equation (BSDE) representation.

Our contribution is that we price a contract that depends on both hedgeable and unhedgeable risk, on which we add ambiguity of the drifts to both types of risk. The objective is to maximise the surplus given a set of alternative models. In terms of the overview of pricing methods discussed in Chapter 1, our approach of “pricing under model ambiguity” belongs to the class that considers a “subset”. In particular, we use a set of coherent risk measures. The dynamic hedging strategy follows from the robust optimisation. Moreover, we use the concept of indifference pricing to obtain the optimal “value” of the liability. To be exact, the utility function we use is the linear expectation of the surplus. We assume that, although the agent is uncertain about the true model, he can make an indication of possible models surrounding his estimation. A risk-neutral and ambiguity averse agent who is maximising his surplus leads to a unique time-consistent optimal price and hedging strategy.

We illustrate our approach in the context of pensions and insurances, though it can be applied more generally. Pricing and hedging in incomplete markets is of great interest in financial and economic modelling. In the literature the focus lies on pricing complex contracts and when allowed for risk aversion, model ambiguity is often swept under the carpet.

Additional to literature review given in Chapter 1, we start with a short recap of the associated literature. In Section 2 of Chapter 3, we deepen into the intuition behind our approach and the derivations towards Theorem 3.1 and Lemma 3.1. Section 3.3 describes some examples. The first two examples show the extreme

settings of a pure complete and a pure incomplete market. The economic interpretations of the optimal solutions are explained in these two subsections as well. The third example presents two correlated factors and is rather general to show the widespread applicability of the methodology. The last example given is related to insurance and actuarial problems in asset-liability management.

3.1.1 Literature Review

The acknowledgement that investors are concerned and do not rely on a single model specification, but instead act as if it is merely an approximation gains attention from both academia and the applied world. Robustness is investigated by Hansen and Sargent by a series of papers and their book Hansen and Sargent (2008). Boyle et al. (2008) search for the robust stochastic discount factor in an incomplete market. Maenhout (2004) extends the consumption and portfolio problem of Merton (1969) by including uncertainty on the return process. The decision rule he uses to evaluate the different models is the one that works reasonably well under misspecification. The misspecification is measured by a (non-symmetric) distance measure between the baseline model and the alternative. This concept is known as ϕ -divergences where Maenhout (2004) uses the Kullback-Leibler divergence. Robustness is implemented by the addition of a penalty term to the utility function. We add an uncertainty constraint.

Cao et al. (2005), Garlappi et al. (2007) and Wang (2005) approach model uncertainty from a Bayesian perspective, which are extensions of the multiple prior model from Gilboa and Schmeidler (1989). In the multiple prior approach of Wang (2005) the investor is uncertain about the prior beliefs in asset-pricing models. The tangency portfolios appear no longer to be optimal in a CAPM setting when the investor is uncertainty averse. The uncertainty in Garlappi et al. (2007) is characterised by confidence intervals around the estimated expected returns. Also Cao et al. (2005) maximise the utility for the worst-case prior, though they concentrate on the impact of market participation on the equity premium. A combination of the classical and Bayesian robustness is what we exploit, to be exempted from the prior specification and including the worst-case perturbation that mother nature can select.

Basak and Chabakauri (2012) minimise the hedging preference or hedging quality in an incomplete market. They define the objective by various criteria in terms of the means and variances of the hedging error, which is the deviation of the hedge from the target value. The intuition coincides with our objective where the

agent maximises the difference between the hedge position and the liability. Moreover, their explanation of incompleteness is represented from a similar viewpoint as in this chapter. The difference between these papers and this chapter is that we focus on the impact of model uncertainty whereas they do not include misspecification. Also Chacko and Viceira (2005) work in an incomplete market setting, where they examine consumption and portfolio-choice problems with a focus on stochastic volatility. Cochrane and Saa-Requejo (2000) allow trades in only those assets that have plausible mean-variances. The Good-Deal-Bounds method states that the existence of assets with extreme market prices of risk are unlikely to be true (“too good to be true”).

The concept of utility indifference pricing we employ is based on the price of a claim such that an agent who is maximising his utility is indifferent between having the claim or not. Musiela and Zariphopoulou (2004) show some recent work on utility indifference pricing, for exponential utilities in specific.

3.2 Optimal Pricing Result

3.2.1 The Agent’s Problem

Before the derivation is obtained in the next subsection, we first introduce the general route we will follow. The agent aims to hedge his liabilities by finding the optimal hedging portfolio. Therefore the liability $L(\cdot)$ is the item to be priced, that might depend on both hedgeable and unhedgeable risk. The hedging portfolio is represented by the positions in the traded assets $A(\cdot)$. This set-up is similar as in an asset-liability management (ALM) problem.

The liabilities may depend on both hedgeable and unhedgeable risk. Hedgeable risk represents the risk that underlies the liquid and traded assets, i.e. the agent can go short or long positions in these assets. That is why this market is called the complete market. Whereas in an incomplete market, the risk factor is not traded and therefore unhedgeable. Let there be n tradeable assets (\mathbf{x}), l untradeable risk factors (\mathbf{y}) and let there be a bank account on which one can go short or long for the risk-free rate. The goal of maximising the profit, payoff or surplus has a widespread applicability and intuition described in the introduction. The agent’s choice variable at every point in time is the division over the different financial assets for the next time step. The function $A(t, \bar{\mathbf{x}}_t)$ is the total value of the hedge position. The agent chooses the hedge position $\bar{\boldsymbol{\theta}}(t)$ in the financial assets that

maximises his expected surplus at terminal time T . The liabilities may depend on both types of risk factors. This is given by

$$\max_{\bar{\theta}(t)} \mathbb{E}[A(T, \bar{\mathbf{x}}_T, \bar{\boldsymbol{\theta}}_T) - L(T, \bar{\mathbf{x}}_T, \mathbf{y}_T)] \quad (3.1)$$

The surplus at terminal time T is a known function. Profit maximisation without trading constraints leads in most cases to an unbounded problem. By specifying a utility function this problem can be overcome. However, this implies that the risk-neutral objective of profit maximisation has been lost and the utility function has to be specified. In the utility literature risk-aversion is heavily investigated, though we consider model rather than risk ambiguity. The addition of ambiguity aversion bounds the problem of the risk-neutral agent.

The robust equivalent of objective (3.1) is

$$\max_{\bar{\theta}(t)} \min_{\mathbb{Q} \in \mathcal{Q}^{k^2}} \mathbb{E}^{\mathbb{Q}}[A(T, \bar{\mathbf{x}}_T, \bar{\boldsymbol{\theta}}_T) - L(T, \bar{\mathbf{x}}_T, \mathbf{y}_T) | \mathcal{F}_t] \quad (3.2)$$

In the robust model the inner part of the optimisation is played by a so called “mother nature” who acts as a malevolent factor and minimises the surplus by choosing the worst-case scenario, whereas the agent searches for the best hedge strategy that is least affected by mother nature’s choice.

We specify model uncertainty by alternative models as a set of different probability measures. By Girsanov Theorem there is a one-to-one correspondence between probability measures and drifts. Therefore the set \mathcal{Q} can be identified as an ellipsoid around the baseline model. The width of this set of alternatives indicates the amount of ambiguity and is represented by the scalar k .

$$\mathcal{K} = \{ \boldsymbol{\mu}(t, \mathbf{z}) + \boldsymbol{\epsilon}(t) | \boldsymbol{\epsilon}(t)' \boldsymbol{\Sigma}^{-1}(t) \boldsymbol{\epsilon}(t) \leq k^2 \} \quad (3.3)$$

“Mother nature” chooses the worst-case model from this set in order to generate the robust solution. The choice of k corresponds to the decision about the size of the set of alternatives. We propose to link k to those models that are indistinguishable from each other. By imposing both a probability on the Type I and Type II error, the bound on the ellipsoid can be made. We focus on the quantification of the set of indistinguishable models in Chapter 4.

In Section 3.2.2, we explain the complete procedure and derivation of our method. The multivariate solution is characterised by the partial differential equation (PDE) given. The solution of this can either be obtained analytically

or numerically. At least the existence of the unique solution is proven by the backward stochastic differential equation link. An agent whose objective is to maximise profit, but who is uncertain about the process, prices by the risk-neutral measure if he is in a complete market setting. If no financial instruments are at his disposal we show that this coincides with the idea of prudent pricing. The last two examples are both a combination of the two types of risk. The analytical solution exists for the described correlated risk factors and for the unit-linked contract.

3.2.2 Financial Market with Model Ambiguity

The formal specification of the model starts by some notational assumptions. Suppose that there are n tradeable assets with prices $\mathbf{x}(t) = \{x_i(t) | i = \{1, \dots, n\}\}$ plus the bank account x_0 and suppose there are l different unhedgeable risk factors which we represent by the observed values $\mathbf{y}(t) = \{y_i(t) | i = \{1, \dots, l\}\}$. The vector \mathbf{x} represent the hedgeable (financial) risk and the vector \mathbf{y} represents the unhedgeable (insurance) risk. The column vector of the two is denoted by $\mathbf{z} = (\mathbf{x}, \mathbf{y})'$. When we add the bank account to the set of hedgeable assets the specific vector has length $n + 1$ and is identified by a bar. Both processes are characterised by the multivariate stochastic differential equation

$$d \begin{pmatrix} \mathbf{x}(t) \\ \mathbf{y}(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_x(t, \mathbf{x}) \\ \boldsymbol{\mu}_y(t, \mathbf{y}) \end{pmatrix} dt + \boldsymbol{\Sigma}^{1/2}(t, \mathbf{x}, \mathbf{y}) d \begin{pmatrix} \mathbf{W}_x(t) \\ \mathbf{W}_y(t) \end{pmatrix} \quad (3.4)$$

If we vectorise this notation and add the bank account process we get

$$d \begin{pmatrix} x_0(t) \\ \mathbf{z}(t) \end{pmatrix} = \begin{pmatrix} r(t, \mathbf{x}(t))x_0(t) \\ \boldsymbol{\mu}(t, \mathbf{z}) \end{pmatrix} dt + \begin{pmatrix} \mathbf{0}' \\ \boldsymbol{\Sigma}^{1/2}(t, \mathbf{z}) \end{pmatrix} \cdot d\mathbf{W}_z(t) \quad (3.5)$$

where $\boldsymbol{\mu}(t, \mathbf{z}) = (\boldsymbol{\mu}_x(t, \mathbf{x}), \boldsymbol{\mu}_y(t, \mathbf{y}))'$, $\mathbf{W}_z(t) = (\mathbf{W}_x(t), \mathbf{W}_y(t))'$ and $\boldsymbol{\Sigma}(t, \mathbf{z}) = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}$ and we assume that $\boldsymbol{\Sigma}$ is invertible. The risk-free rate of the bank account can either be constant or follow a stochastic process. The accumulation over time represents the rate on the bank account.

The agent can trade in the hedgeable risk, but not in the unhedgeable risk. Therefore the amount he invests in the first n assets can be chosen such that he maximises the surplus, while the other l hedge positions with respect to the unhedgeable risk are restricted to zero. Thus we get $\theta_0, \boldsymbol{\theta}_x(t) = \overbrace{[\theta_1(t), \dots, \theta_n(t)]}^n$ and $\boldsymbol{\theta}_y(t) = \overbrace{[0, \dots, 0]}^l$,

where $\bar{\boldsymbol{\theta}}(t) = [\overbrace{\theta_0(t)}^1, \overbrace{\theta_1(t), \dots, \theta_n(t)}^n]$ and $\boldsymbol{\theta}(t) = [\overbrace{\theta_1(t), \dots, \theta_n(t)}^n, \overbrace{0, \dots, 0}^1]$. This can be imposed by the constraint $\mathbf{R}\boldsymbol{\theta} = \mathbf{0}$, where \mathbf{R} is an $[l \times (n + l)]$ matrix equal to $\begin{bmatrix} \mathbf{0}_{[l \times n]} & \mathbf{I}_{[l \times l]} \end{bmatrix}$, on the left a $[l \times n]$ zero-matrix and next to it a $[l \times l]$ identity matrix. How much the agent goes long or short on the bank account is captured by the position $\theta_0(t)$. We do not assume any short selling constraints. Each $\theta_i(t)$ with $i > 0$ represents the number of assets the agents buys of asset $x_i(t)$ (the amount invested is $\theta_i(t) \cdot x_i(t)$ at time t). The total value of the assets $A(t)$ includes the bank account. At every point in time a change in the value of the assets can only occur due to a change in the underlying values, no gain or loss is attainable by a re-allocation over the different assets. We incorporate this in the model by the assumption of the self-financing condition

$$dA(t) = \bar{\boldsymbol{\theta}}(t)' d\bar{\mathbf{x}}(t) \quad (3.6)$$

where the first term of $\bar{\mathbf{x}}_t$ is represented by the bank account $dx_0 = r(t, \mathbf{x}(t))x_0 dt$, accordingly $\bar{\mathbf{x}}$ is a vector of size $[(1 + n) \times 1]$. By Itô's Lemma one can see that a change in asset value may not be caused by a re-allocation of the value over the available assets indirectly follows from the second term of $dA(t) = \bar{\boldsymbol{\theta}}(t)' d\bar{\mathbf{x}}(t) + \bar{\mathbf{x}}(t)' d\bar{\boldsymbol{\theta}}(t)$ being to be zero. The definition of A is

$$A(t) = \bar{\boldsymbol{\theta}}(t)' \bar{\mathbf{x}}(t) \quad (3.7)$$

Rewriting the amount invested in the bank account in terms of the self-financing condition yields

$$A(t) = \theta_0(t)x_0(t) + \boldsymbol{\theta}'_{\mathbf{x}}(t)\mathbf{x}(t) \quad (3.8)$$

$$dA(t) = (\theta_0(t)r(t, \mathbf{x}(t))x_0(t) + \boldsymbol{\theta}'_{\mathbf{x}}(t)\boldsymbol{\mu}_{\mathbf{x}}(t, \mathbf{x}))dt + \boldsymbol{\theta}'_{\mathbf{x}}(t)\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{1/2}(t, \mathbf{x})d\mathbf{W}_{\mathbf{x}}(t) \quad (3.9)$$

Let the bank account $x_0(t)$ be the numéraire, then by Itô's Lemma the following holds

$$d\tilde{A}(t) = d\left(\frac{A}{x_0}\right)(t) \quad (3.10)$$

$$= \frac{1}{x_0(t)} \left\{ (\theta_0(t)r(t, \mathbf{x})x_0 + \boldsymbol{\theta}'_{\mathbf{x}}(t)\boldsymbol{\mu}_{\mathbf{x}}(t, \mathbf{x}) - A(t)r(t, \mathbf{x})) dt + \boldsymbol{\theta}'_{\mathbf{x}}(t)\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{1/2}(t, \mathbf{x})d\mathbf{W}_{\mathbf{x}}(t) \right\} \quad (3.11)$$

The definition (3.8) as function for the amount invested in the bank account ($\theta_0(t) = \frac{1}{x_0(t)}(A(t) - \boldsymbol{\theta}'(t)\mathbf{x}(t))$) can be substituted in (3.11)

$$d\tilde{A}(t) = \tilde{\boldsymbol{\theta}}'(t)d\mathbf{x}(t) \quad (3.12)$$

where

$$\tilde{\boldsymbol{\theta}}(t) = \frac{\boldsymbol{\theta}_x(t)}{x_0(t)} \quad (3.13)$$

Ambiguity in the mean is added by a combined set for the means of the vector process \mathbf{z} ,

$$\mathcal{K} = \{\boldsymbol{\mu}(t, \mathbf{z}) + \boldsymbol{\epsilon}(t) | \boldsymbol{\epsilon}(t)' \boldsymbol{\Sigma}^{-1}(t) \boldsymbol{\epsilon}(t) \leq k^2\} \quad (3.14)$$

where $\boldsymbol{\epsilon}(t) = (\boldsymbol{\epsilon}_x(t, \mathbf{x}), \boldsymbol{\epsilon}_y(t, \mathbf{y}))'$. The motivation behind the ellipsoid specification is that the financial or economic agent indicates his ambiguity of the point-estimation by a region around it. The drift adjustments are proportional to the covariance matrix since by Girsanov Theorem the change of measure results in a change in the drift multiplied by the covariance matrix of the baseline model. The agent believes that the true value of the drift parameters lies in the set \mathcal{K} . Moreover, we assume that k^2 is larger than the Sharpe ratio that is obtained from the baseline model. The theorem derived only holds while

$$k^2 - \mathbf{q}'_x \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{q}_x \geq 0 \quad (3.15)$$

is satisfied. The assumption is equivalent to stating that the risk-free rate should be included among the set of plausible models. This implies that model uncertainty cannot vanish. The geometric interpretation in a two dimensional case is an ellipse, in a three dimensional case an ellipsoid, and in general it has the form that the dot-product of a vector is bounded.²

3.2.3 Optimisation

As an intermediate step we consider the Feynman-Kač equation (FK) whereafter we proceed with the optimisation of it. First we divide the liability by the numéraire as well $\tilde{L}(t, \mathbf{z}_t) = L(t, \mathbf{z}_t)/x_0(t)$. The Feynman-Kač Theorem states

²Note that throughout this thesis we use the term ellipsoid as generalisation for any dimension.

that the conditional expectation is

$$\mathbb{E}^*[\tilde{A}(T) - \tilde{L}(T, \mathbf{z}_T) | \mathcal{F}_t] = \bar{v}(t, \tilde{A}, \mathbf{x}, \mathbf{y}) \quad (3.16)$$

where the star (*) represents the measure

$$\begin{cases} d\mathbf{W}_x(t) + \boldsymbol{\epsilon}_x(t)dt \\ d\mathbf{W}_y(t) + \boldsymbol{\epsilon}_y(t)dt \end{cases} \quad (3.17)$$

and $\bar{v}(t, \tilde{A}, \mathbf{x}, \mathbf{y})$ satisfies the PDE

$$\begin{aligned} & \bar{v}_t + \tilde{\boldsymbol{\theta}}'(t)(\mathbf{q}_x(t, \mathbf{x}) + \boldsymbol{\epsilon}_x(t))\bar{v}_{\tilde{A}} + (\boldsymbol{\mu}_x(t, \mathbf{x}) + \boldsymbol{\epsilon}_x(t))'\bar{\mathbf{v}}_x + \\ & (\boldsymbol{\mu}_y(t, \mathbf{y}) + \boldsymbol{\epsilon}_y(t))'\bar{\mathbf{v}}_y + \frac{1}{2}\text{tr}(\mathbf{H}_{\bar{v}(z)}\boldsymbol{\Sigma}_{\bar{v}}) = 0 \end{aligned} \quad (3.18)$$

where $\mathbf{H}_{\bar{v}} = \begin{bmatrix} \bar{v}_{\tilde{A}\tilde{A}} & \bar{v}_{\tilde{A}x} & \bar{v}_{\tilde{A}y} \\ \bar{v}_{x\tilde{A}} & \bar{v}_{xx} & \bar{v}_{xy} \\ \bar{v}_{y\tilde{A}} & \bar{v}_{yx} & \bar{v}_{yy} \end{bmatrix}$, $\boldsymbol{\Sigma}_{\bar{v}} = \begin{bmatrix} \boldsymbol{\theta}'\boldsymbol{\Sigma}_{xx}\tilde{\boldsymbol{\theta}} & \tilde{\boldsymbol{\theta}}'\boldsymbol{\Sigma}_{xx} & \tilde{\boldsymbol{\theta}}'\boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xx}\tilde{\boldsymbol{\theta}} & \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx}\tilde{\boldsymbol{\theta}} & \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}$ and where the subscripts on $\boldsymbol{\Sigma}$ represent covariance matrices consisting of the block covariances related to \mathbf{x} and \mathbf{y} , and $\mathbf{q} = (\mathbf{q}_x, \mathbf{q}_y) = \boldsymbol{\mu}(t, \mathbf{z}) - r(t, \mathbf{x}(t))\mathbf{z}(t)$.

For every state of the world ω , the PDE remains linear in \tilde{A} . Because for every $\boldsymbol{\theta}(t, \omega)$ and every $\boldsymbol{\epsilon}_x(t, \omega)$ as fixed stochastics the PDE is linear in all directions of $\bar{v}(\cdot)$ and its derivatives.

By adding ambiguity to the PDE, the optimised function is obtained by $\max_{\boldsymbol{\theta}(t)} \min_{\boldsymbol{\epsilon}(t)} \bar{v}(t, \tilde{A}, \mathbf{x}, \mathbf{y}) = v(t, \tilde{A}, \mathbf{x}, \mathbf{y})$. The robust optimisation problem can be interpreted as a two player game where the agent wants to maximise the surplus whereas the robustness role is played by the counter player “mother nature” who minimises the surplus. The robustly optimised value $v(t, \tilde{A}, \mathbf{x}, \mathbf{y})$ is given by the Hamilton-Jacobi-Bellman (HJB) equation,

$$\begin{aligned} & v_t + \max_{\boldsymbol{\theta}(t)} \min_{\boldsymbol{\epsilon}(t)} \left\{ \tilde{\boldsymbol{\theta}}'(t)(\mathbf{q}_x(t, \mathbf{x}) + \boldsymbol{\epsilon}_x(t))v_{\tilde{A}} + \mathbf{v}'_z(\boldsymbol{\mu}(t, \mathbf{z}) + \boldsymbol{\epsilon}(t)) + \right. \\ & \left. \frac{1}{2}v_{\tilde{A}\tilde{A}}\tilde{\boldsymbol{\theta}}(t)'\boldsymbol{\Sigma}_{xx}\tilde{\boldsymbol{\theta}} + \mathbf{v}'_{Ax}\boldsymbol{\Sigma}_{xx}\tilde{\boldsymbol{\theta}}(t) + \mathbf{v}'_{Ay}\boldsymbol{\Sigma}_{yx}\tilde{\boldsymbol{\theta}}(t) \right\} + \\ & \frac{1}{2}\text{tr}(\mathbf{H}_{v(z)}(t)\boldsymbol{\Sigma}) = 0 \end{aligned} \quad (3.19)$$

$$\text{s.t.} \quad \boldsymbol{\epsilon}(t)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}(t) \leq k^2 \quad (3.20)$$

$$\mathbf{R}\boldsymbol{\theta} = \mathbf{0}$$

$$\text{with} \quad v(T, \tilde{A}, \mathbf{x}, \mathbf{y}) = \tilde{A}(T, \mathbf{x}) - \tilde{L}(T, \mathbf{z}) \quad (3.21)$$

where (3.21) is the boundary condition at time T and $\mathbf{H}_{v(z)}(t)$ is the lower right

second order derivative matrix with respect to the variables \mathbf{z} only (no \tilde{A} terms). The HJB is intuitively equal to “the Feynman-Kač with optimal $\tilde{\boldsymbol{\theta}}(t)$ and $\boldsymbol{\epsilon}(t)$ per Δt ”. The difference between FK and HJB is that now $\tilde{\boldsymbol{\theta}}(t)$ and $\boldsymbol{\epsilon}(t)$ may depend on the value function and its derivatives which makes the PDE non-linear. Therefore the value function $v(t, \tilde{A}, \mathbf{x}, \mathbf{y})$ can no longer be represented by a linear expectation. We will prove in Theorem 3.1 that the value for $v(t, \tilde{A}, \mathbf{x}, \mathbf{y})$ can be represented by a non-linear expectation $\mathcal{E}[\cdot]$. Since the terminal condition is linear in \tilde{A} , we propose the Ansatz that $v(\cdot)$ is linear in \tilde{A}

$$v(t, \tilde{A}, \mathbf{x}, \mathbf{y}) = \tilde{A}(t) - \tilde{w}(t, \mathbf{x}, \mathbf{y}) \quad (3.22)$$

Despite the non-linearity caused by the max-min the linearity in \tilde{A} remains due to the specific boundary condition of the surplus. The function $\tilde{w}(\cdot)$ satisfies

$$\begin{aligned} -\tilde{w}_t + \max_{\boldsymbol{\theta}(t)} \min_{\boldsymbol{\epsilon}(t)} \left\{ \tilde{\boldsymbol{\theta}}(t)'(\mathbf{q}_x(t, \mathbf{x}) + \boldsymbol{\epsilon}_x(t)) \right. \\ \left. - \tilde{\mathbf{w}}'_z(\boldsymbol{\mu}(t, \mathbf{z}) + \boldsymbol{\epsilon}(t)) \right\} - \frac{1}{2} \text{tr}(\mathbf{H}_{\tilde{w}(\mathbf{z})}(t)\boldsymbol{\Sigma}) = 0 \end{aligned} \quad (3.23)$$

$$\text{s.t.} \quad \boldsymbol{\epsilon}(t)' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}(t) \leq k^2$$

$$\mathbf{R}\tilde{\boldsymbol{\theta}} = \mathbf{0}$$

$$\text{with} \quad \tilde{w}(T, \mathbf{x}, \mathbf{y}) = \tilde{L}(T, \mathbf{z}) \quad (3.24)$$

The agent maximises for $-\tilde{w}(t, \mathbf{x}, \mathbf{y})$ which is the difference between the non-linear expectation of the surplus and the value of the assets. At time T , $\tilde{w}(T, \mathbf{x}, \mathbf{y})$ is exactly the liability value divided by the numéraire. Note that \tilde{w} depends also on $\boldsymbol{\theta}(t)$ and $\boldsymbol{\epsilon}(t)$. Supporting the mindset of the agent he maximises minus \tilde{w} which can be interpreted as an equivalent for the discounted value of the liabilities. Since the only variables that depend on the maximin problem are $\boldsymbol{\theta}(t)$ and $\boldsymbol{\mu}(t, \mathbf{z})$ where $\boldsymbol{\mu}(t, \mathbf{z})$ can be rewritten in $\boldsymbol{\mu}(t, \mathbf{z}) + \boldsymbol{\epsilon}(t)$, and $\boldsymbol{\epsilon}$ is the dependent variable, the above objective function can be simplified by eliminating the terms that are independent of the decision variables. We concentrate first on the max-min, which we define by

$$m(t, \mathbf{x}, \mathbf{y}) = \max_{\boldsymbol{\theta}(t)} \min_{\boldsymbol{\epsilon}(t)} \left\{ \tilde{\boldsymbol{\theta}}(t)'(\mathbf{q}_x(t, \mathbf{x}) + \boldsymbol{\epsilon}_x(t)) - \tilde{\mathbf{w}}'_z(\boldsymbol{\mu}(t, \mathbf{z}) + \boldsymbol{\epsilon}(t)) \right\} \quad (3.25)$$

We optimise $m(t, \mathbf{x}, \mathbf{y})$ for every time step Δt and plug the optimal $m^*(t, \mathbf{x}, \mathbf{y})$ in

PDE (3.23) to obtain a new non-linear PDE

$$-\tilde{w}_t + m^*(t, \mathbf{x}, \mathbf{y}) - \frac{1}{2} \text{tr} \left(\mathbf{H}_{\tilde{w}(\mathbf{z})}(t) \boldsymbol{\Sigma} \right) = 0 \quad (3.26)$$

Maximising $v(\cdot)$ corresponds with minimising $\tilde{w}(\cdot)$, which is equivalent with maximising $m(\cdot)$. We concentrate on the max-min part $m(\cdot)$.

$$\begin{aligned} \max_{\boldsymbol{\theta}(t)} \min_{\boldsymbol{\epsilon}(t)} \quad & \tilde{\boldsymbol{\theta}}(t)' \cdot (\mathbf{q}_x(t, \mathbf{x}) + \boldsymbol{\epsilon}_x(t)) - \tilde{\mathbf{w}}_z' \boldsymbol{\epsilon}(t) \\ \text{s.t.} \quad & \boldsymbol{\epsilon}(t)' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}(t) \leq k^2 \end{aligned} \quad (3.27)$$

For ease of exposition we suppress the dependence from $\boldsymbol{\theta}$ and $\boldsymbol{\epsilon}$ of t in the derivation. Moreover, maximising over $\boldsymbol{\theta}$ is equivalent with maximising over $\tilde{\boldsymbol{\theta}}$. Thus, the vectorised objective reads as

$$\begin{aligned} \max_{\tilde{\boldsymbol{\theta}}} \min_{\boldsymbol{\epsilon}} \quad & \tilde{\boldsymbol{\theta}}' \cdot \mathbf{q} + \boldsymbol{\epsilon}' (\tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z) \\ \text{s.t.} \quad & \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon} \leq k^2 \\ & \mathbf{R} \tilde{\boldsymbol{\theta}} = \mathbf{0} \end{aligned} \quad (3.28)$$

By the constraint that the last l θ_i s are zero, also the last l components of \mathbf{q} drop out (\mathbf{q}_y). Where $\boldsymbol{\theta}, \mathbf{q}, \boldsymbol{\epsilon}$ are vectors of dimension $[(n+l) \times 1]$.

We solve this optimisation in two steps. First consider the inner minimisation

$$\begin{aligned} \min_{\boldsymbol{\epsilon}} \quad & \tilde{\boldsymbol{\theta}}' \cdot \mathbf{q} + \boldsymbol{\epsilon}' (\tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z) \\ \text{s.t.} \quad & \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon} \leq k^2 \end{aligned} \quad (3.29)$$

This is a linear objective with a quadratic constraint, from which we know that it has a unique minimum. The optimal strategy of mother nature is

$\boldsymbol{\epsilon}^* = -k \frac{\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z)}{\sqrt{(\tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z)' \boldsymbol{\Sigma} (\tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z)}}$. The objective for the agent is

$$\begin{aligned} \max_{\tilde{\boldsymbol{\theta}}} \quad & \tilde{\boldsymbol{\theta}}' \mathbf{q} - k \sqrt{(\tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z)' \boldsymbol{\Sigma} (\tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z)} \\ \text{s.t.} \quad & \mathbf{R} \tilde{\boldsymbol{\theta}} = \mathbf{0} \end{aligned} \quad (3.30)$$

This is a quadric objective with linear constraints. The covariance matrix $\boldsymbol{\Sigma}$ is per definition semi positive-definite. Moreover, since we assume the matrix to be invertible the matrix is positive definite and thus the square root is a convex function. Since the optimal $\boldsymbol{\epsilon}$ is the negative root, the agent maximises a concave function which results in a unique maximum (take $\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} - \tilde{\mathbf{w}}_z$, then $\max_{\hat{\boldsymbol{\theta}}} (\hat{\boldsymbol{\theta}} +$

$\tilde{w}_z)' \mathbf{q} - k \sqrt{\hat{\theta}' \Sigma \hat{\theta}}$ s.t. $\mathbf{R} \hat{\theta} = \mathbf{R} \tilde{w}_z$). See Appendix 3.A for the full derivation. For both control variables we obtain a candidate solution that satisfies the first order conditions (FOC). Since we have shown that the minimisation and maximisation both yield a unique solution the optimal strategies are

$$\begin{aligned} \tilde{\theta}^* &= \begin{bmatrix} \tilde{w}_x + \Sigma_{xx}^{-1} \Sigma_{xy} \tilde{w}_y + \tilde{h} \Sigma_{xx}^{-1} \mathbf{q}_x \\ \mathbf{0} \end{bmatrix} \\ \epsilon^* &= \begin{bmatrix} -\mathbf{q}_x \\ (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \tilde{h}^{-1} \tilde{w}_y - \Sigma'_{xy} \Sigma_{xx}^{-1} \mathbf{q}_x \end{bmatrix} \end{aligned} \quad (3.31)$$

where $\tilde{h} = \sqrt{\frac{\tilde{w}'_y (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \tilde{w}_y}{k^2 - \mathbf{q}'_x \Sigma_{xx}^{-1} \mathbf{q}_x}}$. Hence, without dividing by the numéraire, the optimal hedging strategy and drift distortions are

$$\begin{aligned} \theta^* &= \begin{bmatrix} w_x + \Sigma_{xx}^{-1} \Sigma_{xy} w_y + h \Sigma_{xx}^{-1} \mathbf{q}_x \\ \mathbf{0} \end{bmatrix} \\ \epsilon^* &= \begin{bmatrix} -\mathbf{q}_x \\ (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) h^{-1} w_y - \Sigma'_{xy} \Sigma_{xx}^{-1} \mathbf{q}_x \end{bmatrix} \end{aligned} \quad (3.32)$$

where $h = \sqrt{\frac{w'_y (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) w_y}{k^2 - \mathbf{q}'_x \Sigma_{xx}^{-1} \mathbf{q}_x}}$. This proves Lemma 3.1. Since the optimisation problem is expressed dependent on the numéraire we continue the derivation of the indifference price including the numéraire. The optimal $m^*(t, \mathbf{x}, \mathbf{y})$ is

$$\begin{aligned} m^*(t, \mathbf{x}, \mathbf{y}) &= -\tilde{w}'_x r(t, \mathbf{x}) \mathbf{x} - \tilde{w}'_y (\boldsymbol{\mu}_y - \Sigma'_{xy} \Sigma_{xx}^{-1} \mathbf{q}_x) \\ &\quad - c \sqrt{\tilde{w}'_y (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \tilde{w}_y} \end{aligned} \quad (3.33)$$

Plugging this into PDE (3.26) results in the non-linear PDE

$$\begin{aligned} &-\tilde{w}_t - \tilde{w}'_x r(t, \mathbf{x}(t)) \mathbf{x} - \tilde{w}'_y (\boldsymbol{\mu}_y - \Sigma'_{xy} \Sigma_{xx}^{-1} \mathbf{q}_x) - \frac{1}{2} \text{tr} (\mathbf{H}_{\tilde{w}(z)} \Sigma) \\ &- c \sqrt{\tilde{w}'_y (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \tilde{w}_y} = 0 \end{aligned} \quad (3.34)$$

where $c = \sqrt{k^2 - \mathbf{q}'_x \Sigma_{xx}^{-1} \mathbf{q}_x}$ and $\mathbf{H}_{\tilde{w}(z)} = \begin{bmatrix} \tilde{w}_{xx} & \tilde{w}_{xy} \\ \tilde{w}_{yx} & \tilde{w}_{yy} \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_{yy} \end{bmatrix}$.

The fact that the square root has to be positive leads to the assumption that the bank account is prevented to be strictly better or worse than the risky asset.

$$k^2 - \mathbf{q}'_x \Sigma_{xx}^{-1} \mathbf{q}_x \geq 0 \quad (3.35)$$

If this is violated, then in the agent's worst-case model the excess return is still positive therefore he shall always invest as much as possible in the financial world, because in his belief he faces no model uncertainty and shall never use the bank account. Note that the risk-neutral agent is indifferent with respect to risk since he is maximising the surplus. We concentrate on model misspecification purely. If there is only one hedgeable risk factor in the economy the condition states that the Sharpe ratio should be within the agent's set of alternatives. Then, this corresponds with the idea of the Good-Deal-Bounds methodology to exclude the deals that are overperforming, but here also those that underperform.

Because the financial components \mathbf{x} of the risk vector are perfectly replicated by the delta-hedge, the ambiguity regarding the mean of the hedgeable risk is eliminated and replaced by the risk-free return $r(t, \mathbf{x})$. The ambiguity for the mean of the unhedgeable processes \mathbf{y} are decomposed in the part consisting of the variance of each component of \mathbf{y} plus terms consisting of the covariances with all risk factors. This is known as the Schur complement ($\Sigma_{\mathbf{y}\mathbf{y}} - \Sigma'_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{x}\mathbf{x}}^{-1} \Sigma_{\mathbf{x}\mathbf{y}}$) which is the conditional variance of the unhedgeable risk given the hedgeable risk.

Remember the Ansatz of (3.19) to be

$$v(t, \tilde{A}, \mathbf{x}, \mathbf{y}) = \tilde{A}(t) - \tilde{w}(t, \mathbf{x}, \mathbf{y}) \quad (3.36)$$

The indifference price of the agent between investing the initial value of assets on the bank account without liability contract and the initial value plus the equivalent that is needed to be indifferent to writing the contract is $\pi(t, \mathbf{x}, \mathbf{y})$. Since from (3.36) the growth of the bank account is known, without contract we get an expected surplus of $v(t, \tilde{A}, \mathbf{x}, \mathbf{y}) = \tilde{A}(t)$ at time t (since without contract $\tilde{L}(T, \mathbf{x}, \mathbf{y}) = 0$ the non-linear expectation becomes linear). While the indifference price, the extra cash needed to make the agent indifferent between the liability contract or not, is $\tilde{\pi}(t, \mathbf{x}, \mathbf{y})$. The expected surplus is $v(t, \tilde{A}, \mathbf{x}, \mathbf{y}) = \tilde{A}(t) + \tilde{\pi}(t, \mathbf{x}, \mathbf{y}) - \tilde{w}(t, \mathbf{x}, \mathbf{y})$. Hence by the equality condition the indifference price is determined.

$$\begin{aligned} \tilde{A}(t) &= \tilde{A}(t) + \tilde{\pi}(t, \mathbf{x}, \mathbf{y}) - \tilde{w}(t, \mathbf{x}, \mathbf{y}) \\ \tilde{\pi}(t, \mathbf{x}, \mathbf{y}) &= \tilde{w}(t, \mathbf{x}, \mathbf{y}) \end{aligned} \quad (3.37)$$

Where $\tilde{w}(t, \mathbf{x}, \mathbf{y}) = w(t, \mathbf{x}, \mathbf{y})/x_0(t) = e^{-\int_0^t r(s, \mathbf{x}(s)) ds} w(t, \mathbf{x}, \mathbf{y})$ and $\pi(t, \mathbf{x}, \mathbf{y}) = w(t, \mathbf{x}, \mathbf{y})$.

Rephrasing the PDE in terms of $\pi(t, \mathbf{x}, \mathbf{y})$ gives

$$\begin{aligned} & \pi_t + \boldsymbol{\pi}'_{\mathbf{x}} r(t, \mathbf{x}(t)) \mathbf{x} + \boldsymbol{\pi}'_{\mathbf{y}} (\boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{q}_{\mathbf{x}}) \\ & + \frac{1}{2} \text{tr} (\boldsymbol{\pi}_{\mathbf{x}\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}} + 2\boldsymbol{\pi}_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}} + \boldsymbol{\pi}_{\mathbf{y}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}) \\ & - r(t, \mathbf{x}(t)) \pi + c \sqrt{\boldsymbol{\pi}'_{\mathbf{y}} (\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}}) \boldsymbol{\pi}_{\mathbf{y}}} = 0 \end{aligned} \quad (3.38)$$

Note, that the equation is a semi-linear partial differential equation that describes the behaviour of the liability $\pi(t, \mathbf{z})$ as a function of t and \mathbf{z} . The PDE is subject to the boundary condition $\pi(T, \mathbf{z}) = L(T, \mathbf{x}, \mathbf{y})$ which is the value of the insurance contract at time T . Thus we derived the optimal PDE for which we can prove existence and uniqueness of the solution. The following theorem is proven based on the auxiliary representation of the semi-linear PDE as a BSDE problem.

Theorem 3.1 (Existence and uniqueness)

Let $\pi(t, \mathbf{x}, \mathbf{y})$ be the indifference price of an agent that maximises the expected surplus $\mathbb{E}[A(T, \mathbf{x}) - L(T, \mathbf{x}, \mathbf{y}) | \mathcal{F}_t]$ under model ambiguity in an incomplete market, then the indifference price π is given by the PDE

$$\begin{aligned} & \pi_t + \boldsymbol{\pi}'_{\mathbf{x}} r(t, \mathbf{x}(t)) \mathbf{x} + \boldsymbol{\pi}'_{\mathbf{y}} (\boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{q}_{\mathbf{x}}) + \\ & c \sqrt{\boldsymbol{\pi}'_{\mathbf{y}} (\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}}) \boldsymbol{\pi}_{\mathbf{y}}} + \\ & \frac{1}{2} \text{tr} (\boldsymbol{\pi}_{\mathbf{x}\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}} + 2\boldsymbol{\pi}_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}} + \boldsymbol{\pi}_{\mathbf{y}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}) - r(t, \mathbf{x}(t)) \pi = 0 \end{aligned} \quad (3.39)$$

with terminal value $\pi(T, \mathbf{x}, \mathbf{y}) = L(T, \mathbf{x}, \mathbf{y})$, where $c = \sqrt{k^2 - \mathbf{q}'_{\mathbf{x}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{q}_{\mathbf{x}}}$. The solution π of the semi-linear PDE exists and is unique.

Proof of Theorem 3.1 In case of a semi-linear multivariate PDE we can interpret it as a backward stochastic differential equation (BSDE) from which we know the existence and uniqueness of the solution by Pardoux and Peng (1990). We can express the PDE by

$$\begin{aligned} & \pi_t + \begin{bmatrix} \boldsymbol{\pi}_{\mathbf{x}} \\ \boldsymbol{\pi}_{\mathbf{y}} \end{bmatrix}' \begin{bmatrix} r(t, \mathbf{x}(t)) \mathbf{x} \\ (\boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{q}_{\mathbf{x}}) \end{bmatrix} + \frac{1}{2} \text{tr} \left(\begin{bmatrix} \boldsymbol{\pi}_{\mathbf{x}\mathbf{x}} & \boldsymbol{\pi}_{\mathbf{x}\mathbf{y}} \\ \boldsymbol{\pi}'_{\mathbf{x}\mathbf{y}} & \boldsymbol{\pi}_{\mathbf{y}\mathbf{y}} \end{bmatrix} \boldsymbol{\Sigma} \right) - \\ & r(t, \mathbf{x}(t)) \pi + \\ & c \sqrt{\begin{bmatrix} \boldsymbol{\pi}_{\mathbf{x}} \\ \boldsymbol{\pi}_{\mathbf{y}} \end{bmatrix}' \begin{bmatrix} 0 & 0 \\ 0 & (\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\pi}_{\mathbf{x}} \\ \boldsymbol{\pi}_{\mathbf{y}} \end{bmatrix}} = 0 \end{aligned} \quad (3.40)$$

where the linear, the Feynman-Kač part can be denoted by the operator

$$\begin{aligned} \mathcal{L}_{\mu,\sigma}\pi &= \left(\frac{\partial}{\partial t} + r(t, \mathbf{x}(t))\mathbf{x}' \frac{\partial}{\partial \mathbf{x}} + (\boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{q}_{\mathbf{x}})' \frac{\partial}{\partial \mathbf{y}} + \right. \\ &\quad \left. \frac{1}{2}\text{tr} \left(\frac{\partial^2}{\partial \mathbf{x}\mathbf{x}'} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}} + 2\frac{\partial^2}{\partial \mathbf{x}\mathbf{y}'} \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}} + \frac{\partial^2}{\partial \mathbf{y}\mathbf{y}'} \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}} \right) \right) \pi \end{aligned} \quad (3.41)$$

and the non-linear part by $g(\cdot)$. Moreover, both representations are linked by

$$\mathbf{X} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \text{ and } Y = \pi. \text{ Hence,}$$

$$\left. \begin{array}{l} \text{PDE} \\ \mathcal{L}_{\mu,\sigma}\pi + \\ g(t, \mathbf{X}_t, Y_t, \mathbf{Z}_t) = 0 \\ \pi(T, X) = L(T, X) \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \text{BSDE} \\ d\mathbf{X}_t = \boldsymbol{\mu} dt + \boldsymbol{\Sigma}^{1/2} d\mathbf{W}_t \\ dY_t = -g(t, \mathbf{X}_t, Y_t, \mathbf{Z}_t) + \mathbf{Z}_t d\mathbf{W}_t \\ g(t, \mathbf{X}_t, Y_t, \mathbf{Z}_t) = c\sqrt{\boldsymbol{\pi}'_{\mathbf{y}} \mathbf{S} \boldsymbol{\pi}_{\mathbf{y}}} - r(t, \mathbf{x}(t))Y_t \\ Y_T = L(T, X) \end{array} \right.$$

The first equation is the forward process where $\boldsymbol{\mu}$ are the drifts of \mathbf{x} and \mathbf{y} under the optimal measure, $\boldsymbol{\Sigma}$ is the covariance matrix that belongs to the trace terms of the PDE. The second equation is the backward process with the driver $g(\cdot)$, c is a scalar, the square root of restriction (3.35), \mathbf{S} is the Schur complement and $\mathbf{Z} = \nabla_{\mathbf{X}} Y_t \cdot \boldsymbol{\Sigma}(t, \mathbf{X}_t)$. Since $g(\cdot)$ is Lipschitz continuous as $g(t, \mathbf{Z}) \leq C|\mathbf{Z}|$, $g(\cdot, \mathbf{y}, Y, \mathbf{Z})$ is an \mathcal{F}_t -adapted process and $\int_0^T |g(t, \mathbf{y}, 0, \mathbf{0})| dt \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ the L2 space, the BSDE has a solution which is unique. By applying Itô's Lemma to π under the new measure we get

$$\begin{aligned} d\pi &= \pi_t dt + \boldsymbol{\pi}'_{\mathbf{x}}(r\mathbf{x} dt + \boldsymbol{\Sigma}_{\mathbf{x}} d\mathbf{W}_{\mathbf{x}}) + \boldsymbol{\pi}'_{\mathbf{y}}((\boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{q}_{\mathbf{x}}) dt + \\ &\quad \boldsymbol{\Sigma}_{\mathbf{y}} d\mathbf{W}_{\mathbf{y}}) + \frac{1}{2}\text{tr}(\boldsymbol{\pi}_{\mathbf{x}\mathbf{x}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}} + 2\boldsymbol{\pi}_{\mathbf{x}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}} + \boldsymbol{\pi}_{\mathbf{y}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}) \end{aligned} \quad (3.42)$$

Furthermore there is the relation given

$$dY = -g(t, \mathbf{X}_t, Y_t, \mathbf{Z}_t) dt + \mathbf{Z}_t d\mathbf{W} \quad (3.43)$$

The terms related to the Brownian motions are combined into the \mathbf{Z}_t term. Thus $\mathbf{Z}_t = (\boldsymbol{\pi}'_{\mathbf{x}}\boldsymbol{\Sigma}_{\mathbf{x}}, \boldsymbol{\pi}'_{\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}})'$. Since we have set $\pi = Y$, the definition of \mathbf{Z} as the derivatives of Y is fulfilled. The g -function equals

$$\begin{aligned} -g(t, \mathbf{X}_t, Y_t, \mathbf{Z}_t) &= \pi_t + \boldsymbol{\pi}'_{\mathbf{x}} r\mathbf{x} + \boldsymbol{\pi}'_{\mathbf{y}}((\boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{q}_{\mathbf{x}}) + \\ &\quad \frac{1}{2}\text{tr}(\boldsymbol{\pi}_{\mathbf{x}\mathbf{x}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}} + 2\boldsymbol{\pi}_{\mathbf{x}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}} + \boldsymbol{\pi}_{\mathbf{y}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}) \end{aligned} \quad (3.44)$$

By (3.38) the driver equals

$$g(t, \mathbf{X}_t, Y_t, \mathbf{Z}_t) = r(t, \mathbf{x}(t))\pi - c\sqrt{\pi'_y (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \pi_y} \quad (3.45)$$

■

A shorthand notation for the BSDE solution and the indifference price is

$$\pi(t, \mathbf{x}, \mathbf{y}) = e^{-\int_t^T r(s, \mathbf{x}(s)) ds} \mathcal{E}^g [L(T, \mathbf{x}, \mathbf{y}) | \mathcal{F}_t] \quad (3.46)$$

Moreover, we also know the optimal hedging portfolio θ^* and the robust deviations ϵ^* that belongs to optimisation problem of Theorem 3.1.

Lemma 3.1 (Optimal hedging portfolio and robustness adjustments)

Let $\pi(t, \mathbf{x}, \mathbf{y})$ be the indifference price of an agent that maximises the expected surplus $\mathbb{E}[A(T, \mathbf{x}) - L(T, \mathbf{x}, \mathbf{y}) | \mathcal{F}_t]$ under model ambiguity in an incomplete market, then the agent's robust optimal dynamic hedging portfolio is

$$\theta^* = \begin{bmatrix} \pi_x + \Sigma_{xx}^{-1} \Sigma_{xy} \pi_y + h \Sigma_{xx}^{-1} q_x \\ \mathbf{0} \end{bmatrix} \quad (3.47)$$

and the ambiguity ensures the optimal solution to be robust for drift adjustments

$$\epsilon^* = \begin{bmatrix} -q_x \\ (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) h^{-1} \pi_y - \Sigma'_{xy} \Sigma_{xx}^{-1} q_x \end{bmatrix} \quad (3.48)$$

where $h = \sqrt{\frac{\pi'_y (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \pi_y}{k^2 - q'_x \Sigma_{xx}^{-1} q_x}}$.

Proof of Lemma 3.1 Expression (3.31) is obtained by solving the first order conditions implied by the Lagrangian of the optimisation problem. Subsequently, w and its derivatives are expressed in term of π by relation (3.38). ■

The connection between PDE's and BSDE's proves the uniqueness and solvability of the theorem. Algorithms to find numerical solutions of BSDE's are well-known (Gobet et al., 2005; Gobet and Labart, 2007, 2010), see for applications of BSDE's El Karoui et al. (2009). Time-consistent ambiguity averse preferences including jump processes are priced by Laeven and Stadje (2014). Pelsser and Stadje (2014) arrive at the same class of pricing operators by imposing time- and market-consistency. The time-consistency is in our case directly fulfilled by the ini-

tial HJB formulation of the optimisation problem. Time-consistent coherent risk measures have a Lipschitz driver which coincides with the ellipsoid uncertainty constraint.

The agent who wants to maximise his surplus and acknowledges the ambiguity of the underlying model, acts by the “robust method of pricing” we derived. A practitioner’s methodology to price in incomplete markets is the industry standardised Cost-of-Capital method (Keller and Luder, 2004). This method, that insurance companies use, quantifies the market value of the replicating portfolio plus a mark-up for the unhedgeable risks. It relies mostly on the subjective quantification of risk. Therefore see Filipović and Vogelpoth (2008) for a critical discussion of the Swiss Solvency Test on which the Cost-of-Capital method (CoC) is based. The CoC method leads to a pricing operator that has similar characteristics as our result. The indifference pricing operator from Theorem 3.1 can be interpreted as a best estimate, which is the conditional expectation (if $g(\cdot) = 0$ then we have a linear PDE), plus a constant (c) times the standard deviation of the unhedgeable component. The standard deviation per Δt is a penalty that is added. If $\Delta t \rightarrow 0$ this is normally distributed and hence the penalty consists of the quantiles from the normal distribution. The normality assumption is fulfilled since we are in a diffusion setting. The decomposition of the result from Theorem 3.1 corresponds with the interpretation of the Cost-of-Capital method, where $c\sqrt{\pi'_y S \pi_y} = \text{“CoC”}$ per Δt . Ghalehjooghi and Pelsser (2014) shows that in a diffusion setting the CoC method and the standard deviation pricing principle have the same limit. The interpretation of Lemma 3.1 is discussed in detail in the next section. Each part gets more or less attention is the specific set-up of the examples considered.

3.3 Examples

We show four examples implied by the theorem and lemma derived. The first two examples are the extreme boundary cases. On the one hand we consider a market with only hedgeable risk and on the other hand we consider a market with purely unhedgeable risk. The third example is based on an underlying asset that can be quite general, from real estate to weather indicators. This generality goes beyond the asset-liability management of insurance companies or pension funds only. The fourth example is an insurance contract.

3.3.1 Example I: Pure Hedgeable Risk

The two most extreme cases we can encounter is either purely hedgeable risk or purely unhedgeable risk. The first case is the complete market setting, where we assume that the bank account generates a constant risk-free rate r and thus $dx_0 = rx_0 dt$. If we assume that there are no untradeable assets in the market, $l = 0$. The indifference pricing operator is given by the linear PDE

$$\pi_t + \pi'_x r x + \frac{1}{2} \text{tr}(\pi_{xx} \Sigma_{xx}) - r\pi = 0 \quad (3.49)$$

Note that the drift of the process x is r , the constant risk-free rate, and not μ_x . The pricing process of the replicating portfolio is what is known as “risk-neutral pricing”. By Lemma 3.1 the optimal hedging strategy is

$$\theta^* = \pi_x \quad (3.50)$$

which is the delta-hedging strategy that perfectly replicates the derivative contract. The optimal robustness factors are

$$\epsilon_x^* = -q_x \quad (3.51)$$

which imply the change of the drift to the risk-free rate.

3.3.2 Example II: Pure Unhedgeable Risk

The other extreme is the case when there are no tradeable assets at the agent’s disposal, then $n = 0$. This implies that there are no assets available to hedge. The agent has only the bank account and unhedgeable assets at his disposal. The agent’s best indifference price is to solve the PDE

$$\pi_t + \pi'_y \mu_y + k \sqrt{\pi'_y \Sigma_{yy} \pi_y} + \frac{1}{2} \text{tr}(\pi_{yy} \Sigma_{yy}) - r\pi = 0 \quad (3.52)$$

The optimal hedging portfolio is zero, since the agent is unable to invest in the tradeable assets since there are none. Whereas the unhedgeable risk cannot be traded by definition. The response of mother nature is

$$\epsilon_y^* = \frac{k}{\sqrt{\pi'_y \Sigma_{yy} \pi_y}} \Sigma_{yy} \pi_y \quad (3.53)$$

The drift adjustment ϵ^* is the adjustment of the drift of the unhedgeable process (μ_y) in the “prudent” direction. Mother nature adds a penalty term to the drift proportional to the standard deviation of the unhedgeable risk multiplied by the ambiguity specification k . The adjusted drift is equivalent to the robust pricing method that is known as “actuarial pricing”.

3.3.3 Example III: Correlated Risk

In some cases, if there is only one risk of each type and the partial derivative π_y is either monotonically increasing or decreasing, then the absolute sign of $\sqrt{\pi_y^2} = |\pi_y|$ can be replaced by $\pm\pi_y$ depending on the sign of π_y . The PDE becomes linear and we can express the solution, using the Feynman-Kac formula

$$\pi(t, x, y) = e^{-r(T-t)} \mathbb{E}^{\mathbb{M}} [L(T, x, y) | \mathcal{F}_t] \quad (3.54)$$

where \mathbb{M} is the measure with the adjusted means for both risk factors, i.e. $r \cdot x$ for the x process and $\mu_y + \epsilon_y$ for the y process. The terminal condition is $\pi(T, x, y) = L(T, x, y)$. In the general multidimensional case, the new measure \mathbb{M} can be interpreted as the intersection of the ellipsoid and the risk-neutral measures $\mathcal{Q}^{k^2} \cap \mathcal{Q}^{RN}$, corresponding with the inf-convolution of Barrieu and El Karoui (2005). By specifying the ellipsoid this intersection determines the optimal solution.

Consider a non-traded asset that is correlated with a traded asset. Assume there is a risky asset x , the bank account B and a non-traded asset y with the following dynamics:

$$d \begin{bmatrix} x_t \\ y_t \\ B_t \end{bmatrix} = \begin{bmatrix} \mu_x x_t \\ \mu_y y_t \\ r B_t \end{bmatrix} dt + \begin{bmatrix} \left[\begin{array}{cc} \sigma_x^2 x_t^2 & \rho \sigma_x x_t \sigma_y y_t \\ \rho \sigma_x x_t \sigma_y y_t & \sigma_y^2 y_t^2 \end{array} \right]^{1/2} \\ \mathbf{0}' \end{bmatrix} d \begin{bmatrix} W_x \\ W_y \end{bmatrix} \quad (3.55)$$

By Theorem 3.1 we know that the process of the risky asset will be priced by the risk-free rate, i.e. the drift term changes $\mu_x \rightarrow r$ due to $\epsilon_x = -(\mu_x x_t - r x_t)$. However the ambiguity of the unhedgeable asset is displayed by the ellipsoid $\epsilon' \Sigma^{-1} \epsilon \leq k^2$. Now that ϵ_x is known, i.e. this corresponds with the market price of risk, the adjusted drift term of the unhedgeable asset can be expressed in term of the constant k^2 and the market price of risk. Due to the incompleteness the pricing operator is non-linear, therefore the drift adjustment ϵ_y determines two prices that can be interpreted as either going long or short or as a bid-ask price.

In case the PDE is linear the price is uniquely determined. By Lemma 3.1 the optimal drift adjustments are

$$\begin{bmatrix} \epsilon_x^* \\ \epsilon_y^* \end{bmatrix} = \begin{bmatrix} -\mu_x x_t + r x_t \\ -\rho \sigma_y y_t \frac{\mu_x - r}{\sigma_x} \pm \sigma_y y_t \sqrt{1 - \rho^2} \sqrt{k^2 - \left(\frac{\mu_x - r}{\sigma_x}\right)^2} \end{bmatrix} \quad (3.56)$$

The optimal hedge position is

$$\theta_x^* = \pi_x + \pi_y \frac{\sigma_y y_t}{\sigma_x x_t} \rho \pm \pi_y \frac{\sigma_y y_t}{\sigma_x x_t} \sqrt{\frac{1 - \rho^2}{k^2 - \left(\frac{\mu_x - r}{\sigma_x}\right)^2}} \frac{\mu_x - r}{\sigma_x} \quad (3.57)$$

where the first term π_x is the delta-hedging part linked to purely hedgeable risk, the second term $\pi_y \frac{\sigma_y y_t}{\sigma_x x_t} \rho$ is delta-hedging for the unhedgeable risk weighted by the relative standard deviations, and the last term is the product of the residual of the second term, the unhedgeable risk, $\pi_y \frac{\sigma_y y_t}{\sigma_x x_t} \sqrt{1 - \rho^2}$ and $\frac{(\mu_x - r)/\sigma_x}{\sqrt{k^2 - \left(\frac{\mu_x - r}{\sigma_x}\right)^2}}$. Note that $\frac{(\mu_x - r)/\sigma_x}{\sqrt{k^2 - \left(\frac{\mu_x - r}{\sigma_x}\right)^2}}$ goes to infinity when the market price of risk approaches k , while for small Sharpe ratio's it is approximately the Sharpe ratio scaled down by a factor k . Thus for Sharpe ratio's that are approaching the ones that are extreme, the agent invests huge amounts in the underlying asset.

Since mother nature ensures the robustness of the result, the plus sign in the drift of (3.58) changes the drift term upwards, hence the price will be at its highest point, whereas the minus sign leads to the lower bound. Floroiu and Pelsser (2013) consider a call option on this non-traded asset in particular and highlight the decreasing value of the option in incomplete markets for increasing volatility (Miao and Wang (2007)). Depending on the position of the agent, the robust result is either one of the two bounds. The robust price of the non-traded asset can be characterised by the stochastic process

$$\begin{aligned} dy_t = & \left(\mu_y - \rho \sigma_y \frac{\mu_x - r}{\sigma_x} \pm \sigma_y \sqrt{1 - \rho^2} \sqrt{k^2 - \left(\frac{\mu_x - r}{\sigma_x}\right)^2} \right) y_t dt + \\ & \sigma_y y_t \left(\rho dW_x + \sqrt{1 - \rho^2} dW_y \right) \end{aligned} \quad (3.58)$$

As Floroiu and Pelsser (2013) show, if the variables μ_y , σ_y and ϵ_y are constants, the process y follows a log-normal distribution and corresponds with a dividend paying stock from which the analytical solution is known.

3.3.4 Example IV: Life Insurance Contract

Consider a life insurance contract with one traded and one non-traded asset. In this unit linked contract, the survivors receive the value of the stock at time T bounded by a minimum guarantee (g). In this case $n = 1$, $l = 1$, and S is the stock price that follows a log-normal distribution. For ease of exposition, N , the number of survivors in the policy, also follows a log-normal distribution.³ The stochastic processes are

$$d \begin{bmatrix} S_t \\ N_t \end{bmatrix} = \begin{bmatrix} \mu S_t \\ \alpha N_t \end{bmatrix} dt + \begin{bmatrix} \sigma S_t & 0 \\ 0 & \beta N_t \end{bmatrix} \begin{bmatrix} dW_S \\ dW_N \end{bmatrix} \quad (3.59)$$

We assume no correlation between the two processes, that is no causal relation between the price of the stock and the number of survivors. Since $\frac{\partial \pi}{\partial N} > 0$ at time T , π is monotone increasing in N . Consequently, the PDE is

$$\begin{aligned} \pi_t + \pi'_S r S + \pi_N \left(\alpha N + \sqrt{k^2 - q_S^2 / (\sigma^2 S^2)} \beta N \right) + \\ \frac{1}{2} (\pi_{S^2} \sigma^2 S^2 + \pi_{N^2} \beta^2 N^2) - r \pi = 0 \end{aligned} \quad (3.60)$$

where $q_S = \mu S - r S$ and the terminal condition is

$$L(T, S_T, N_T) = \max(S_T, g) N_T \quad (3.61)$$

By Theorem 3.1, the optimal hedging portfolio is

$$\theta_S^* = \pi_S + \pi_N \frac{\beta N_t}{\sigma S_t} \frac{(\mu - r) / \sigma}{\sqrt{k^2 - \left(\frac{\mu - r}{\sigma}\right)^2}} \quad (3.62)$$

Similar as in the previous example the optimal hedging portfolio consists of a delta-hedging part (π_S), and now that there is no correlation between the hedgeable and unhedgeable risk, the investment in the traded asset proportional to the correlation is zero, the second term of the hedge position $\left(\pi_N \frac{\beta N_t}{\sigma S_t} \frac{(\mu - r) / \sigma}{\sqrt{k^2 - \left(\frac{\mu - r}{\sigma}\right)^2}} \right)$ is the Sharpe ratio relation with k proportional to the relative standard deviations.

We need to solve this for the unknown function $\pi(t, S_t, N_t)$. By solving the PDE we can find the function of $\pi(T - \Delta t, S_t, N_t)$ and its derivatives recursively until time t . The solution can be written as a conditional expectation by the

³Assumption of log-normality for ease of exposition. Can be unrealistic since non-zero probability of $N(t) < N(t + \Delta t)$. However, if the drift is negative (enough) this probability becomes small.

Feynman-Kaç formula

$$\pi(t, S, N) = e^{-r(T-t)} \mathbb{E}^{\mathbb{M}} [\pi(T, S_T, N_T) | \mathcal{F}_t] \quad (3.63)$$

Note that the expectation is taken under the probability measure \mathbb{M} that belongs to the adjusted risk-neutral mean for the hedgeable component and the additional prudence factor for the unhedgeable part. Measure \mathbb{M} belongs to the processes having drifts

$$\begin{cases} rS_t \\ \alpha_0 N_t + \beta N_t \sqrt{k^2 - ((\mu - r)/\sigma)^2} \end{cases} \quad (3.64)$$

The interpretation of risk measure \mathbb{M} is that the financial risk is perfectly replicated such that the ambiguity is eliminated and the mean is replaced by the risk-free return. The ambiguity of the unhedgeable process is now the intersection of the ambiguity set \mathcal{K} and the line r . The intersection of this line and the ellipsoid has two solutions, corresponding with the sign of the liability function. In this case the zero correlation between the two processes causes the ellipsoid to be an exact circle. For this example, we can solve the conditional expectation analytically.

$$\begin{aligned} \pi(t, S_t, N_t) &= e^{-r(T-t)} \mathbb{E}^{\mathbb{M}} [\max(S_T, g) N_T | \mathcal{F}_t] \\ &= \left(S_t \Phi(d_1) - e^{-r(T-t)} g \Phi(d_2) + e^{-r(T-t)} g \right) \cdot \\ &\quad N_t e^{(\alpha + \beta \sqrt{k^2 - ((\mu - r)/\sigma)^2})(T-t)} \end{aligned} \quad (3.65)$$

where $d_{1,2} = \frac{1}{\sigma \sqrt{T-t}} \left(\ln \left(\frac{S_t}{g} \right) + \left(r \pm \frac{\sigma^2}{s} \right) (T-t) \right)$.

3.4 Conclusion

For an agent who wants to maximise his expected surplus, is uncertain about the economy and is allowed to hedge, we obtain the following results. We find a semi-linear PDE for the pricing operator, that can be interpreted as a BSDE with a Lipschitz driver. Hence, the PDE is uniquely solvable for the indifference price. The combination of model ambiguity in a complete market and hedging leads to risk-neutral pricing, implicitly the agent's response eliminates the model ambiguity. Model ambiguity in combination with purely unhedgeable risk results in the action of "actuarial pricing". This is a conservative way of pricing where the uncertain drift is adjusted in the "prudent" direction. The penalty term that

mother nature adds is proportional to the standard deviation of the unhedgeable risk multiplied by the ambiguity specification (the size of the ellipsoid). In a model with both hedgeable and unhedgeable risk the agent will price market-consistently plus actuarial prudentially. The traded risky assets are priced by the risk-free rate. The drift adjustments of the non-traded assets are twofold, either negative or positive. The prudent direction depends on the payoff structure of the agent and can be interpreted as the bid or the ask price. The optimal hedging portfolio consists of a delta-hedging part linked to the purely hedgeable risk and the same delta-hedging term proportional to the correlation between the traded and non-traded assets, plus the product of the residual of the correlated delta-hedging part and a market confidence term. The market confidence term causes the hedging portfolio in the traded risky assets to go to infinity when the Sharpe ratio's become extreme, while moderate market prices of risk lead to an additional term of the Sharpe ratio itself scaled down by the size of the uncertainty set. For some special cases we can solve the pricing semi-linear PDE explicitly.

3.A Optimisation

The optimisation problem from (3.28) is

$$\begin{aligned}
 \max_{\tilde{\theta}} \min_{\epsilon} \quad & \tilde{\theta}' \cdot q + \epsilon'(\tilde{\theta} - \tilde{w}_z) \\
 \text{s.t.} \quad & R\tilde{\theta} = 0 \\
 \text{s.t.} \quad & \epsilon' \Sigma^{-1} \epsilon \leq k^2
 \end{aligned} \tag{3.66}$$

The Lagrangian is

$$L(\tilde{\theta}, \epsilon, \lambda_0, \lambda) = \tilde{\theta}' q + \epsilon'(\tilde{\theta} - \tilde{w}_z) - \lambda_0 \frac{1}{2}(\epsilon' \Sigma^{-1} \epsilon - k^2) - \lambda'(R\tilde{\theta} - 0)$$

where λ has dimension $[l \times 1]$ and λ_0 is a scalar.

The FOC are

$$\frac{\partial L}{\partial \tilde{\theta}} = q + \epsilon - R'\lambda = 0 \tag{3.67}$$

$$\frac{\partial L}{\partial \epsilon} = -\tilde{w}_z + \tilde{\theta} - \lambda_0 \Sigma^{-1} \epsilon = 0 \tag{3.68}$$

$$\frac{\partial L}{\partial \lambda} = -R\tilde{\theta} = 0 \tag{3.69}$$

$$\frac{\partial L}{\partial \lambda_0} = -\frac{1}{2}(\epsilon' \Sigma^{-1} \epsilon - k^2) = 0 \tag{3.70}$$

The first three block of equations are linear. Note that although R is rank deficient, $R\Sigma^{-1}R'$ has full rank.

$$\lambda^* = (R\Sigma^{-1}R')^{-1}R(\Sigma^{-1}q - \lambda_0^{-1}\tilde{w}_z) \tag{3.71}$$

$$\tilde{\theta}^* = \tilde{w}_z + \lambda_0 \Sigma^{-1}(R'\lambda^* - q) \tag{3.72}$$

$$\epsilon^* = R'\lambda^* - q \tag{3.73}$$

Now we can plug in λ^* in ϵ^* and solve the constraint $\epsilon' \Sigma^{-1} \epsilon = k^2$ for λ_0 .

$$\lambda_0^{-1} = \pm \sqrt{\frac{q' \Sigma^{-1} R' (R \Sigma^{-1} R')^{-1} R \Sigma^{-1} q - q' \Sigma^{-1} q + k^2}{\tilde{w}'_z R' (R \Sigma^{-1} R')^{-1} R \tilde{w}_z}} \tag{3.74}$$

Since mother nature is minimising the second order derivative with respect to ϵ , ϵ has to be positive to ensure a minimum, therefore $\lambda_0^{-1} = -\sqrt{\cdot}$. Let $X = R'(R\Sigma^{-1}R')^{-1}R$ be the $[(n+l) \times (n+l)]$ null matrix with the Schur complement

on the lower right $[l \times l]$ block, $\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy} \end{bmatrix}$. Then

$$\begin{aligned} \epsilon^* &= \mathbf{X} \left(-\sqrt{\frac{\mathbf{q}' \Sigma^{-1} \mathbf{X} \Sigma^{-1} \mathbf{q} - \mathbf{q}' \Sigma^{-1} \mathbf{q} + k^2}{\tilde{\mathbf{w}}'_z \mathbf{X} \tilde{\mathbf{w}}_z}} \tilde{\mathbf{w}}_z + \Sigma^{-1} \mathbf{q} \right) - \mathbf{q} \quad (3.75) \\ &= \begin{bmatrix} -\mathbf{q}_x \\ (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \sqrt{\frac{k^2 - \mathbf{q}'_x \Sigma_{xx}^{-1} \mathbf{q}_x}{\tilde{\mathbf{w}}'_y (\Sigma_{yy} - \Sigma'_{xy} \Sigma_{xx}^{-1} \Sigma_{xy}) \tilde{\mathbf{w}}_y}} \tilde{\mathbf{w}}_y - \Sigma'_{xy} \Sigma_{xx}^{-1} \mathbf{q}_x \end{bmatrix} \end{aligned}$$

where the second equality follows from $\mathbf{X} \Sigma^{-1} = \begin{bmatrix} 0 & 0 \\ -\Sigma'_{xy} \Sigma_{xx}^{-1} & 1 \end{bmatrix}$ and $\mathbf{q}' \Sigma^{-1} \mathbf{X} \Sigma^{-1} \mathbf{q} - \mathbf{q}' \Sigma^{-1} \mathbf{q} = [-\mathbf{q}'_x \Sigma_{xx}^{-1} \mathbf{q}_x, \mathbf{0}]'$. The necessary condition, the fact that we prevent the bank account to be strictly better or worse than the risky asset, follows from the admissibility of the square root

$$k^2 - \mathbf{q}'_x \Sigma_{xx}^{-1} \mathbf{q}_x \geq 0 \quad (3.76)$$

Chapter 4

Sets of Indistinguishable Models for Robust Optimisation

Models can be wrong and recognising their limitations is important in financial and economic decision making under uncertainty. Finding the explicit specification of the uncertainty set has been difficult so far. We develop a method that provides a plausible set of models to use in robust decision making. The choice of the specific size of the uncertainty region is what we focus on. We use the Neyman-Pearson Lemma to characterise a set of models that cannot be distinguished statistically from a baseline model. The set of indistinguishable models can explicitly be obtained for a given probability on the Type I and II error.¹

¹This chapter is based on the paper Balter and Pelsser (2015b).

4.1 Introduction

Since models are simplifications of real world processes and situations, there is the risk of a mismatch between the model and the implied strategy. In asset pricing, model uncertainty has implications for the valuation of derivatives and long-dated contracts. In such situations the value of an asset is not unique but falls in a range, requiring extensions of standard pricing methods. We develop a method that provides a plausible set of models surrounding the baseline model to use in robust decision making. Making financial decisions robust calls for applying worst-case scenarios. Therefore even if an optimisation problem is considered under the allowance of model uncertainty, still the question remains how to determine the set of alternative models. The choice of the uncertainty set is what we focus on, contrary to most literature in which a robust control problem is solved for a given set of alternatives. In the second review of Chapter 1 we discussed other methods in the literature that cope with uncertainty. The shared intuition is that an agent is concerned about model misspecification. Consequently, plausible alternatives have to be identified to derive the impact of the uncertainty. Intuitively, these models come from a set surrounding the baseline model. Our objective is to determine explicitly the set that incorporates all plausible alternative models using statistical testing theory. We use the Neyman-Pearson Lemma and impose a Type I and II error to construct the set of indistinguishable models.

The outline is as follows, the introductory section is continued by a short recap of the associated literature that is discussed in Section 1.2. Thereafter the intuition underlying our contribution is described. In Section 4.2 the model is introduced and the indistinguishable set of models is derived for deterministic alternatives that serves as an intuitive illustration. Then an example of a stochastic alternative is shown that fundamentally changes the probability distribution. In Section 4.3 our main contribution is presented in Theorem 4.1 where we create the set of indistinguishable models for stochastic time-consistent alternatives *ex ante*. The link to a numerical bound on several divergences is displayed in Section 4.4.

4.1.1 Literature Review

Hansen and Sargent (2008) proceed on the incorporation of model ambiguity along with the specification of the optimisation problem. Therefore the method depends on the initial objective whereas we concentrate on the creation of a general set of indistinguishable models. However, the motivation is similar. Namely, that alter-

natives surrounding the baseline model up to a constant should be evaluated by the uncertain agent. Hansen and Sargent use bounded entropy (i.e. with bounded Kullback-Leibler divergence).

The several ϕ - (i.e. f -) divergences used as constraint or penalty to admit model uncertainty is what we quantify in Section 4.4. The founder of these functions, that measure the difference between two probability distributions, is Csiszár (1963). Breuer and Csiszár (2013a) solve robust optimal control problems for several divergences that they translate to the moment problem. However the determination of a plausible uncertainty parameter has to be chosen by hand.

The algorithm of removing implausible models from a prespecified set is developed by Hansen et al. (2011). This concept is called the Model Confidence Set approach. This chapter's contribution is the creation of the set of plausible models at time zero without the need of introducing a set of possible models. Moreover, the specification of the test statistic is circumvented.

It is possible to simply construct a confidence interval based on the estimated parameter around the baseline model. However, the limitation is that only a very specific class of models are considered. Namely only those models with parameters that are supposed to be constant over the observation period. The alternative models are models with other constant parameter values. In this chapter we would like to consider alternative models with different structures explicitly, specifically time-dependent and stochastic parameters. By this, the set of alternatives incorporates a large class of different models. Moreover, the confidence interval approach performs the test *ex post* and imposes a Type I error to construct the set of alternative models. We emphasize on the construction of a set of models *ex ante* by imposing both the probability on the Type I and Type II errors.

4.1.2 Intuition

Our goal is to obtain the explicit characterisation of the set of indistinguishable models based on statistical testing theory. We use the Neyman-Pearson Lemma (Neyman and Pearson, 1933) to characterise a set of models that cannot be distinguished statistically from a baseline model. Therefore the set of indistinguishable models can explicitly be obtained *ex ante*, for a given Type I and II error.

Suppose we have an optimisation problem which extends over the time interval $[0, T]$. With other words, suppose there would be T years of extra information available. Our baseline model is specified on a filtered probability space

$(\Omega, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \mathbb{P})$, where \mathbb{P} denotes the probability measure that corresponds to our baseline model. The idea is that we define the plausible set of alternative models, as those models that cannot be distinguished statistically from the baseline model if one would take the observations accumulated over $[0, T]$ into consideration. In other words, we want to exclude ex ante (at time 0) those models which could possibly be rejected by a statistical test procedure at time T with a reasonable level of confidence. We want to use the most powerful tests possible, which are likelihood ratio tests, as stated by the Neyman-Pearson Lemma.

4.2 Statistically Indistinguishable Models

Let us make our model set-up more specific. We assume that we are considering models that can be described by diffusion processes. This means that we are considering stochastic processes X that are described by stochastic differential equations of the form

$$dX(t) = \mu(t, \omega) dt + \sigma(t, \omega) dW(t) \quad (4.1)$$

For the specification of possible alternative models, we consider Brownian motion with a (stochastic) drift process $dW(t) + \lambda(t, \omega) dt$. The ω indicates that λ can be stochastic and may depend on the whole historic path. Such an alternative model specification of the Brownian motion can be captured as a change in probability measure from \mathbb{P} to a new probability measure \mathbb{Q} . With slight abuse of notation we will denote both the alternative model and the alternative probability measure by \mathbb{Q} .²

The likelihood ratio $H_0 : \mathbb{P}$ versus $H_A : \mathbb{Q}$ (based on the information over the interval $[0, T]$) is given by the value of the Radon-Nikodym derivative $R(T)$ at time T . In our diffusion model setting, we know from Girsanov's Theorem that the likelihood ratio (i.e. the Radon-Nikodym derivative) is a stochastic process $R(t) = \frac{d\mathbb{Q}}{d\mathbb{P}}$ which is given by the stochastic differential equation

$$dR(t) = \lambda(t, \omega) R(t) dW^{\mathbb{P}}(t) \quad (4.2)$$

The superscript \mathbb{P} denotes the probability measure we are considering. The solu-

²Note that \mathbb{P} and \mathbb{Q} do not represent the physical and risk-neutral measure respectively. \mathbb{P} stands for the baseline model and \mathbb{Q} for the alternative.

tion to the stochastic differential equation (4.2) can be represented as

$$R(T) = \exp \left\{ -\frac{1}{2} \int_0^T \lambda(t, \omega)^2 dt + \int_0^T \lambda(t, \omega) dW^{\mathbb{P}}(t, \omega) \right\} \quad (4.3)$$

Hence, the value at time T of the likelihood ratio $R(T, \omega)$ is completely determined by the realisation ω of a path of the Brownian motion $\{W^{\mathbb{P}}(t, \omega)\}_{0 \leq t \leq T}$ and the process $\{\lambda(t, \omega)\}_{0 \leq t \leq T}$ of the alternative model \mathbb{Q} along this path.

Based on the realised path of the Brownian motion at time T we could test if model \mathbb{P} should be rejected in favour of model \mathbb{Q} . We are testing two simple hypotheses, and the Neyman-Pearson Lemma tells us that the most powerful test is a likelihood ratio test. The form of the optimal test procedure is that we reject model \mathbb{P} if $R(T)$ is larger than the critical value γ . The critical value γ is determined by the equation

$$\mathbb{P}[R(T) \geq \gamma] = \alpha \quad (4.4)$$

We set the critical value γ such that probability of incorrectly rejecting model \mathbb{P} when model \mathbb{P} is the true model is equal to α . This is known as the Type I error. The probability α is the significance level of the test, and is typically set at 0.05.³

We should also be worried about the Type II error: this is the error of incorrectly rejecting model \mathbb{Q} when model \mathbb{Q} is the true model. This probability is typically denoted by β and can be computed as

$$\mathbb{Q}[R(T) < \gamma] = \beta \quad (4.5)$$

The complement of the Type II error is the probability of accepting model \mathbb{Q} when model \mathbb{Q} is the true model. This is known as the *power* of the statistical test. The power can be computed as

$$\mathbb{Q}[R(T) \geq \gamma] = 1 - \beta \quad (4.6)$$

A typical value for β is 0.20, leading to a statistical power of 0.80.

The Type I and Type II probabilities can be computed ex ante at time 0 for a given alternative model \mathbb{Q} . The model selection procedure we propose is based

³Inclusion or exclusion of the equality sign in the expression $R(T) \geq \gamma$ is potentially relevant when there are point-masses in the probability distribution of $R(T)$. For ease of exposition, we assume that this is not the case. When we do have point-masses we can still handle this mathematically by considering randomised tests.

on the Type II error of the likelihood ratio test. The intuition is as follows. For small values of $\lambda(t, \omega)$ we will have a model \mathbb{Q} that is “close” to the baseline model \mathbb{P} . This closeness can be identified by the fact that the likelihood ratio $R(T)$ will be a random variable with a probability distribution tightly concentrated around the value $R(T) = 1$. Even though we can define a critical value γ for any model \mathbb{Q} , for models that are “close” there will be almost no difference between the \mathbb{P} -probability and the \mathbb{Q} -probability of the event $R(T) \geq \gamma$. Hence, the power of the statistical test will be very low. In the limiting case when $\mathbb{P} = \mathbb{Q}$ the power of the (randomised) likelihood ratio test will be as low as α .

Hence, our model selection criterion will include all models for which the statistical power $\mathbb{Q}[R(T) \geq \gamma]$ is below $1 - \beta$. We consider these models to be *statistically indistinguishable* from the baseline model \mathbb{P} . By imposing the α , for each deviation λ the critical value γ is defined. If the associated power is too high the λ is excluded from the set of indistinguishable models and vice versa.

We can express $\mathbb{Q}[R(T) \geq \gamma]$ also as $\mathbb{E}^{\mathbb{P}}[R(T)\mathbf{1}(R(T) \geq \gamma)]$, and we obtain an interpretation for the \mathbb{P} -expectation as the Tail-Value-at-Risk (TVaR) or Conditional-Value-at-Risk (CVaR) of the random variable $R(T)$ with a confidence level of α . Hence, if we put an upper bound of $1 - \beta$ on the power of the likelihood ratio test, this is equivalent to restricting the TVaR of the Radon-Nikodym derivative $R(T)$ to $1 - \beta$. TVaR is a *coherent risk measure* (Artzner et al. (1999) and Rockafellar and Uryasev (2000, 2002)), which has attractive properties. For example, the acceptance set (that is the set of all $R(T)$ for which $\text{TVaR}(R(T)) \leq 1 - \beta$) is a closed convex set. Hence, we obtain immediately that our set of indistinguishable models is also a closed convex set.

For deterministic $\lambda(t)$ we can compute everything explicitly, though our main goal is to find the set of indistinguishable models for stochastic $\lambda(t, \omega)$. To illustrate our general idea we first consider the deterministic case.

4.2.1 Deterministic Drift Term

When mother nature is only allowed to use alternative models with a deterministic drift term $\lambda(t)$, we can compute the probability distribution of the Radon-Nikodym derivative defined in equation (4.2) explicitly. For this case we obtain

$$R(T) = \exp \left\{ -\frac{1}{2} \int_0^T \lambda(t)^2 dt + \int_0^T \lambda(t) dW^{\mathbb{P}}(t) \right\} \quad (4.7)$$

In particular, $\ln R(T)$ has a normal distribution with mean $-\frac{1}{2} \int_0^T \lambda(t)^2 dt$ and variance $\int_0^T \lambda(t)^2 dt$. The likelihood ratio test procedure $R(T) \geq \gamma$ is equivalent to performing a test on the statistic

$$r(T) = \int_0^T \lambda(t) dW^{\mathbb{P}}(t) \quad (4.8)$$

This test statistic is intuitively appealing: we compute the inner product between the model-drift $\lambda(t)$ and the realised changes $dW(t)$ in the Brownian motion along the whole path $[0, T]$. Every time $\lambda(t)$ and $dW(t)$ have the same sign, this increases the value of $r(T)$. Hence, if model \mathbb{Q} is true, then $r(T)$ will on average have a positive value. *Ex post*, the realisation of the path of $W(T)$ is observed and indicates the likelihood whether it was generated by model \mathbb{P} or \mathbb{Q} . However, *ex ante* the test will not be conducted but rather serves as a hypothetical test. Note that before time T the test statistic is a random variable.

The test statistic $r(T)$ (under model \mathbb{P}) has a normal distribution with mean 0 and variance $\int_0^T \lambda(t)^2 dt$. The hypothesis \mathbb{P} is rejected if $r(T) \geq \gamma$. By imposing a significance level α the critical value γ can be derived analytically. Under the alternative model \mathbb{Q} the test statistic has a normal distribution with mean $\int_0^T \lambda(t)^2 dt$ and variance $\int_0^T \lambda(t)^2 dt$. The power of the test can be computed explicitly as

$$\mathbb{Q}[r(T) \geq \gamma] = \Phi \left(\Phi^{-1}(\alpha) + \left(\int_0^T \lambda(t)^2 dt \right)^{\frac{1}{2}} \right) \quad (4.9)$$

For the case $\lambda(t) \equiv 0$ we see that the power is $\Phi(\Phi^{-1}(\alpha)) = \alpha$. For non-zero values of $\lambda(t)$ the expression $(\int_0^T \lambda(t)^2 dt)^{\frac{1}{2}}$ is strictly positive and therefore the power will be larger than α . If we consider all models with a power below $1 - \beta$ as indistinguishable, then the class of indistinguishable models (with deterministic $\lambda(t)$) is given by all models for which the L_2 -norm $(\int_0^T \lambda(t)^2 dt)^{\frac{1}{2}}$ is below a certain threshold.

If we take for example $\alpha = 0.05$, then $\Phi^{-1}(\alpha) = -1.64$. If we take $\beta = 0.20$ then the power is 0.80 and we have $\Phi^{-1}(0.80) = 0.84$. Hence, the class of all indistinguishable models is then given by all models that satisfy $(\int_0^T \lambda(t)^2 dt)^{\frac{1}{2}} \leq 0.84 - (-1.64) = 2.48$, if one would have T years of extra data.

The deterministic example we have formulated can be generalised easily to the *multi-dimensional case*. For a vector-valued Brownian motion all alternative mod-

els are specified by the deterministic vector-valued process $\lambda(t)$. The test statistic $r(t)$ is then given by

$$r(T) = \int_0^T \lambda(t) \cdot dW^{\mathbb{P}}(t) \quad (4.10)$$

This is also a random variable with mean 0 and variance $\int_0^T |\lambda(t)|_2^2 dt$, where $|\lambda(t)|_2$ denotes the L_2 -norm of the vector $\lambda(t)$. Hence, in the multi-dimensional case the set of indistinguishable models is given by all models for which $(\int_0^T |\lambda(t)|_2^2 dt)^{\frac{1}{2}}$ is below the same threshold as in the one-dimensional case (e.g. $2.48/\sqrt{T}$). We may choose the power dependent on the number of dimensions.

4.2.2 Stochastic Drift Term

The deterministic $\lambda(t)$ serves as an intuitive illustration, but our ambition is to consider a much larger class of alternative models: $\lambda(t, \omega)$. If we allow for stochastic $\lambda(t, \omega)$ then a very large class of alternative models is accessible over an interval $[0, T]$. By the Martingale Representation Theorem *any* probability distribution (with support on whole \mathfrak{R}) can be attained over an interval $[t, t + \varepsilon]$ with $\varepsilon > 0$.

Let us consider a model with stochastic $\lambda(t, \omega)$. Suppose we consider the random variable

$$R(T) := e^{-\frac{1}{2}a^2T} \cosh(aW(T)) = \frac{1}{2} \left(e^{-\frac{1}{2}a^2T+aW(T)} + e^{-\frac{1}{2}a^2T-aW(T)} \right) \quad (4.11)$$

which is strictly positive and has expectation $\mathbb{E}[R(T)] = 1$. Hence, this is a valid Radon-Nikodym derivative. This $R(T)$ corresponds to a \mathbb{Q} -model where the probability distribution of $W(T)$ at time T is given by a mixture distribution of two normal distributions with mean $+aT$ and $-aT$, the same variance T , and mixing probabilities $\frac{1}{2}$. Note that this mixture distribution is not a normal distribution, and has mean 0 and variance equal to $T + (aT)^2$ (see Appendix 4.A), which is larger than the variance T under the \mathbb{P} -model. Hence, mother nature considers alternatives with fatter tails relative to the baseline. Mean-repelling processes are plausible models in finance and economics.

Although this is a very simple example, it shows explicitly that with a stochastic $\lambda(t, \omega)$ we can fundamentally alter the properties of the probability distribution of $W(T)$, beyond only changing the mean of the normal distribution.

4.3 Stochastic and Time-Consistent Set of Indistinguishable Models

The deterministic case and the stochastic example served as an illustration of our approach. However, we are interested in the generalisation of this method. In other words, we would like to allow for a wide class of alternative models. Since stochastic alternatives can lead to fundamentally different models, we would like to characterise the set of stochastic alternative models surrounding the baseline model that are indistinguishable based on an insufficient power and sufficient probability on the Type I error.

4.3.1 Time-Consistency

One of the main motivations for studying the set of statistically indistinguishable sets are robust solutions to stochastic optimal control problems in economics and in financial markets.

When we are solving optimal control problems, we want to consider solutions that are *time-consistent*. This means that the optimal solution at any time-point $0 < t < T$ does not depend on the history of the process between $[0, t]$. In other words, the optimal policy devised at time 0 for the interval $[0, T]$ is still valid at time t given the information \mathcal{F}_t .

The set of indistinguishable models we have defined thus far is *not* time-consistent: the set is defined as those models that have sufficiently low power at time T using the information over the whole path $[0, T]$. We have established in Section 4.2 that the set of indistinguishable models defines a coherent risk measure. But, this risk measure is not time-consistent since it is “static” at time 0 and at time $0 < t < T$ there is no connection between the Type I and II errors and the characterisation of the set of alternative models.

However, we can look at a smaller class of risk measures: the class of time-consistent risk measures. This class has been extensively studied in recent years, and we know how to characterise this class of risk measures. Delbaen (2006) proves that time-consistent (coherent and convex) risk measures are generated by m -stable sets of probability measures. A similar structure (albeit with less mathematical rigour) was already proposed by Epstein and Schneider (2003). An alternative characterisation of time-consistent risk measures has been provided by Gianin (2006). She proves that every time-consistent risk measure is equivalent

to a g -expectation $\mathcal{E}^g[\cdot]$. These non-linear g -expectations can be computed as the solution of a backward stochastic differential equation (BSDE) with a driver $g(t, Y, Z)$. A further characterisation has been provided by Barrieu and El Karoui (2009): they prove that time-consistent coherent risk measures are generated by drivers $g(t, Z)$ that satisfy a Lipschitz growth constraint in Z , and time-consistent convex risk measures are generated by drivers that satisfy a quadratic growth constraint in Z .

Hence, we propose to intersect the class of indistinguishable models (which are coherent, but not time-consistent) with the collection of time-consistent coherent risk measures. Since the objective can be interpreted as the coherent risk measure TVaR/CVaR. We then obtain the set of time-consistent indistinguishable models. The question is now: how can we obtain an explicit characterisation of this intersection?

4.3.2 Maximum Power Calculation

We obtain an explicit characterisation in the following way. The class of time-consistent coherent risk measures are generated by BSDE's with drivers that satisfy the Lipschitz growth condition $g(t, Y, Z) \leq k|Z|$. This is equivalent to the class of Radon-Nikodym derivatives with kernels $|\lambda(t, \omega)| \leq k$ (Barrieu and El Karoui, 2009).

We want to investigate the maximum power that can be achieved within the class of Radon-Nikodym derivatives with $|\lambda(t, \omega)| \leq k$, such that the Type I error is equal to α . We can formulate this as a stochastic optimisation problem of the form

$$\begin{aligned} \max_{\gamma, |\lambda(t, \omega)| \leq k} & \mathbb{E}[R(T)\mathbf{1}(R(T) \geq \gamma)] & \text{(MP)} \\ \text{s.t.} & \mathbb{E}[\mathbf{1}(R(T) \geq \gamma)] = \alpha \\ & dR = \lambda(t, \omega)RdW, R_0 = 1 \end{aligned}$$

The objective function is the power of the test $R(T) \geq \gamma$ formulated as a P-expectation. The second line gives the Type I error (also formulated as a P-expectation), the third line describes the stochastic process for the Radon-Nikodym derivative given the control variable $\lambda(t, \omega)$. The optimisation problem is non-convex due to the indicator function. Since it is easier to work with convex

functions, we introduce the auxiliary function

$$\begin{aligned}
 F_\alpha(R, \gamma) &= \alpha\gamma + \mathbb{E} \left[(R - \gamma)^+ \right] \\
 &= \alpha\gamma + \mathbb{E} [R\mathbf{1}(R(T) \geq \gamma)] - \gamma\mathbb{E} [\mathbf{1}(R(T) \geq \gamma)] \\
 &= \mathbb{E} [R\mathbf{1}(R(T) \geq \gamma)] + (\alpha - \mathbb{E} [\mathbf{1}(R(T) \geq \gamma)])\gamma
 \end{aligned} \tag{4.12}$$

where $(R - \gamma)^+$ denotes $\max(R - \gamma, 0)$. The functional $F_\alpha(R, \gamma)$ is convex in and continuous as a function in $\gamma \in \mathbb{R}$ and $R(T) \in L_2(T) : \mathbb{E}[R(T)^2] < \infty$. This is also shown by Rockafellar and Uryasev (2000, 2002) who introduce a similar auxiliary function to minimise the Conditional-Value-at-Risk (CVaR).

In order to solve the constrained optimisation problem (MP) we solve

$$\begin{aligned}
 \max_{|\lambda(t,\omega)| \leq k} \min_{\gamma} F_\alpha(R, \gamma) & \tag{MaMi} \\
 \text{s.t. } dR = \lambda(t, \omega)RdW &
 \end{aligned}$$

The optimisation (MaMi) is equivalent to (MP), this is proven by Rockafellar and Uryasev (2000, 2002). They prove that the CVaR can be obtained by rewriting the optimisation problem in terms of a convex auxiliary function. The Type I error constraint is satisfied by minimisation of the auxiliary function over γ

$$\frac{\partial F_\alpha(R, \gamma)}{\partial \gamma} = \alpha - \mathbb{E} [\mathbf{1}(R(T) \geq \gamma)] = 0 \tag{4.13}$$

Since $F_\alpha(R, \gamma)$ is convex in γ , the extreme value is a minimum. In particular, after minimisation we have

$$\begin{aligned}
 \min_{\gamma} F_\alpha(R, \gamma) &= \min_{\gamma} \mathbb{E} [R(T)\mathbf{1}(R(T) \geq \gamma)] + (\alpha - \alpha)\gamma \\
 &= \mathbb{E} [R(T)\mathbf{1}(R(T) \geq \gamma)]
 \end{aligned} \tag{4.14}$$

This is the power that we would like to maximise. Henceforth, we continue by maximising the minimised $F_\alpha(R, \gamma)$. However, because this is nontrivial we would prefer to change the order of the optimisation. Solving the initial nonconvex optimisation problem is identical with solving the max-min. The optimal value from the (MaMi) is always lower than or equal to the reversed order of the optimisation, by the max-min inequality.

Let (MiMa) be

$$\begin{aligned} \min_{\gamma} \max_{|\lambda(t,\omega)| \leq k} F_{\alpha}(R, \gamma) & \quad (\text{MiMa}) \\ \text{s.t. } dR = \lambda(t, \omega) R dW & \end{aligned}$$

Then, we can summarise the relations between the different optimisation formulations by

$$(\text{MP}) = (\text{MaMi}) \leq (\text{MiMa}) \quad (4.15)$$

If we solve the (MiMa) we find an upper bound on (MaMi) and (MP). First we solve the (MiMa). In specific, we start with the inner maximisation problem

$$\begin{aligned} \max_{|\lambda(t,\omega)| \leq k} F_{\alpha}(R, \gamma) & \quad (4.16) \\ \text{s.t. } dR = \lambda(t, \omega) R dW & \end{aligned}$$

The inner maximisation is solved by formulating it as a Hamilton-Jacobi-Bellman (HJB) problem. Note that for every fixed $\gamma \in \mathbb{R}$ the function $F_{\alpha}(R, \gamma)$ is convex in R . We introduce the value function $V(t, r, \gamma) = \mathbb{E}[F_{\alpha}(R(T), \gamma) | R(t) = r]$ with boundary condition $V(T, r, \gamma) = F_{\alpha}(R(T), \gamma)$.

For any value of α and γ the optimised function $V(t, r, \gamma)$ with respect to $\lambda(t, \omega)$ for $t \leq T$ solves the HJB-equation

$$V_t + \max_{|\lambda(t,r)| \leq k} \frac{1}{2} \lambda(t, r)^2 r^2 V_{rr} = 0 \quad (4.17)$$

Optimal control problems of this sort have been studied in the literature in the context of uncertain volatility models (Avellaneda et al. (1995) and Vanden (2006)), and are called *Black-Scholes-Barenblatt* equations. The general solution is

$$|\lambda^*(t, r)| = \begin{cases} k & \text{for } V_{rr}(t, r, \gamma) \geq 0 \\ 0 & \text{for } V_{rr}(t, r, \gamma) < 0 \end{cases} \quad (4.18)$$

We propose $|\lambda(t, \omega)| = k$ as candidate solution. The analytical expression for the value function is

$$V(t, r, \gamma) = \alpha\gamma + rN(d_1) - \gamma N(d_2) \quad (4.19)$$

where $d_1 = \frac{1}{k\sqrt{T-t}} \left(\ln\left(\frac{r}{\gamma}\right) + \frac{1}{2}k^2(T-t) \right)$ and $d_2 = d_1 - k\sqrt{T-t}$. This function

solves the HJB-equation for the boundary condition. The value function is convex as

$$V_{rr}(t, r, \gamma) = \frac{n(d_1)}{rk\sqrt{T-t}} \geq 0 \quad \forall r, t < T \quad (4.20)$$

where $n(\cdot)$ is the standard normal density function.

The Verification Theorem 11.2.2 of Øksendal (2003) states that if $V(t, r, \gamma)$ is uniformly integrable then a solution of the HJB is an optimal control. The value function $V(t, r, \gamma)$ is indeed uniformly integrable. Since $R(T)$ itself is uniformly integrable for all $\lambda(t, \omega)$, consequently also the partial moment $V(t, r, \gamma) = \mathbb{E}[F_\alpha(R, \gamma) | R(t) = r]$ is uniformly integrable. The Radon-Nikodym derivative is uniformly integrable because the expectation $\mathbb{E}[R(T)^2] \leq e^{k^2 T} < \infty$ as $|\lambda(t, \omega)| \leq k$. Thus $|\lambda(t, \omega)| = k$ is an optimal control that leads to the maximum power of (4.16).

The optimal value function is $V(0, 1, \gamma) = \mathbb{E}[F_\alpha(R(T), \gamma) | \mathcal{F}_0] = \alpha\gamma + \mathbb{E}[(R(T) - \gamma)^+ | R(0) = 1]$. Hence, this is the optimal power and can be interpreted as a log-normal expectation equal to a constant plus the expected value of a Black-Scholes call option.

The remaining part is the outer minimisation

$$\min_{\gamma} V(0, 1, \gamma) \quad (4.21)$$

This simple calculation leads to the optimal γ^* to be the $(1-\alpha)$ -quantile of the log-normal $R(T)$. This proves that the Radon-Nikodym derivative with $|\lambda(t, \omega)| \equiv k$ for all $0 \leq t \leq T$ achieves the highest possible power within the constraint on the Type I error. Note that the log-normal Radon-Nikodym derivative is not the unique solution that leads to the optimum. Though all that is needed is the optimal value of the norm of $\lambda(t, \omega)$ which is unique.

We can conclude that the upper bound on (MaMi) is the log-normal power of (MiMa). The original problem (MP) is equivalent (MaMi). Since $|\lambda^*(t, R^*)| = k$ is a feasible solution of the original problem, the upper bound on the power is attained. Hence, the inequality between (MaMi) and (MiMa) becomes an equality and the maximum power possible is obtained for $|\lambda^*(t, R^*)| = k$. ■

Based on the derivation for the maximum power, all $\lambda(t, \omega)$ s that yield a lower power are indistinguishable. We summarise this in Theorem 4.1.

Theorem 4.1 (Sets of indistinguishable models)

The set of time-consistent models on $[0, T]$ with

$$|\lambda(t, \omega)| \leq \frac{\Phi^{-1}(1 - \beta) - \Phi^{-1}(\alpha)}{\sqrt{T}} \quad (4.22)$$

is indistinguishable from the baseline model for a Type I error of α and a Type II error of β .

Proof of Theorem 4.1 Given that the optimal $R^*(T)$ is a log-normal martingale with volatility k , then the optimal value for γ is equal to the $(1 - \alpha)$ -quantile of $R^*(T)$. The optimised power at time $t = 0$ with $R(0) = 1$ is therefore equal to

$$\mathbb{E} [R^*(T)\mathbf{1}(R^*(T) \geq \gamma^*)] = \mathbb{Q} [R^*(T) \geq \gamma^*] = \Phi \left(\Phi^{-1}(\alpha) + k\sqrt{T} \right) \quad (4.23)$$

If we want to construct a class of time-consistent coherent risk measures with stochastic $\lambda(t, \omega)$ such that every element in the class is statistically indistinguishable, then a sufficient condition is to choose a k such that the “worst-case” power (4.23) does not exceed $(1 - \beta)$. ■

If we set $\alpha = 0.05$ and $\beta = 0.20$, Theorem 4.1 implies that all $\lambda(t, \omega)$'s are considered as plausible alternatives that are in between $-2.48/\sqrt{T}$ and $2.48/\sqrt{T}$. All previous derivations also hold for vector Brownian motions and vector $\lambda(t, \omega)$. The multivariate equivalent is a bound on the L2-norm; $|\lambda(t, \omega)| \leq \frac{2.48}{\sqrt{T}}$.

4.3.3 Sufficiency

The log-normal Radon-Nikodym derivative with maximal power for stochastic alternatives is a sufficient rather than a necessary condition. This implies that there exist alternative models with a norm above k and still admit a power below $1 - \beta$. For deterministic deviations the log-normal solution is both sufficient and necessary. Henceforth, larger deviations are allowed in the set of indistinguishable models.

The stochastic example of Section 4.2.2 illustrates that the bound on the norm of the stochastic alternatives is a sufficient condition to determine the set of indistinguishable models. For the hyperbolic cosine as choice for $R(T)$ we can explicitly compute the conditional expectation $R(t) = \mathbb{E}^{\mathbb{P}} [R(T) | \mathcal{F}_t] = e^{-\frac{1}{2}a^2t} \cosh(aW(t))$.

If we apply Itô's Lemma to $R(t)$ we obtain the stochastic differential equation

$$dR(t) = ae^{-\frac{1}{2}a^2t} \sinh(aW(t))dW(t) = a \tanh(aW(t))R(t)dW(t) \quad (4.24)$$

Hence, this Radon-Nikodym derivative corresponds to a model where $\lambda(t, \omega) = a \tanh(aW(t))$. For positive values of $W(t)$ we have a positive and increasing drift, and the drift is bounded by $+a$. For negative values of $W(t)$ we have a negative and increasing drift, and the drift is bounded by $-a$. Hence, the alternative model \mathbb{Q} is a “mean-repelling” process, which will increase the variance of $W(T)$ under model \mathbb{Q} .

The likelihood ratio test will reject model \mathbb{P} if $R(T) \geq \gamma$. As $\cosh(aW(T))$ is symmetric around $W(T) = 0$ and strictly increasing in $|W(T)/\sqrt{T}|$, the rejection set defined by $R(T) \geq \gamma$ is equivalent to the rejection set $|W(T)/\sqrt{T}| \geq \gamma'$. If we want to test at a significance level of $\alpha = 0.05$ then $\gamma' = -\Phi^{-1}(\alpha/2) = 1.96$.

The power of the likelihood ratio test can be computed as $\mathbb{Q}[|W(T)/\sqrt{T}| \geq \gamma']$ which can be expressed as $\mathbb{E}^{\mathbb{Q}}[\mathbf{1}(|W(T)/\sqrt{T}| \geq \gamma')] = \mathbb{E}^{\mathbb{P}}[R(T)\mathbf{1}(W(T)/\sqrt{T} \geq \gamma')] + \mathbb{E}^{\mathbb{P}}[R(T)\mathbf{1}(W(T)/\sqrt{T} \leq -\gamma')]$. A direct computation of the expectations yields

$$\mathbb{Q}[|W(T)/\sqrt{T}| \geq \gamma'] = \Phi(\Phi^{-1}(\alpha/2) + a\sqrt{T}) + \Phi(\Phi^{-1}(\alpha/2) - a\sqrt{T}) \quad (4.25)$$

see Appendix 4.B for the full derivation. If we solve this last equation (numerically) for $a\sqrt{T}$ with $\alpha = 0.05$ and $1 - \beta = 0.80$ then we find $a\sqrt{T} = 2.80$. Hence, for the tanh example we find the result that all models that are indistinguishable from the baseline model $a = 0$ are given by $|a\sqrt{T}| \leq 2.80$. This set is larger than the “constant lambda” set $|\lambda\sqrt{T}| \leq 2.48$. Thus the conclusion for this example is that the bound $|\lambda\sqrt{T}| \leq 2.48$ gives a sufficient condition for statistical indistinguishability.

4.4 Bounds on Divergences

Ben-Tal et al. (2013) discuss several possible divergences (non-symmetric distance measures) to generate robust results in optimisation problems. ϕ -Divergence (or f -divergence) functions measure the distance between two probability distributions weighted by the specific function. The choice which measure should be picked is an unanswered issue in their and many other papers, plus the question when the distance is too far is rarely investigated. In this chapter we explicitly focus on the

size of the set of alternatives. We can directly link this to a critical value for each measure. Note that Ben-Tal et al. (2013) consider the discrete versions, (4.26), whereas we consider the continuous divergences, (4.27) and (4.28), as a function of the Radon-Nikodym derivative. The numerical value for the set of time-consistent models that cannot be distinguished between with a Type I error of 5% and a power less than 80% is displayed in the last column of Table 4.1. The derivations can be found in Appendix 4.C.

The general discrete ϕ -divergence is defined as

$$D_\phi(p, q) = \sum_{i=1}^m q_i \phi\left(\frac{p_i}{q_i}\right) \quad (4.26)$$

We use the continuous version

$$D_\phi(p, q) = \mathbb{E}^{\mathbb{Q}} \left[\phi\left(\frac{1}{R(T)}\right) \right] \quad (4.27)$$

$$= \mathbb{E}^{\mathbb{P}} \left[R(T) \phi\left(\frac{1}{R(T)}\right) \right] \quad (4.28)$$

where the functions for $\phi(\cdot)$ are given for each measure and by definition $\phi(\cdot)$ is convex. Note that we use $\tilde{\phi}(t) = t\phi\left(\frac{1}{t}\right)$ which are, for the divergences considered, convex as well as every $\tilde{\phi}$ coincides with a ϕ . The variable t is set here to $R(T)$.

The lay-out of the proof for each divergence function is similar as the proof of the maximum power calculation. First we postulate the log-normal Radon-Nikodym derivative as candidate solution. Then the analytical expression for the value function is used to prove its convexity, which implies that the proposed solution $|\lambda(t, \omega)| = k$ solves the HJB. By the Verification Theorem 11.2.2 of Øksendal (2003) and the uniform integrability of the value function the optimal control is obtained. ■

Table 4.1: ϕ -Divergences

Divergence	$\phi(t)$	for $ \lambda(t, \omega) = k$	$k\sqrt{T} = 2.48$
Kullback-Leibler	$t \ln t - t + 1$	$\frac{1}{2}k^2T$	3.08
Burg entropy	$-\ln t + t - 1$	$\frac{1}{2}k^2T$	3.08
J-divergence	$(t - 1) \ln t$	k^2T	6.15
χ^2 -distance	$\frac{1}{t}(t - 1)^2$	$e^{k^2T} - 1$	467.90
Modified χ^2 -distance	$(t - 1)^2$	$e^{k^2T} - 1$	467.90
Hellinger distance	$(\sqrt{t} - 1)^2$	$2 - 2e^{-\frac{1}{8}k^2T}$	1.07
Variation distance	$ t - 1 $	$4N(\frac{1}{2}k\sqrt{T}) - 2$	1.57
χ -divergence of order $\theta > 1$	$ t - 1 ^\theta$	–	see Table 4.2
Cressie-Read $\theta \neq 0, 1$	$\frac{1 - \theta + \theta t - t^\theta}{\theta(1 - \theta)}$	see below	see Table 4.2

Both the χ -divergence of order $\theta > 1$ and the Cressie-Read divergence depend on the additional parameter θ . The Cressie-Read divergence can analytically be expressed by

$$\frac{1}{\theta(1 - \theta)} \left(1 - e^{-\frac{1}{2}k^2\theta(1 - \theta)T} \right) \quad (4.29)$$

and the χ -divergence can be derived analytically for integer values of θ as well. For $k\sqrt{T} = 2.48$ we get the bounds on both measures displayed in Table 4.2.

Table 4.2: Numerical bounds

Divergence θ	1.5	2.0	2.5	3.0
χ -divergence of order θ	10.40	467.90	1.02×10^6	1.03×10^8
Cressie-Read	12.05	233.95	2.72×10^4	1.72×10^7

Possible applications of these measures are optimal control problems where the agent is worried about possible misspecification of the baseline model. The uncertainty could be described by an additional constraint such that the Kullback-Leibler divergence between the baseline model and all alternatives has to be bounded. The cutoff point to differentiate plausible alternatives from implausible ones is what we derived. If one would impose a probability of incorrectly rejecting the baseline model of 5% and if one would need at least a power of 80% to correctly accept the alternative, then for all models that yield a lower power, the information is insufficient to distinguish between \mathbb{P} and \mathbb{Q} . Hence the cutoff point on the Kullback-Leibler divergence equals 3.08. The same reasoning holds for the other divergences.

4.5 Conclusion

By imposing probabilities on the Type I and II error of the likelihood ratio test, we are able to quantify uncertainty explicitly. Hence if an agent acknowledges that his model might be misspecified, he would like to evaluate the optimal decision rule against the worst-case model among the plausible alternatives to incorporate robustness. Applications that build upon the uncertainty need the specific set of indistinguishable models. Examples can be found on a wide range to price and hedge in incomplete markets, for instance long-dated insurance contracts or illiquid assets.

We used the Neyman-Pearson Lemma to characterise a set of models that cannot be distinguished statistically from a baseline model. Both deterministic and time-consistent stochastic alternatives are proven to have maximal power for a log-normal Radon-Nikodym derivative with bounded volatility. Allowing for stochastic alternatives yields a tremendous enlargement of the class of alternative models that will be considered to be indistinguishable. The set of indistinguishable models can explicitly be obtained *ex ante*, for given Type I and II probabilities. The result can be linked to quantify bounds on ϕ -divergences such as the Kullback-Leibler divergence.

4.A Variance Hyperbolic Cosine

Test

$$\begin{aligned}
 H_0 : \mathbb{P} & \quad \text{versus} \quad H_A : \mathbb{Q} \\
 H_0 : W(T) & \quad \text{versus} \quad H_A : W(T) + \int_0^T \lambda(s, \omega) ds
 \end{aligned} \tag{4.30}$$

Where $W(T) \sim N(0, T)$. The moment generating function is

$$\begin{aligned}
 M_{\mathbb{P}}(t) &= \mathbb{E}^{\mathbb{P}} \left[e^{tW(T)} \right] \\
 &= e^{\frac{1}{2}Tt^2} \\
 \mathbb{E}[W(T)] &= \frac{\partial M}{\partial t}(0) = \left. Tte^{\frac{1}{2}Tt^2} \right|_{t=0} = 0 \\
 \mathbb{E}[W(T)^2] &= \frac{\partial^2 M}{\partial t^2}(0) = \left. T^2t^2e^{\frac{1}{2}Tt^2} + Te^{\frac{1}{2}Tt^2} \right|_{t=0} = T
 \end{aligned} \tag{4.31}$$

And under \mathbb{Q}

$$\begin{aligned}
 M_{\mathbb{Q}}(t) &= \mathbb{E}^{\mathbb{P}} \left[e^{tW(T)} R(T) \right] \\
 &= \mathbb{E}^{\mathbb{P}} \left[e^{tW(T)} \frac{1}{2} \left(e^{-\frac{1}{2}a^2T+aW(T)} + e^{-\frac{1}{2}a^2T-aW(T)} \right) \right] \\
 &= \frac{1}{2} \left(e^{-\frac{1}{2}a^2T+(a+t)^2\frac{1}{2}T} + e^{-\frac{1}{2}a^2T+(-a+t)^2\frac{1}{2}T} \right) \\
 \mathbb{E}[W(T)] &= \frac{\partial M}{\partial t}(0) = \frac{1}{2} \left((a+t)Te^{-\frac{1}{2}a^2T+(a+t)^2\frac{1}{2}T} \right) + \\
 & \quad \frac{1}{2} \left((-a+t)Te^{-\frac{1}{2}a^2T+(-a+t)^2\frac{1}{2}T} \right) \Big|_{t=0} \\
 &= 0 \\
 \mathbb{E}[W(T)^2] &= \frac{\partial^2 M}{\partial t^2}(0) = \frac{1}{2} \left((a(a+t)T^2 + T)e^{-\frac{1}{2}a^2T+(a+t)^2\frac{1}{2}T} \right) + \\
 & \quad \frac{1}{2} \left((-a(-a+t)T^2 + T)e^{-\frac{1}{2}a^2T+(-a+t)^2\frac{1}{2}T} \right) \Big|_{t=0} \\
 &= (aT)^2 + T
 \end{aligned} \tag{4.32}$$

4.B Power Hyperbolic Cosine

Under \mathbb{Q} the probability distribution of $W(T)/\sqrt{T}$ is the mixture distribution $W(T)/\sqrt{T} \sim \frac{1}{2}N(a\sqrt{T}, 1) + \frac{1}{2}N(-a\sqrt{T}, 1)$. Then

$$\begin{aligned}
 \mathbb{Q} \left[|W(T)/\sqrt{T}| \geq \gamma' \right] &= \mathbb{Q} \left[W(T)/\sqrt{T} \geq \gamma' \right] + \mathbb{Q} \left[W(T)/\sqrt{T} \leq -\gamma' \right] \\
 \mathbb{Q} \left[W(T)/\sqrt{T} \geq \gamma' \right] &= \frac{1}{2} \mathbb{Q} \left[W(T)/\sqrt{T} - a\sqrt{T} \geq \gamma' - a\sqrt{T} \right] + \\
 &\quad \frac{1}{2} \mathbb{Q} \left[W(T)/\sqrt{T} + a\sqrt{T} \geq \gamma' - a\sqrt{T} \right] \\
 &= \frac{1}{2} \mathbb{Q} \left[W(T)/\sqrt{T} - a\sqrt{T} \leq -\gamma' + a\sqrt{T} \right] + \\
 &\quad \frac{1}{2} \mathbb{Q} \left[W(T)/\sqrt{T} + a\sqrt{T} \leq -\gamma' + a\sqrt{T} \right] \\
 &= \frac{1}{2} \Phi(-\gamma' + a\sqrt{T}) + \frac{1}{2} \Phi(-\gamma' - a\sqrt{T}) \\
 \mathbb{Q} \left[W(T)/\sqrt{T} \leq -\gamma' \right] &= \frac{1}{2} \mathbb{Q} \left[W(T)/\sqrt{T} - a\sqrt{T} \leq -\gamma' - a\sqrt{T} \right] + \\
 &\quad \frac{1}{2} \mathbb{Q} \left[W(T)/\sqrt{T} + a\sqrt{T} \leq -\gamma' - a\sqrt{T} \right] \\
 &= \frac{1}{2} \Phi(-\gamma' - a\sqrt{T}) + \frac{1}{2} \Phi(-\gamma' + a\sqrt{T})
 \end{aligned}$$

Since $\lambda(t, \omega) = a \tanh(aW(t))$ and we derived that $|a\sqrt{T}| \leq 2.80$, it follows that $|a| \leq 2.80/\sqrt{T}$. Because $-a \leq a \tanh(aW(t)) \leq a$, the necessary condition can be compared with the sufficient condition $|\lambda\sqrt{T}| \leq 2.48$.

4.C Divergences

4.C.1 Kullback-Leibler

The Kullback-Leibler divergence (also known as *entropy*) is defined as

$$D(\mathbb{P} \parallel \mathbb{Q}) = \mathbb{E}^{\mathbb{P}}[-\ln R(T) - 1 + R(T)] = \mathbb{E}^{\mathbb{Q}} \left[-\frac{\ln R(T) + 1 - R(T)}{R(T)} \right] \quad (4.33)$$

We want to investigate the maximum entropy that can be achieved within the class of Radon-Nikodym derivatives with $|\lambda(t, \omega)| \leq k$. We can formulate this as a stochastic optimisation problem of the form

$$\begin{aligned}
 \max_{|\lambda(t, \omega)| \leq k} \mathbb{E}[-\ln R(T) - 1 + R(T)] & \quad (4.34) \\
 \text{s.t. } dR = \lambda(t, \omega) R dW &
 \end{aligned}$$

This optimisation problem admits the following HJB representation. If we set $V(t, r) := \mathbb{E}[-\ln R(T) - 1 + R(T) \mid R(t) = r]$, then the optimised value function $V(t, r)$ for $t \leq T$ is given by the HJB-equation

$$V_t + \max_{|\lambda(t, r)| \leq k} \frac{1}{2} \lambda(t, r)^2 r^2 V_{rr} = 0 \quad (4.35)$$

The terminal condition $V(T, r) = -\ln r - 1 + r$ is a convex payoff in r . Hence we propose $|\lambda(t, \omega)| = k$ to solve the HJB. The implied value function for the candidate solution is

$$V(t, r) = -\ln r + \frac{1}{2} k^2 (T - t) - 1 + r \quad (4.36)$$

This function solves the HJB-equation for the boundary condition and maximises the objective in (4.35) as V is convex in r . Similarly as in the maximum power calculation in Section 4.3.2, the value function $V(t, r)$ is uniformly integrable and thus the optimal control is $|\lambda(t, \omega)| = k$. Hence, the log-normal Radon-Nikodym derivative with $|\lambda(t, \omega)| \equiv k$ for all $0 \leq t \leq T$ achieves the maximal entropy of $V(0, 1) = \frac{1}{2} k^2 T$. Thus we can characterise the class of time-consistent indistinguishable models with $|\lambda(t, \omega)| \leq 2.48/\sqrt{T}$ by the implied maximum attainable entropy of $\frac{1}{2} (2.48)^2 = 3.08$.

4.C.2 Burg Entropy

Burg entropy or also called minimum discrimination information is defined as

$$D(\mathbb{Q} \parallel \mathbb{P}) = \mathbb{E}^{\mathbb{Q}}[\ln R(T) + \frac{1}{R(T)} - 1] = \mathbb{E}^{\mathbb{P}}[R(T) \ln R(T) + 1 - R(T)] \quad (4.37)$$

We can formulate this as a stochastic optimisation problem of the form

$$\begin{aligned} \max_{|\lambda(t, \omega)| \leq k} \mathbb{E}[R(T) \ln R(T) + 1 - R(T)] \\ \text{s.t. } dR = \lambda(t, \omega) R dW \end{aligned} \quad (4.38)$$

This optimisation problem admits the following HJB representation. If we set $V(t, r) := \mathbb{E}[R(T) \ln R(T) + 1 - R(T) \mid R(t) = r]$, then the optimised value function $V(t, r)$ for $t \leq T$ is given by the HJB-equation

$$V_t + \max_{|\lambda(t, R)| \leq k} \frac{1}{2} \lambda(t, r)^2 r^2 V_{rr} = 0 \quad (4.39)$$

As the terminal condition $V(T, r) = r \ln r + 1 - r$ is a strictly convex function in r , we can follow the same procedure as the proof of the maximum power and entropy calculation. The log-normal Radon-Nikodym derivative with $|\lambda(t, \omega)| \equiv k$ for all $0 \leq t \leq T$ achieves the maximal Burg entropy of $V(0, 1) = \frac{1}{2}k^2T$. Since for $q \in \mathbb{R}$

$$\begin{aligned} \mathbb{E}[R(T)^q \mid \mathcal{F}_t] &= \mathbb{E}\left[r^q e^{-\frac{1}{2}qk^2(T-t)+qk(W^{\mathbb{P}}(T)-W^{\mathbb{P}}(t))} \mid \mathcal{F}_t\right] \\ &= r^q e^{\frac{1}{2}k^2(T-t)(q^2-q)} \end{aligned} \quad (4.40)$$

The derivative with respect to q on both sides leads to

$$\begin{aligned} \frac{\partial \mathbb{E}[R(T)^q \mid \mathcal{F}_t]}{\partial q} &= \mathbb{E}[R(T)^q \ln R(T) \mid \mathcal{F}_t] \\ \frac{\partial r^q e^{\frac{1}{2}k^2(T-t)(q^2-q)}}{\partial q} &= r^q \ln r e^{\frac{1}{2}k^2(T-t)(q^2-q)} + \\ &\quad r^q e^{\frac{1}{2}k^2(T-t)(q^2-q)} \frac{1}{2}k^2(T-t)(2q-1) \end{aligned} \quad (4.41)$$

For $q = 1$ we get $V(t, r) = r \left(\ln r + \frac{1}{2}k^2(T-t)\right) + 1 - r$. Hence, we can characterise the class of time-consistent indistinguishable models with $|\lambda(t, \omega)| \leq 2.48/\sqrt{T}$ by the implied maximum attainable Burg entropy of $\frac{1}{2}(2.48)^2 = 3.08$.

4.C.3 J-Divergence

Jeffreys'(1946) J-divergence is $D(\mathbb{P} \parallel \mathbb{Q}) + D(\mathbb{Q} \parallel \mathbb{P})$ which equals $\mathbb{E}^{\mathbb{P}}[-\ln R(T)] + \mathbb{E}^{\mathbb{P}}[R(T) \ln R(T)]$. We can formulate this as a stochastic optimisation problem of the form

$$\begin{aligned} \max_{|\lambda(t, \omega)| \leq k} &\mathbb{E}[-\ln R(T)] + \mathbb{E}[R(T) \ln R(T)] \\ \text{s.t. } &dR = \lambda(t, \omega)RdW \end{aligned} \quad (4.42)$$

Again the associated value function is convex in r at terminal time T , $V(T, r) = (r-1) \ln r$. The log-normal candidate leads to $V(t, r) = (r-1) \ln r + \frac{1}{2}k^2(T-t)(r+1)$, and hence the maximum J-divergence is obtained for $V(0, 1) = k^2T = (2.48)^2 = 6.15$.

4.C.4 χ^2 -Distance

Actually for all $\frac{\partial^2 \phi(r)}{\partial r^2} \geq 0$ the associated value function is convex for $r > 0$ and hence $|\lambda(t, \omega)| = 2.48/\sqrt{T}$ leads to the maximum attainable distance. The \mathbb{P} -expectation of the χ^2 -distance is

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[R(T) \phi \left(\frac{1}{R(T)} \right) \right] &= \mathbb{E}^{\mathbb{P}} [R(T)^2 - 2R(T) + 1] \\ &= \mathbb{E}^{\mathbb{P}} [R(T)^2] - 2\mathbb{E}^{\mathbb{P}} [R(T)] + 1 \end{aligned} \quad (4.43)$$

For the associated optimisation problem the HJB representation is similar as the ones before. If we set $V(t, r) := \mathbb{E} [R(T)^2 - 2R(T) + 1 \mid R(t) = r]$, then the terminal condition $V(T, r) = r^2 - 2r + 1$ is convex. For $|\lambda(t, \omega)| = k$ the value function is

$$\begin{aligned} V(t, r) &= r^2 e^{k^2(T-t)} - 2r + 1 \\ V(0, 1) &= e^{2.48^2} - 1 = 467.90 \end{aligned} \quad (4.44)$$

4.C.5 Modified χ^2 -Distance

The objective function of the modified χ^2 -distance under \mathbb{P} is

$$\mathbb{E}^{\mathbb{P}} \left[R(T) \phi \left(\frac{1}{R(T)} \right) \right] = \mathbb{E}^{\mathbb{P}} \left[\frac{1}{R(T)} - 2 + R(T) \right] \quad (4.45)$$

The value function is $V(t, r) := \mathbb{E}^{\mathbb{P}} \left[\frac{1}{R(T)} - 2 + R(T) \mid R(t) = r \right]$ which has a convex terminal condition. For the log-normal Radon-Nikodym derivative the modified χ^2 -distance's value function is

$$V(t, r) = \frac{1}{r} e^{k^2(T-t)} - 2 + r \quad (4.46)$$

Henceforth, $V(0, 1) = e^{k^2 T} - 1 = 467.90$ is the bound that includes the indistinguishable models for a 5% probability that a Type I error occurs and have a power less than 80%. In order to derive (4.46) we applied Itô's Lemma, $d \left(\frac{1}{R(t)} \right) = \frac{1}{R(t)} (k^2 dt - kW^{\mathbb{P}}(t))$.

4.C.6 Hellinger Distance

The Hellinger distance equals

$$\mathbb{E}^{\mathbb{P}} \left[R(T) \phi \left(\frac{1}{R(T)} \right) \right] = \mathbb{E}^{\mathbb{P}} \left[1 - 2\sqrt{R(T)} + R(T) \right] \quad (4.47)$$

The associated value function is $V(t, r) := \mathbb{E}^{\mathbb{P}} \left[1 - 2\sqrt{R(T)} + R(T) \mid R(t) = r \right]$. Since

$$\begin{aligned} \sqrt{R(T)} &= \sqrt{R(t)} e^{-\frac{1}{4}k^2(T-t) + \frac{1}{2}k(W^{\mathbb{P}}(T) - W^{\mathbb{P}}(t))} \\ d\sqrt{R(T)} &= -\frac{1}{8}k^2\sqrt{R(t)}dt + \frac{1}{2}k\sqrt{R(t)}dW^{\mathbb{P}}(t) \end{aligned} \quad (4.48)$$

The optimal control solves for $|\lambda(t, \omega)| = k$

$$\begin{aligned} V(t, r) &= 1 - 2\sqrt{r}e^{-\frac{1}{8}k^2T} + r \\ V(0, 1) &= 1.07 \end{aligned} \quad (4.49)$$

4.C.7 Variation Distance

The variation distance can be decomposed into a call and put option and is defined as follows

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[R(T) \phi \left(\frac{1}{R(T)} \right) \right] &= \mathbb{E}^{\mathbb{P}} \left[\left| \frac{1}{R(T)} - 1 \right| R(T) \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[\max \left(\frac{1}{R(T)} - 1, 0 \right) R(T) \right] + \\ &\quad \mathbb{E}^{\mathbb{P}} \left[\max \left(1 - \frac{1}{R(T)}, 0 \right) R(T) \right] \\ &= \mathbb{E}^{\mathbb{P}} [\max(1 - R(T), 0)] + \mathbb{E}^{\mathbb{P}} [\max(R(T) - 1, 0)] \end{aligned} \quad (4.50)$$

The objective is convex by convexity of the max-operator. The associated value function is $V(t, r) := \mathbb{E}^{\mathbb{P}} \left[\left| \frac{1}{R(T)} - 1 \right| R(T) \mid R(t) = r \right]$. Therefore we propose $dR(t) = kR(t)dW^{\mathbb{P}}(t)$ for the value function

$$\begin{aligned} V(t, r) &= N(d_1)r - N(d_2) + N(-d_2) - N(-d_1)r \\ d_1 &= \frac{1}{k\sqrt{T-t}} \left(\ln r + \frac{1}{2}k^2(T-t) \right) \\ d_2 &= d_1 - k\sqrt{T-t} \end{aligned} \quad (4.51)$$

The second derivative is positive thus the optimal value is obtained for

$$\begin{aligned}
 V(0, 1) &= N(d_1) - N(d_2) + N(-d_2) - N(-d_1) & (4.52) \\
 d_1 &= \frac{1}{2}k\sqrt{T} \\
 d_2 &= -d_1 \\
 N(-d_1) &= 1 - N(d_1) \\
 V(0, 1) &= 4N(d_1) - 2
 \end{aligned}$$

For $d_1 = \frac{1}{2}k\sqrt{T} = 1.24$, the bound on the variation distance is $V(0, 1) = 1.57$.

4.C.8 χ -Divergence of Order $\theta > 1$

For general θ strictly above 1 the χ -divergence can only be solved numerically, though for integers analytically. The divergence functions is defined by

$$\mathbb{E}^{\mathbb{P}} \left[R(T)\phi \left(\frac{1}{R(T)} \right) \right] = \mathbb{E}^{\mathbb{P}} \left[\left| \frac{1}{R(T)} - 1 \right|^{\theta} R(T) \right] \quad (4.53)$$

For different values of $\theta = \{1.5, 2, 2.5, 3\}$ and $k^2T = 2.48^2$, we integrate

$$\mathbb{E}^{\mathbb{P}} \left[\left| \frac{1}{R(T)} - 1 \right|^{\theta} R(T) \right] = \int_{-\infty}^{\infty} \left| e^{\frac{1}{2}k^2T - k\sqrt{T}z} - 1 \right|^{\theta} e^{-\frac{1}{2}T + k\sqrt{T}z} n(z) dz$$

If $\theta = 2$ this coincides with modified χ^2 -divergence.

4.C.9 Cressie-Read

The Cressie-Read divergence is

$$\begin{aligned}
 \mathbb{E}^{\mathbb{P}} \left[R(T)\phi \left(\frac{1}{R(T)} \right) \right] &= \mathbb{E}^{\mathbb{P}} \left[\frac{1 - \theta + \theta \frac{1}{R(T)} - \left(\frac{1}{R(T)} \right)^{\theta}}{\theta(1 - \theta)} R(T) \right] & (4.54) \\
 &= \frac{1}{\theta(1 - \theta)} \mathbb{E}^{\mathbb{P}} [R(T) - \theta R(T) + \theta - R(T)^{1-\theta}]
 \end{aligned}$$

This is convex in r therefore if $R(T)$ is log-normal distributed with volatility k ,

$$V(t, r) = \frac{1}{\theta(1 - \theta)} \left(r - \theta \cdot r + \theta - r^{1-\theta} e^{-\frac{1}{2}k^2\theta(1-\theta)(T-t)} \right) \quad (4.55)$$

At time 0 this equals

$$\begin{aligned}V(0,1) &= \frac{1}{\theta(1-\theta)} \left(1 - e^{-\frac{1}{2}k^2\theta(1-\theta)T}\right) \\ &= \frac{1}{\theta(1-\theta)} \left(1 - e^{-3.08\theta(1-\theta)}\right) \\ &= \frac{1}{\theta(1-\theta)} \left(1 - 0.05e^{\theta(1-\theta)}\right)\end{aligned}\tag{4.56}$$

Chapter 5

Conclusion

*“Not everything that can be counted counts,
and not everything that counts can be counted.”*

-Albert Einstein (1879-1955)

This chapter provides an overall conclusion of this thesis, more detailed conclusions can be found at the end of each individual chapter. Recognising that models serve as approximations of the reality only, supports the acknowledgement of model uncertainty. This in combination with robustness, has been widely used in physics and engineering. In this research however, both subjects are used for and make use of topics in econometrics, finance, actuarial science, game theory and operations research. This thesis has argued that we are in an incomplete market. Incompleteness itself indicates already that one is uncertain since multiple equivalent martingale measures exist that all lead to different possible values. Models including the incompleteness assumptions imitate real world patterns better, since not every liability is (liquidly) traded. Moreover, the acknowledgement of uncertainty about a single model specification diminishes the gap between practice and modelling as well. A literature review on pricing in incomplete markets and on model uncertainty is given in Chapter 1.

In Chapter 2 parameter uncertainty is implemented in a Bayesian framework, whereas in Chapter 3 and 4 the model uncertainty is tackled from the frequentist viewpoint. Specifically, in Chapter 2, interest rates are modelled by an affine term structure model, namely the multivariate Vasicek process. The agent is uncertain about the specific parameter estimations and incorporates this by assuming that these parameters are drawn from probability distributions. By drawing many simulations from the conditional posterior distributions we are able to indicate the uncertainty of the specific parameters and the term structure as a whole. Furthermore, models often used in the industry seem to undermine the convexity effect. To find the hedge position that minimises the gap between the assets and liabilities we solve this optimisation problem in an incomplete market setting in the next chapter.

In Chapter 3, an optimal control problem is explored where an agent is uncertain about a model that is represented by a stochastic differential equation. Since he distrusts his model he would like to make a robust decision such that in bad times the loss is limited. The following game accomplished this strategy; the agent wants to maximise his surplus against a counter player who wants to minimise the surplus. The counter player has a choice among a set of models to pick the worst-case scenario. This set has to possess plausible alternatives. An intuitive choice is a region around the baseline model, graphically interpretable by an ellipsoid. In Chapter 3 this optimisation problem is solved for the multivariate case. That is, mother nature picks the worst-case drift distortions whereas the

optimal hedging portfolio is a combination of a hedging position related to both hedgeable and unhedgeable risk and a speculative part. The specific choice of the size of the set of plausible sets is contributed to in Chapter 4.

The fourth chapter focusses on the set of indistinguishable models. For robust stochastic optimisation problems a set of alternatives is needed as input. However, how this set should look like and how large it should be is often not questioned but assumed to be given. Both deterministic and stochastic deviations from the drift are considered. Allowing for stochastic alternatives yields a wide class of models beyond solely changing the drift term, as these imply fundamentally different distributions. The deviations that cannot be distinguished from the null hypothesis, *ex ante*, are included in the set. The Type I and II error are imposed on the likelihood ratio test. The alternatives are identified by a power that is too low to distinguish the alternative from the null hypothesis. For stochastic deviations we have proven that the power is maximised if the volatility is equal to the constant bound that was imposed to refrain time-inconsistency. Hence, for both deterministic and stochastic deviations the Radon-Nikodym derivative is log-normal distributed. Therefore for a given Type I and II error, i.e. $\alpha = 5\%$ and $1 - \beta = 80\%$, the deviations surrounding the main model smaller than $2.48/\sqrt{T}$ are indistinguishable if one would have T years of extra data. If T increases, thus if more information is available, the agent is less uncertain and the set of alternatives becomes smaller.

To conclude, models are simplifications of real world processes and situations. Therefore, there is the risk of a mismatch between the model and the implied strategy. Yet, this risk has not often been quantified and consequently it has not been taken into account. In this thesis several approaches are considered that allow for the concern that models might be wrong. The impact on long-term interest rates, on the hedging strategy and on the price is derived. Moreover, the set of plausible alternative models is identified. Hence, the effect of *Model Uncertainty on Robustness, Estimation and Stochastic Optimisation* is investigated.

Chapter 6

Valorisation

*“The policies with respect to the melting of the icebergs
face the same problem of extrapolation on the extreme long end
as those of pension funds.”*

-Anne Balter (Assertions accompanying the thesis)

In this addendum I outline the knowledge valorisation of this dissertation, by discussing the (social and economic) relevance of model uncertainty and the respective results, potential target groups, translated services, processes and activities, the innovativeness of the research and the implementation of the valorisation plans. Firstly, I discuss the relevance of this thesis to actuarial science, specifically to pension funds and insurance companies. Since the concepts derived are quite general, the usefulness extends to a wide variety of applications. After giving some examples in which misspecification plays a role, I highlight the general applicability once more by means of the section about climatology that serves both literally and figuratively as the tip of the iceberg.

6.1 Pension Funds

Since pension funds have liabilities on very long horizons, they need instruments with corresponding maturities in order to hedge the risk or at least to quantify the present value of all outstanding liabilities. The present value of the liabilities, together with the present value of the assets, is an indicator of the pension funds' health. The problem is that financial instruments are liquid up to maturities of 20 years whereafter the liquidity declines rapidly, while pensioners can live as long as a century. In the past, actuaries in The Netherlands used a fixed rate of 4% to discount their liabilities. However, soon one realised that this might not be market-consistent. Since then, discount curves have been used, but prices are not quoted at the long end. The ultimate forward rate (UFR) methodology, also proposed for insurance companies in Europe, states that in the long run, at a maturity of 60 years, the rates should converge to a fixed UFR. These fixed values seem to be artificial and should be investigated. The current debate is all about how to translate the obligations that have to be paid in the future into its value today. With other words "how much money is needed today to meet the future obligations?". In Chapter 2 the UFR method is implemented. Another standard in the industry, the Nelson-Siegel method, is also implemented. Both methods are compared with the Vasicek model.

Chapter 2 is about extrapolating the term structure of interest rates for maturities with declining liquidity and especially about measuring the uncertainty. We quantified the effect of parameter uncertainty in a one-factor affine term structure model in a Bayesian way. We applied this model to European and American data and found that the convexity term is an important factor whose role is often

neglected in practice. Our conclusion is that convergence happens very slowly, implying that the convexity effect does not vanish within a century. The other finding is that taking uncertainty into account results in an economically plausible, though rather wide range.

The societal impact of a high curve is a low present value of the liabilities. Therefore less is needed to save for the future and more can be spend today. This results in either no cutting of the pension payments or an increase in the indexation. Hence, this is beneficial for the current retirees and disadvantageous for the younger generations. Saving less might have a negative effect for the latter group if the model deviations are different than expected. Our simulations show the uncertainty explicitly. Therefore we do not propose a policy that advises a point estimate, but instead we quantify the uncertainty. The policy could use this information to reserve a buffer. Trivially, the effect is reversed if the expected rates are low. Moreover, irrespective of the funding ratio, which is the present value of the assets divided by the present value of the liabilities, different divisions among age groups harm specific cohorts at the expense of other cohorts.

So far, this has illustrated the social and economic relevance of Chapter 2 for employees and retirees. Another “target group” are the regulators and politicians. This research offers insight into the ongoing debate about the pension system. The paper on which this chapter is based has been used as input for the committee UFR. This committee¹ informed the minister on the UFR policy. The Dutch Central Bank (De Nederlandsche Bank) implemented the proposals from July 2015 onwards as mandatory discount method for pension funds. Moreover the paper *What does a term structure model imply about very long-term discount rates?* has been presented by the DNB to actuaries.

6.2 Insurance Companies

Chapter 3 specifies the effect of model uncertainty on pricing and hedging in incomplete markets. The set-up of the optimisation problem resembles a game in which an agent wants to maximise his surplus by choosing a hedging strategy whereas the worst-case scenario will be picked by a so called mother nature who minimises the surplus. This ensures a robust strategy which is the one that is least sensitive to perturbations of the model. Intuitively, the agent proposes a region of alternative models around his point estimate. We prove that the price of the liabil-

¹Prof. dr. A.A.J. Pelsser, my promoter, is a member of the committee UFR.

ity that may depend on both hedgeable and unhedgeable risk, is uniquely solvable. Moreover, we explain the economic interpretation of the hedging strategy which is identical to delta-hedging if the market is complete, i.e. risk-neutral pricing. In case of purely unhedgeable risk, the drift is adjusted in the prudent direction known as actuarial pricing. And in case of a mixed multivariate composition of tradeable and untradeable assets, the hedging strategy consists of a purely delta-hedging part, an extra delta-hedging part that captures the correlation between the hedgeable and unhedgeable risk and a speculative part.

A practitioner's methodology to price in incomplete markets is the industry standardised Cost-of-Capital method. Insurance companies use this quantifies the market value of the replicating portfolio plus a mark-up for the unhedgeable risk, which relies mostly on the subjective quantification of risk. The CoC method leads to a pricing operator that has similar characteristics as our result. The indifference pricing operator from Theorem 3.1 can be interpreted as a best estimate, which is the conditional expectation, plus a constant times the standard deviation of the unhedgeable component. The European Insurance and Occupational Pensions Authority (EIOPA) proposes the use of the CoC method. This is incorporated in the guidelines for Solvency II that is scheduled to come in effect on January 1, 2016 as mandatory pricing rule for insurance companies in Europe.

Consequently actuarial practices can be justified by the theorem developed in this chapter. Buffers imposed by regulators affect the financial behaviour of pension funds and insurance companies. But also any profit, or non-profit organisation can utilise this to support certain decisions concerning hedging and concerning the level of prudence or risk it wants to take. Moreover, the derivations in Chapter 3 are based on profit maximisation, resembling the core goal of many companies. Additionally, pricing in incomplete markets implies that the value for a non-traded item is deduced. Henceforth, not only insurance products benefit from this insight but one may think of *any* item to be priced. For instance, contracts that pay out during natural disasters, extreme low or high temperatures, the number of survivors, etcetera. Note that catastrophe bonds and mortality swaps are traded, which would imply the possibility to replicate cash flows. However the liquidity of these instruments is rather limited.

6.3 Misspecification

Chapter 4 is a theoretical chapter that serves the valorisation indirectly. It continues on the topic akin to model uncertainty. The idea that models might be misspecified is an improvement towards approaching real world examples by mathematical models. Chapter 4 deals with a general method for determining all alternative models that cannot be distinguished from each other. This set is widely applicable to various problems in different areas that allow for model uncertainty. Applying this to Chapter 3 leads to the quantification of the uncertainty bounds on the ellipsoid. Other literature that deals with uncertainty, but remains silent about a realistic choice of plausible models, can make use of these cutoff points.

Robust optimisation is relevant for both social and economic problems since the quantification of uncertainty can be implemented in many social questions that have been solved for an unknown ambiguity value. An example is the consumption and portfolio problem, where an agent wants to maximise his final wealth or the consumption during a specific time span and has to choose how much to consume, how much to save on the bank account and how much to invest in the risky assets. If an agent acknowledges that the underlying model might be misspecified, he would like to evaluate the optimal decision rule such that the strategy is least sensitive to perturbations of the model. As such a direct link with the constraint on the Kullback-Leibler divergence of the investment problem is established. Moreover, we can link this with a penalty on the entropy as well. Chapter 3 and 4 can also be used to measure uncertainty about interest rates. Affine deviations from the baseline model appear to be among the plausible stochastic alternatives and yield the transition to a mean-reverting model. Ben-Tal et al. (2009) consider a robust solution with respect to how much raw materials a company should buy to produce drugs, where the exact amount of raw material needed entails uncertainty. In another example they optimise the location of an antenna that transmits isotropic harmonic oscillator emitting spherical monochromatic electromagnetic waves. The objective is to send the electromagnetic field that is invoked by the oscillator, as close as possible to the target. Therefore uncertainty comes from both the positioning of the antenna, which is influenced by temperature, wind, etcetera, and the actuation errors, which emerge due to fact that invoking the oscillator cannot be implemented exactly.

6.4 Climatology

Taking one step back, the theory of forecasting the uncertainty based on a panel data set for an autoregressive process allows itself for more than interest rate modelling. The mean-reverting process of Chapter 2, a process that has a level to which it tends to move, can be applied to other fields than finance. Research in climatology can be conducted by use of Chapter 2. The model can be applied to a panel data set, which refers to two-dimensional data where the observations are measured for several objects or individuals through time, such as historic temperature or emission data. Forecasting these variables includes uncertainty which is recognised by the approach of Chapter 2. The widespread interest in these forecasts stems from concern about rising temperatures.

The link between Chapter 2 and the climate issue can also be drawn from an environmental economics viewpoint. How much climate change costs and at whose expense it is has received considerable attention in the field of environmental economics. Since long-term interest rates have an effect on the present value of the costs, climate change policies are influenced by these. Gollier (2013) shows in his book *Pricing the Planet's Future* the impact of the discount rate on extreme long horizons. With a fixed interest rate of 4% the value one needs today to have €100 in 200 years is €0.03 while for a rate of 2% one would need €1.83, almost 55 times as much. Hence, any cash flow that needs to be valued to the present, such as in a cost-benefit analysis, can make use of Chapter 2 when the horizon stretches until the extreme long end.

Bibliography

“Either write something worth reading or do something worth writing.”
-Benjamin Franklin (1706-1790)

- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9(3), 203–228.
- Artzner, P., F. Delbaen, J.-M. Eber, D. Heath, and H. Ku (2007). Coherent multiperiod risk adjusted values and Bellmans principle. *Annals of Operations Research* 152(1), 5–22.
- Avellaneda, M., A. Levy, and A. Parás (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance* 2(2), 73–88.
- Balter, A. G. and A. A. J. Pelsser (2015a). Pricing and hedging in incomplete markets with model ambiguity. *Maastricht University Working Papers*.
- Balter, A. G. and A. A. J. Pelsser (2015b). Sets of indistinguishable models for robust optimisation. *Maastricht University Working Papers*.
- Balter, A. G., A. A. J. Pelsser, and P. C. Schotman (2015). What does a term structure model imply about very long-term discount rates? *Maastricht University Working Papers*.
- Barrieu, P. and N. El Karoui (2005). Inf-convolution of risk measures and optimal risk transfer. *Finance and Stochastics* 9(2), 269–298.
- Barrieu, P. M. and N. El Karoui (2009). Pricing, hedging and optimally designing derivatives via minimization of risk measures. In *Indifference Pricing: Theory and Applications*, pp. 77–146. Princeton University Press.
- Basak, S. and G. Chabakauri (2012). Dynamic hedging in incomplete markets: a simple solution. *Review of Financial Studies* 25(6), 1845–1896.
- Becherer, D. (2009). From bounds on optimal growth towards a theory of good-deal hedging. *Advanced Financial Modelling* 8, 27–51.
- Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* 59(2), 341–357.
- Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust optimization*. Princeton University Press.

- Björk, T. and I. Slinko (2006). Towards a general theory of good-deal bounds. *Review of Finance* 10(2), 221–260.
- Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy* 81(3), 637–654.
- Boyle, P., S. Feng, W. Tian, and T. Wang (2008). Robust stochastic discount factors. *Review of Financial Studies* 21(3), 1077–1122.
- Breuer, T. and I. Csiszár (2013a). Measuring distribution model risk. *Mathematical Finance*, DOI: 10.1111/mafi.12050.
- Breuer, T. and I. Csiszár (2013b). Systematic stress tests with entropic plausibility constraints. *Journal of Banking & Finance* 37(5), 1552–1559.
- Cao, H. H., T. Wang, and H. H. Zhang (2005). Model uncertainty, limited market participation, and asset prices. *Review of Financial Studies* 18(4), 1219–1251.
- Černý, A. and S. Hodges (2002). The theory of good-deal pricing in financial markets. In *Mathematical Finance Bachelier Congress 2000*, pp. 175–202. Springer.
- Chacko, G. and L. M. Viceira (2005). Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. *Review of Financial Studies* 18(4), 1369–1402.
- Cheridito, P., F. Delbaen, and M. Kupper (2005). Coherent and convex monetary risk measures for unbounded Cadlag processes. *Finance and Stochastics* 9(3), 369–387.
- Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics* 164(1), 4–20.
- Cochrane, J. and J. Saa-Requejo (2000). Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy* 108(1), 79–119.
- Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance* 16(3), 519–547.
- Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Science* 8, 85–108.

-
- Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *The Journal of Finance* 55(5), 1943–1978.
- De Jong, F. (2000). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics* 18(3), 300–314.
- Delbaen, F. (2006). The structure of m -stable sets and in particular of the set of risk neutral measures. In *In Memoriam Paul-André Meyer*, pp. 215–258. Springer.
- Delbaen, F., P. Grandits, T. Rheinländer, D. Samperi, M. Schweizer, and C. Stricker (2002). Exponential hedging and entropic penalties. *Mathematical Finance* 12(2), 99–123.
- Delbaen, F. and W. Schachermayer (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen* 300(1), 463–520.
- Diebold, F. and G. Rudebusch (2013). *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. Princeton University Press.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance* 57(1), 405–443.
- Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance* 6(4), 379–406.
- El Karoui, N., S. Hamadène, and A. Matoussi (2009). Backward stochastic differential equations and applications. In *Indifference Pricing: Theory and Applications*, pp. 267–320. Princeton University Press.
- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics* 75(4), 643–669.
- Epstein, L. G. and M. Schneider (2003). Recursive multiple-priors. *Journal of Economic Theory* 113(1), 1–31.
- Filipović, D. and N. Vogelpoth (2008). A note on the Swiss solvency test risk measure. *Insurance: Mathematics and Economics* 42(3), 897–902.
- Floroiu, O. and A. A. J. Pelsser (2013). Closed-form solutions for options in incomplete markets. *Maastricht University Working Papers*.

- Föllmer, H. and P. Leukert (1999). Quantile hedging. *Finance and Stochastics* 3(3), 251–273.
- Föllmer, H. and P. Leukert (2000). Efficient hedging: cost versus shortfall risk. *Finance and Stochastics* 4(2), 117–146.
- Föllmer, H. and A. Schied (2002). Convex measures of risk and trading constraints. *Finance and Stochastics* 6(4), 429–447.
- Garlappi, L., R. Uppal, and T. Wang (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies* 20(1), 41–81.
- Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 571–578.
- Ghalehjooghi, A. S. and A. A. J. Pelsser (2014). Time-consistent actuarial valuations. *Maastricht University Working Papers*.
- Gianin, E. R. (2006). Risk measures via g-expectations. *Insurance: Mathematics and Economics* 39(1), 19–34.
- Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2), 141–153.
- Gobet, E. and C. Labart (2007). Error expansion for the discretization of backward stochastic differential equations. *Stochastic Processes and their Applications* 117(7), 803–829.
- Gobet, E. and C. Labart (2010). Solving BSDE with adaptive control variate. *SIAM Journal on Numerical Analysis* 48(1), 257–277.
- Gobet, E., J.-P. Lemor, and X. Warin (2005). A regression-based Monte Carlo method to solve backward stochastic differential equations. *The Annals of Applied Probability* 15(3), 2172–2202.
- Goldfarb, D. and G. Iyengar (2003). Robust portfolio selection problems. *Mathematics of Operations Research* 28(1), 1–38.
- Gollier, C. (2013). *Pricing the planet’s future: the economics of discounting in an uncertain world*. Princeton University Press.

- Guidolin, M. and F. Rinaldi (2013). Ambiguity in asset pricing and portfolio choice: A review of the literature. *Theory and Decision* 74(2), 183–217.
- Gundel, A. and S. Weber (2007). Robust utility maximization with limited downside risk in incomplete markets. *Stochastic Processes and their Applications* 117(11), 1663–1688.
- Hansen, L. P. and T. J. Sargent (2001). Robust control and model uncertainty. *American Economic Review* 91(2), 60–66.
- Hansen, L. P. and T. J. Sargent (2008). *Robustness*. Princeton University Press.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.
- Harrison, J. M. and D. M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20(3), 381–408.
- Henderson, V. (2002). Valuation of claims on nontraded assets using utility maximization. *Mathematical Finance* 12(4), 351–373.
- Henderson, V. (2005). Explicit solutions to an optimal portfolio choice problem with stochastic income. *Journal of Economic Dynamics and Control* 29(7), 1237–1266.
- Henderson, V. and D. Hobson (2009). Utility indifference pricing—an overview. In *Indifference Pricing: Theory and Applications*, pp. 44–74. Princeton University Press.
- Hodges, S. D. and A. Neuberger (1989). Optimal replication of contingent claims under transaction costs. *Review of Futures Markets* 8(2), 222–239.
- Hu, Y., P. Imkeller, and M. Müller (2005). Utility maximization in incomplete markets. *The Annals of Applied Probability* 15(3), 1691–1712.
- Huber, P. J. (1981). *Robust statistics*. New York: John Wiley and Sons.
- Jaschke, S. and U. Küchler (2001). Coherent risk measures and good-deal bounds. *Finance and Stochastics* 5(2), 181–200.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. In *Mathematical, Physical and Engineering Sciences*, Volume 186, pp. 453–461. The Royal Society.

- Keller, P. and T. Luder (2004). White paper of the Swiss solvency test. *Swiss Federal Office of Private Insurance, Switzerland*.
- Klöppel, S. and M. Schweizer (2007). Dynamic utility-based good deal bounds. *Statistics & Decisions* 25(4), 285–309.
- Kreps, D. M. (1981). Arbitrage and equilibrium in economies with infinitely many commodities. *Journal of Mathematical Economics* 8(1), 15–35.
- Laeven, R. J. and M. Stajda (2014). Robust portfolio choice and indifference valuation. *Mathematics of Operations Research* 39(4), 1109–1141.
- Maenhout, P. J. (2004). Robust portfolio rules and asset pricing. *Review of Financial Studies* 17(4), 951–983.
- Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 247–257.
- Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory* 3(4), 373–413.
- Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society* 41(5), 867–887.
- Miao, J. and N. Wang (2007). Investment, consumption, and hedging under incomplete markets. *Journal of Financial Economics* 86(3), 608–642.
- Musiela, M. and T. Zariphopoulou (2004). An example of indifference prices under exponential preferences. *Finance and Stochastics* 8(2), 229–239.
- Nelson, C. and A. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business*, 473–489.
- Neyman, J. and E. S. Pearson (1933). The testing of statistical hypotheses in relation to probabilities a priori. *Mathematical Proceedings of the Cambridge Philosophical Society* 29(04), 492–510.
- Øksendal, B. (2003). *Stochastic differential equations*. Springer.
- Palm, F. C. and A. Zellner (1992). To combine or not to combine? Issues of combining forecasts. *Journal of Forecasting* 11(8), 687–701.

- Pardoux, E. and S. Peng (1990). Adapted solution of a backward stochastic differential equation. *Systems & Control Letters* 14(1), 55–61.
- Pelsser, A. A. J. and M. Stadje (2014). Time-consistent and market-consistent evaluations. *Mathematical Finance* 24(1), 25–65.
- Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of Risk* 2, 21–42.
- Rockafellar, R. T. and S. Uryasev (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* 26(7), 1443–1471.
- Rogers, L. (2003). Duality in constrained optimal investment and consumption problems: a synthesis. In *Paris-Princeton Lectures on Mathematical Finance 2002*, pp. 95–131. Springer.
- Schied, A. and H. Föllmer (2011). *Stochastic Finance; An Introduction in Discrete Time*. Walter De Gruyter.
- Schweizer, M. (1992). Mean-variance hedging for general claims. *The Annals of Applied Probability* 2(1), 171–179.
- Shen, S., A. A. J. Pelsser, and P. C. Schotman (2014). Robust hedging in incomplete markets. *Maastricht University Working Papers*.
- Smith, A. and T. Wilson (2001). *Fitting Yield curves with long Term Constraints*. Research Notes, Bacon and Woodrow.
- Thomas, M. and E. Maré (2007). Long term forecasting and hedging of the South African yield curve. In *2007 ASSA Convention, The Actuarial Society of South Africa*.
- Vanden, J. M. (2006). Exact superreplication strategies for a class of derivative assets. *Applied Mathematical Finance* 13(01), 61–87.
- Wang, Z. (2005). A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies* 18(2), 673–705.
- Young, V. R. and T. Zariphopoulou (2002). Pricing dynamic insurance risks using the principle of equivalent utility. *Scandinavian Actuarial Journal* 2002(4), 246–279.
- Zariphopoulou, T. (2001). A solution approach to valuation with unhedgeable risks. *Finance and Stochastics* 5(1), 61–82.

Nederlandse Samenvatting

*“If you talk to a man in a language he understands, that goes to his head.
If you talk to him in his language, that goes to his heart.”*
-Nelson Mandela (1918-2013)

Modellen zijn slechts vereenvoudigingen van de werkelijkheid. Het erkennen van modelonzekerheid is belangrijk bij economische besluitvormingen. Het doel van dit promotieonderzoek is het karakteriseren van de set van plausibele alternatieve modellen en het bepalen van het effect van onzekerheid op beslisregels.

Het proefschrift begint in hoofdstuk 1 met een literatuuroverzicht. Verschillende mogelijke manieren van prijzen in incomplete markten worden besproken, waarna op risico en onzekerheid dieper wordt ingegaan. In een complete markt zijn alle toekomstige kasstromen repliceerbaar en kan er frictieloos gehandeld worden. Wanneer de liquiditeit afneemt, spreken we van een incomplete markt. Risico staat voor de onzekerheid dat een bepaalde gebeurtenis voorkomt, waarbij de kans op deze gebeurtenis bekend is. Bij onzekerheid is deze kans onbekend.

In hoofdstuk 2 wordt parameteronzekerheid behandeld volgens de Bayesiaanse methoden. Aangezien er geen renteproducten zijn met extreem lange looptijden, maar deze wel nodig zijn om toekomstige obligaties te verdisconteren, bestaat er onzekerheid met welke disconteringsvoet gerekend moet worden. De afwezigheid van liquide middelen definieert een incomplete markt. In dit hoofdstuk extrapoleren we de rente op basis van een lineair rentetermijnstructuur model, specifiek het Vasicek proces. Hierbij nemen we aan dat er onzekerheid is over de parameters van het model, waaraan specifieke kansverdelingen ten grondslag liggen. Vervolgens kunnen we met behulp van simulaties de onzekerheid in kaart brengen. Het blijkt dat een 100-jaars rente nog voor een groot deel beïnvloed wordt door het convexiteitseffect. Dit komt voort uit de correcte afleiding van het wiskundige model in tegenstelling tot alternatieve extrapolatie methoden die vaak gebruikt worden in de praktijk. Daarnaast is zowel de onzekerheid per parameter, als ook de onzekerheid van de totale rentetermijnstructuur geïdentificeerd. De investeringsstrategie, die het tekort tussen de aandelen en de obligaties in een incomplete markt minimaliseert, wordt afgeleid in het volgende hoofdstuk.

In hoofdstuk 3 gaan we in op een optimaal controle probleem, waarbij een agent onzeker is over het model. Hij wil een robuuste beslissing maken, zodat ook in slechte tijden zijn verlies beperkt blijft. Dit kan worden nagebootst door het volgende spel; de agent wilt zijn winst maximaliseren, terwijl de tegenspeler deze wilt minimaliseren. Zij, de zogenaamde moeder natuur, kan dit bewerkstelligen door het slechtste scenario te kiezen uit een set van plausibele alternatieve modellen. Intuïtief geeft de agent de voorkeur aan alternatieve modellen in een gebied rondom het verwachte model, grafisch voorgesteld als een ellipsoïde. In het derde hoofdstuk wordt dit voor het multidimensionale geval opgelost. Dit houdt

in dat moeder natuur de slechtste deviatie kiest. De optimale strategie van de agent, die het minst gevoelig is voor perturbaties van het onderliggende model, is een combinatie van delta-hedgen gerelateerd aan zowel de verhandelbare als niet-verhandelbare producten en een speculatief deel proportioneel aan het vertrouwen in de markt. De specifieke keuze van de grootte van de onzekerheidsset is het onderwerp van hoofdstuk 4.

Voor een robuust optimalisatie probleem is het nodig dat de set van alternatieven gegeven is. Het karakteriseren van deze set van alternatieve plausibele modellen is de focus van het meer theoretische hoofdstuk 4. Zowel deterministische als stochastische deviaties van de drift worden bekeken. Met behulp van stochastische deviaties is het mogelijk om alternatieven te beschouwen die fundamenteel anders zijn dan het basis model. Het alternatief model kan een andere distributie of variantie hebben. Het doel is om, ex ante, alle deviaties te vinden die niet te onderscheiden zijn van het basis model. De Type I fout is het onterecht verwerpen van de nul hypothese. Traditioneel wordt de kans op de Type I fout, de “alpha”, gelijk aan 5% gekozen. De Type II fout is het onterecht verwerpen van de alternatieve hypothese. Deze kans, de “beta”, wordt gewoonlijk gelijk aan 20% gekozen. De power van een test is het complement van de Type II fout, dit is de kans waarmee de alternatieve terecht wordt geaccepteerd. De power is dus de kracht waarmee twee hypothesen met genoeg kans juist geïdentificeerd kunnen worden. Wanneer deze power te laag is vinden we precies die modellen waarnaar we op zoek zijn. Door het opleggen van de kans op de Type I fout en de minimale acceptatie voor de power, kan de set van niet-te-onderscheiden modellen samengesteld worden. Concluderend, voor een gegeven Type I en II fout van respectievelijk 5% en 20%, zijn alle afwijkingen rondom het basis model die kleiner zijn dan $2.48/\sqrt{T}$ niet te onderscheiden, als er T jaar aan extra data zou zijn. Hoe meer informatie beschikbaar is, dus hoe hoger T , des te zekerder is de agent en des te kleiner is de set van alternatieven.

Hoofdstuk 2, 3 en 4 bevatten elk een gedetailleerde conclusie. Hoofdstuk 5 omvat de algehele conclusie van dit proefschrift.

Hoofdstuk 6 beschrijft de toepasbaarheid van dit proefschrift in de samenleving. De sociale impact van hoofdstuk 2 en 3 vindt directe toepassing bij zowel pensioenfondsen als verzekeringsmaatschappijen. De concepten zijn dermate algemeen waardoor deze ook voor andere sociale of economische problematiek gebruikt kunnen worden. Om het beleid met betrekking tot klimaatverandering te bepalen, dienen de kosten over een lange horizon verdisconteert te worden. Daarnaast kan

het naar-het-gemiddelde-terugkerend model van hoofdstuk 2 toegepast worden op meteorologische data. De valorisatie van hoofdstuk 4 is indirecter van aard.

Curriculum Vitae

"Don't forget your history nor your destiny."
-Bob Marley (1945-1981)

Anne Balter was born on 4 December, 1989 in Sittard, The Netherlands. She attended high school between 2002 and 2008 at St. Janscollege in Hoensbroek, where she received her Gymnasium diploma with distinction (*cum laude*). Subsequently, she studied Econometrics and Operations Research at Maastricht University. During her undergraduate studies, she spent one semester at Queensland University in Brisbane, Australia as a visiting student. In August 2012 she obtained her M.Sc. degree with a specialisation in Actuarial Sciences with distinction (*cum laude*).

After graduation, Anne joined the Department of Quantitative Economics as a Ph.D. candidate in September 2012, under the supervision of Prof. dr. Antoon Pelsser and Prof. dr. Peter Schotman. The results of her research are presented in this thesis. Anne presented her work at various international conferences, such as the World Congress of the Bachelier Finance Society in Brussels, the Actuarial and Financial Mathematics Conference (AFMath) in Brussels, the Annual Conference of the Society of Financial Econometrics (SoFiE) in Toronto, the Quantitative Methods in Finance Conference (QMF) in Sydney, the Nordic Econometric Meeting (NEM) in Helsinki, the General Advanced Mathematical Methods in Finance (AMaMeF) and Swissquote Conference in Lausanne, the International Conference on Computational and Financial Econometrics (CFE) in London, the International Finance Meeting in Paris, and Netspar meetings at different places in The Netherlands and in Venice.

Model Uncertainty

The Effect on Robustness, Estimation and Stochastic Optimisation

This thesis considers the effect of the acknowledgement of model uncertainty in an incomplete market setting on robust pricing techniques. Moreover, it characterises the set of alternative models. A literature review on pricing in incomplete markets and on model uncertainty is given in the first chapter. The second chapter deals with extrapolating the term structure of interest rates on the extreme long end, where uncertainty is quantified by the Bayesian method. The convexity effect appears to be an important element in no-arbitrage term structure models. Theorem 3.1 states the existence and uniqueness of the price that is least sensitive to perturbations of the underlying model. The quantification of uncertainty is explored in the fourth chapter. Where the ellipsoid around the estimation was used by the agent in the

third chapter, here the set is explicitly characterised. The set of alternative models are those models that cannot be distinguished from each other with enough power. Hence Theorem 4.1 states that all, both deterministic and stochastic, deviations from the null hypothesis for a given size obtain the maximal power at a certain level. The level is determined ex ante for a desired power. This thesis ends with a short conclusion, a valorisation section, a Dutch summary and an academic curriculum vitae.



Anne Balter

