

Theory and Application of Model Risk Quantification

Yu Feng

July 13, 2019

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Yu Feng declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Business at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 13/07/2019

DEDICATION

To my beloved family

ACKNOWLEDGEMENT

There are many people who contribute to this work indirectly. I am really grateful for the support that my wife and my parents provided during my study. Also thanks to my dear daughter who has brought me great pleasure in my life.

I would like to express my deep gratitude to my supervisor Prof. Erik Schlögl. He has provided me the greatest support and assistance in my exploration of the thesis topic. I am grateful for his trust and consistent teaching that allow me to complete this challenging theoretical project. Our academic talking has inspired me a lot and leads me to fruitful research outcomes.

I would like to thank one of my co-supervisors Dr. Hardy Hulley who helped me especially on the work of dynamic model risk measurement. His sense of mathematical rigor led to a major improvement of this work. Inspired by it I learned some pure mathematics, which has opened a new world for me.

Abstract

The renowned statistician George E. P. Box wrote that “essentially, all models are wrong, but some are useful.” (Box and Draper 1987) This is certainly true in finance, where many models and techniques that have been extensively empirically invalidated remain in widespread use, not just in academia, but also (perhaps especially) among practitioners. Incorrect models, and model misuse, represent a source of risk that is being increasingly recognised — this is called “model risk.”

Following on the non-parametric approach of Glasserman and Xu (2014) to model risk quantification, we develop new theory and methods for a variety of applications. The work consists of three parts. The first part focuses on the risk in applying option pricing models. In particular, there are two aspects of model risk: the inability of a chosen model to fit observed market prices at a given point in time (calibration error) and the model risk due to recalibration of model parameters (in contradiction to the model assumptions). We quantify these two sources of model risk in a common framework, and consider the trade-offs between them when choosing a model and the frequency with which to recalibrate to the market. We illustrate this approach applied to the models of Black and Scholes (1973) and Heston (1993), using option data for Apple (AAPL) and Google (GOOG).

The second part involves construction of a theory that quantifies model risk for path-dependent losses. The proposed theory generalises the relative-entropic approach of Glasserman and Xu (2014) to any f -divergence. It provides an unified treatment for all underlying dynamics and path-dependency. Three powerful tools are proposed for financial practitioners to quantify model risk. Just like derivative pricing, model risk can also be evaluated using martingale or tree approaches, or by solving partial differential equations.

The third part proposes a new approach to model risk measurement based on the Wasserstein distance between two probability measures. It formulates the theoretical motivation resulting from the interpretation of a fictitious adversary in robust risk management. The proposed approach accounts for all alternative models. It provides practically feasible results that overcome the restriction and the integrability issue imposed by the nominal model. The Wasserstein approach suits for all types of model risk problems, ranging from the single-asset hedging problem to the multi-asset allocation problem. The robust capital allocation line, accounting for the correlation risk, is not achievable with the relative-entropic approach.

Contents

List of Figures	5
List of Tables	8
1 Introduction	9
2 Literature Review	11
2.1 Market Risk Management	11
2.1.1 Coherent Risk Measures	11
2.1.2 Delta-Gamma Approach	14
2.1.3 Monte Carlo simulation	17
2.1.4 Historical Simulation	18
2.2 Risk Neutral Density and Option Pricing	19
2.2.1 Risk Neutral Probability	19
2.2.2 Derivation of Risk Neutral Density from Option Prices	21
2.2.3 Calibration of Risk Neutral Density with Information Criteria	21
2.3 Dynamic Models for Option Pricing	23
2.3.1 Black-Scholes Model	24
2.3.2 Stochastic Volatility Model	25
2.4 Parametric Approach to Model Risk Measurement	27
2.4.1 Classification of Model Risk	27
2.4.2 Model Risk Quantification by Model Comparison	28
2.4.3 Practical Model Risk Management in Derivative Pricing	33
2.4.4 Practical Model Risk Management in Hedging	35
2.5 Non-parametric Approach to Model Risk Measurement	36
2.5.1 Entropic Model Risk Measurement	36
2.5.2 Minimisation of Model Risk under Constraints	38
2.5.3 α -Divergence as a Generalised Relative Entropy Measure	39
2.5.4 Applications of Entropic Model Risk Measure	40
2.5.5 Distribution Model Risk Measurement	44
2.5.6 Wasserstein Metric and Model Uncertainty	46
2.6 Towards Dynamic Theory of Model Risk	49
2.6.1 Robust Control and Model Uncertainty	49
2.6.2 Functional Ito Calculus	54
3 Quantifying Model Risks in Option Pricing Models	57
3.1 Introduction	57

3.2	Calibration error, model risk due to recalibration, and treatment of latent state variables	60
3.3	Numerical implementation	65
3.4	Examining the trade-off between calibration error and model risk due to recalibration	67
3.5	Conclusion	72
4	Theory of Dynamic Model Risk Measurement	76
4.1	Problem Formulation	77
4.2	Characterising the Worst-Case Expected Loss	81
4.3	General Result of Model Risk Measurement	84
4.4	Model Risk Measurement with Continuous Semimartingales	96
4.5	Concluding Remarks	103
5	Application on Dynamic Model Risk Measurement	104
5.1	Introduction	104
5.2	Martingale and Tree Approaches to Model Risk Quantification	105
5.3	Differential Equations for Model Risk Quantification	108
5.4	Solutions to Problems with Linear and Quadratic Loss Functions	111
5.5	Robust Pricing of Equity Swap	114
5.6	Robust European Option Pricing	117
5.7	Robust American Option Pricing	122
5.8	Model Risk for Dynamic Portfolio Management	125
5.9	Model Risk of Portfolio Performance under loss constraint	129
5.10	Concluding Remarks	133
6	Model Risk Measurement Using Wasserstein Metric	134
6.1	Introduction	134
6.2	Basic Concepts	136
6.3	Theory	138
6.3.1	Wasserstein Formulation of the Model Risk Problem	138
6.3.2	Entropy Constraint on Transportation Plan	140
6.3.3	Main Result and Discussion	142
6.3.4	Practical Considerations	143
6.4	Application	146
6.4.1	Jump risk under a diffusive reference model	146
6.4.2	Volatility Risk and Variance Risk	147
6.4.3	Model Risk in Portfolio Variance	149
6.4.4	Robust Portfolio Optimisation and Correlation Risk	152
6.4.5	Model Risk in Dynamic Hedging	156
6.5	Conclusion	159
6.6	Appendix	160
6.6.1	A. Derivation of Eq. 6.3.10	160
6.6.2	B. Derivation of Eq. 6.3.16 and 6.3.20	161

6.6.3	C. Jump Risk and Variance Risk	162
6.6.4	D. Worst-case Portfolio Variance	163
6.6.5	E. The support of a multivariate normal distribution	166
6.6.6	F. Verification of the Wasserstein approach	170
6.6.7	G. Robust MVO Portfolio (Kullback-Leibler divergence)	172
6.6.8	H. Robust MVO Portfolio (Wasserstein approach)	174

List of Figures

3.2.1	Graphic illustration of the mathematical definitions of model risks for (a) observable state variable, (b) latent state variable.	64
3.4.1	Decomposition of model risks of AAPL options using the Black/Scholes model as the nominal model, for a selection of option maturity dates (given as the title of each graph).	74
3.4.2	Decomposition of model risks of GOOG options using the Heston model as the nominal model, for a selection of option maturity dates (given as the title of each graph).	75
5.5.1	(Upper Left) Robust price range of the hypothetical equity swap as function of the relative entropy η at different time horizons T . (Upper Right) Price range as function of time horizons at three η levels. (Lower Left) Price range as function of the volatility σ . (Lower Right) Price range as function of the interest rate r	116
5.5.2	Robust pricing for the hypothetical equity swap under the CEV model with (Left) $\gamma = 0.5$ and (Right) $\gamma = 1.1$. Upper figures are price ranges changing with the equity price at three different parametric significance levels. Middle figures show the corresponding cost measured by the relative entropy. Bottom figures show price ranges at three different entropic significance levels.	117
5.6.1	(a) worst-case risk, $\tilde{v}(0, X_0)$, at different spot prices X_0 in comparison to the nominal risk which follows the Black-Scholes PDE, (b) risks for a fixed spot price, $\tilde{v}(t, X_0)$ where $X_0 = 5000$, at different times to maturity $T - t$	118
5.6.2	Two-dimensional grids of (a) the worst-case risk $\tilde{v}(t, x)$, (b) the penalized worst-case risk $\tilde{u}(t, x)$, (c) the drift term for the change of measure, (b) the relative entropy $\vartheta(\tilde{v}(x, t) - \tilde{u}(x, t))$	119
5.6.3	8-layer trees of (a) the price process of the underlying asset, (b) the budget process calculated by $\eta_t^\vartheta = \vartheta(\tilde{V}_t^\vartheta - \tilde{U}_t^\vartheta)$, (c) the value process \tilde{U}_t^ϑ , and (d) the risk process \tilde{V}_t^ϑ	120
5.6.4	Convergence of (a) \tilde{V}_0^ϑ and \tilde{U}_0^ϑ , and (b) η_0^ϑ , with the number of layers in a CRR binomial tree, on the robust pricing of an European option.	120
5.6.5	Convergence of (a) the nominal risk and (b) the worst-case risk simulated with different numbers of paths.	121

5.7.1	8-layer trees of (a) the price process of the underlying asset, (b) the budget process η_t^ϑ , (c) the value process \tilde{U}_t^ϑ , and (d) the risk process \tilde{V}_t^ϑ .	123
5.7.2	Convergence of (a) \tilde{V}_0^ϑ and \tilde{U}_0^ϑ , and (b) η_0^ϑ , with the number of layers in a CRR binomial tree, on the robust pricing of an American option. . . .	124
5.7.3	(a) worst-case risk, $V_0^\vartheta = \tilde{v}(0, X_0)$, at different spot prices X_0 in comparison to the nominal risk which follows the Black-Scholes PDE, (b) risks for a fixed spot price, $\tilde{v}(t, X_0)$ where $X_0 = 5000$, at different times to maturity $T - t$	124
5.7.4	Two-dimensional grids of (a) the worst-case risk $\tilde{v}(t, x)$, (b) the penalized worst-case risk $\tilde{u}(t, x)$, (c) the drift term for the change of measure, (b) the relative entropy $\vartheta(\tilde{v}(x, t) - \tilde{u}(x, t))$	125
5.8.1	Expected loss (in terms of negative logarithmic return) over the investment horizon (time 0 to T) under the reference model, as a function of the rebalancing strategy $\Delta(x) = a + bx$	128
5.8.2	Expected loss (Left) and the associated relative entropy (Right) over the investment horizon (time 0 to T) under the worst-case model, as a function of the rebalancing strategy $\Delta(x) = a + bx$. Upper figures are at the parametric significance level of $\vartheta = 5$. Lower figures are at $\vartheta = 8$. Parameter choices are interest rate $r = 0.05$, drift $\mu = 0.1$, volatility $\sigma = 0.2$ and time horizon $T = 1$	129
5.9.1	Two-dimensional grids of (Upper Left) the worst-case risk $\tilde{v}(t, x)$, (Upper Right) the penalized worst-case risk $\tilde{u}(t, x)$, (Bottom Left) the drift term for the change of measure, (Bottom Right) the relative entropy. $\vartheta = 5$, $\mu_t = 0.05$, $\sigma_t = 0.2$ and $l = 0.2$	130
5.9.2	(Upper Left) Worst-case loss (measured by negative logarithmic return) and (Upper Right) relative entropy at three ϑ levels. (Middle Left) Worst-case loss and (Upper Right) relative entropy under three loss constraints. (Bottom Left) Worst-case loss changes with the time horizon T at three η levels. (Bottom Right) Worst-case loss changes with the loss constraint l at three η levels. The parameter values are $\vartheta = 5$, $\mu_t = 0.05$, $\sigma_t = 0.2$ and $l = 0.2$ unless specified in the figures.	132
6.1.1	(a) Dirac measure has a support of a single point. An alternative model with a widespread distribution cannot be related to the reference model using f -divergence. (b) State transition in a metric space. Since f -divergence does not involve the metric, Approaches to model risk measurement based on f -divergence would treat the transitions from State 1 to 2 and 1 to 3 as if they have the cost.	135
6.4.1	Worst-case volatility as a function of time under the (a) Wasserstein approach, (b) KL divergence.	148
6.4.2	Multivariate nominal distributions (a) reference model, (b) worst case under the KL divergence, (c) worst case under the Wasserstein approach (as a measure-preserving transform).	150

6.4.3	Multivariate nominal distributions (a) reference model, (b) worst case under the KL divergence, when the support is a low-dimensional subspace. Worst-case multivariate nominal distributions under the Wasserstein approach (c) $\theta = 0$ (d) $\theta = 0.5$	151
6.4.4	The normalised optimal composition of a portfolio consisting of two securities, calculated by a^* divided by $\lambda/2$. The normalised optimal composition under the reference model is give by a constant vector $\Sigma^{-1}\mu$, while those under the worst-case models are dependent on λ . In particular, the Kullback-Leibler approach reduces both compositions proportionally, while the Wasserstein approach reduces compositions in a nonlinear way.	154
6.4.5	Robust capital allocation lines (CALs) using (a) the Kullback-Leibler divergence and (b) the Wasserstein approach.	155
6.4.6	(a) Worst-case hedging risk under the KL divergence, and (b) hedging risk simulated by randomly sampling volatilities.	157
6.4.7	(a) Worst-case hedging risk under the Wasserstein approach.	158
6.4.8	(a) Sample paths generated for the Wasserstein approach, (b) convergence of the worst-case hedging risk.	158
6.6.1	Eigenvalue x^* of the effective covariance matrix Σ^* increases by a greater amount when the original eigenvalue x gets closer to zero.	176

List of Tables

3.4.1	Model risks (in terms of relative entropy) under different models and recalibration frequencies for AAPL.	70
3.4.2	Model risks (in terms of relative entropy) under different models and recalibration frequencies for GOOG.	71
3.4.3	Model risks (in terms of relative entropy) under different models, by maturity buckets, for AAPL (daily recalibration frequency).	72
3.4.4	Model risks (in terms of relative entropy) under different models, by maturity buckets, for GOOG (daily recalibration frequency).	73
6.3.1	Worst-case probability density function at different (α, β) combinations. p is the nominal distribution and u is the uniform distribution. δ denotes the Dirac δ -function and T^* is the transportation map given by Eq. 6.3.10.	143
6.3.2	Worst-case density function with prior q_0 at different (α, β) combinations. p is the nominal distribution. δ denotes the Dirac δ -function and T^* is the transportation map given by Eq. 6.3.10.	145

Chapter 1

Introduction

Risk management is a key objective in the financial industry. Traditionally, practitioners mainly focus on the market risk that is easier to quantify. The risk that originates from the mis-specification of the models is largely ignored. However, due to an increasing application of financial models, such model risk is no longer negligible. In fact, inappropriate assumptions underlying derivative pricing models and credit risk models may have contributed to the financial crisis. For practitioners, a general framework of measuring and hence managing model risk becomes essential in the industry that relies heavily on models.

The measurement of model risk also has theoretical importance, and the problem of model ambiguity requires advanced technologies on topics such as stochastic analysis. A general theory should be applicable to measuring the model risks associated with a variety of problems, both path-independent and path-dependent. The path dependency of some problems requires the use of the most recently developed functional Ito calculus. A general theory should also cover the possible losses under alternative models in a non-parametric sense. This means that the model risk should not be quantified by simply changing the values of the parameters, or by switching to another parametric model. Instead, one needs to account for all possible models described by some probability measures. The key idea is to impose a constraint to all probability measures, thus obtaining a non-parametric set of legitimate measures, based on some sort of statistical distance, or divergence. One could formulate the problem as a robust control problem, which has wide applications in both finance and economics.

In this thesis, we will approach the problem of model risk measurement from three directions. The first proposes an empirical approach to quantify the model risk by relative entropy, which is especially useful in option pricing. The second approach formulates the general theory of dynamic model risk measurement that is applicable to all f -divergence and path-dependent problems. The third one attempts to step away from the restriction of f -divergence in order to cover all probability measures, rather than only measures equivalent to some reference measure. The Wasserstein metric and optimal transport theory take an important role in this work.

This thesis is structured as follows. Chapter 2 reviews the relevant literature in the area of risk management. This includes the standard approaches of measuring

market risks, the parametric way of quantifying model risk, and the recently proposed model risk measurement by relative entropy. Chapter 2 also reviews the models widely used in the area of option pricing, both non-parametric and parametric. Hansen and Sargent's work on the problem of robust control and Cont's work on functional Ito calculus are also reviewed for their applications in following chapters. Chapter 3 proposes an empirical approach of finding the level of two types of model risk in option pricing, calibration error and model risk due to recalibration, and presents a method to quantify these in a unified framework. Chapter 4 and 5 formulate the theory of dynamic model risk measurement and apply it to path-dependent problems. Chapter 6 introduces an approach to model risk measurement based on the Wasserstein metric.

Chapter 2

Literature Review

2.1 Market Risk Management

Financial risk refers to the possibility of incurring financial losses, due to changes in certain risk factors. Risk management is the measurement and control of financial risk using some quantitative (probabilistic) approaches. (McNeil, Frey and Embrechts 2015)

Financial risk management involves several processes, including identification of risk factors, quantification of risk exposure, allocation of risk capital, risk mitigation and controlling, etc. Based on the source of risk factors, financial risks are commonly classified into market risk, credit risk, liquidity risk, operational risk, model risk, etc. Market risk refers to the possible loss from adverse movements in the mark-to-market value of a portfolio. Credit risk originates from the unexpected failure of a counterparty to fulfil its contractual obligations. Liquidity risk is the risk of significantly increased transaction costs, difficulties in converting assets to cash, or difficulties accessing funding in the market. Model risk is associated with the use of imperfect models in the financial industry, and has been widely recognised as a risk which also needs to be managed.

In this section, we will specifically discuss the conventional ways of quantifying market risk using some risk measures. The change of the market value of a portfolio can be attributed to multiple risk factors. Delta-Gamma approach, Monte Carlo simulation and historical simulation are the three dominant methods that model the risk factors and their impact on the portfolio value. It is noted that the market risk measurement conventionally does not account for possible losses due to model uncertainty. From a practical point of view, model risk needs to be added on top of the market risk.

2.1.1 Coherent Risk Measures

As a well-known risk measure, Value at Risk (VaR) is widely used by financial practitioners. Despite of that, other risk measures, such as expected shortfall (ES) and entropic value at risk (EVAR), may have certain advantages in some cases. In this

section the general requirement of a coherent risk measure is discussed.

A risk measure $m(X)$ assigns a number to any portfolio whose market return is a random variable X . Such a risk measure is coherent as long as it satisfies the following five properties:(Artzner, Delbaen, Eber and Heath 1999)

1. Normalized: $m(0) = 0$, which means that the risk of holding no assets is zero.
2. Monotonicity: $m(X) \geq m(Y)$ if $X \leq Y$. This means that the risk measured for one portfolio is higher than the other, if this portfolio has systematically lower returns.
3. Subadditivity: $m(X + Y) \leq m(X) + m(Y)$. This diversification principle implies a lower risk by adding more assets.
4. Homogeneity: $m(\alpha X) = \alpha m(X)$ if $\alpha \geq 0$. This means that the risk measure should be linearly scalable.

Sometimes the notion of coherence is too restrictive. Instead a more general set of risk measures can be viewed as being reasonable. In a convex risk measure, the properties of subadditivity and homogeneity are relaxed and replaced by a convexity condition:(Föllmer and Schied 2002)

$$m(\lambda X + (1 - \lambda)Y) \leq \lambda m(X) + (1 - \lambda)m(Y), \lambda \in [0, 1] \quad (2.1.1)$$

The most common risk measure, Value at Risk (VaR), is defined as the maximum loss over a target horizon within a specified confidence interval. It is simply given by the percentile of the profit & loss distribution:

$$P(L > VaR) = 1 - c \quad (2.1.2)$$

where L is the possible loss and c denotes the confidence level. A 99% confidence level means that the probability of experiencing loss greater than VaR is less than 1%. It is well-known that in general VaR is not a coherent risk measure as it does not always obey the subadditivity condition (Artzner et al. 1999). However, VaR is coherent if the possible loss is elliptically distributed (e.g. normally distributed), though under such distributions VaR is equivalent to a variance risk measure.

On the other hand, Expected Shortfall (ES) is a coherent risk measure. It is defined as the "average" VaR under confidence levels that are lower than a given level α :(Acerbi and Tasche 2002)

$$ES_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_\gamma(X) d\gamma \quad (2.1.3)$$

where $VaR_\gamma(X)$ is the VaR of the same portfolio at a certain confidence level (γ is the complement of the confidence level, i.e. $\gamma = 0.01$ for a 99% confidence level). If the loss follows a continuous distribution, then ES is equal to the expected value of loss, conditional to losses that exceed the VaR level:

$$ES_\alpha(X) = E[-X | X \leq -VaR_\alpha(X)] \quad (2.1.4)$$

Apart from ES, the entropic value at risk (EVaR) is also a coherent risk measure. In a dual representation, EVaR is defined by (Ahmadi-Javid 2012)

$$EVaR_{1-\alpha}(X) = \sup_{Q \in \Theta} (E^Q(-X)) \quad (2.1.5)$$

where Θ is a set of probability measures on the probability space (Ω, F) , which is constraint by a certain “distance” from the nominal probability measure P (i.e. the probably measure used to estimate VaR and ES). Here the “distance” is defined by the relative entropy between the two probability measures Q and P :

$$\Theta = \{Q : D(Q||P) \leq -\ln \alpha\} \quad (2.1.6)$$

where

$$D(Q||P) = \int \frac{dQ}{dP} \left(\ln \frac{dQ}{dP} \right) dP \quad (2.1.7)$$

By introducing a Lagrange multiplier θ^{-1} , we could convert the constraint maximisation problem Eq. 2.1.5 to a dual problem:

$$EVaR_{1-\alpha}(X) = \inf_{\theta > 0} \sup_{Q \in \Theta} [E_Q(-X) - \theta^{-1} (D(Q||P) + \ln \alpha)] \quad (2.1.8)$$

$$= \inf_{\theta > 0} \left[\theta^{-1} \ln E \left(\frac{e^{-\theta X}}{\alpha} \right) \right] \quad (2.1.9)$$

It is noted that by definition EVaR is equivalent to the entropic model risk measure introduced by Glasserman and Xu (2014), which is the main topic studied in this thesis. There is only a slight conceptual difference, as the EVaR measures risks that are not only due to imperfect models. It is also used for measuring market risks, as EVaR can be viewed as a measure of tail loss alternative to VaR and ES. Instead of simply using the quantile a (where $P(-X \geq a) = \alpha$) as the risk measure, EVaR provides the lower boundary of a set of risk measures Φ :

$$EVaR_{1-\alpha}(X) = \inf_{a \in \Phi_\alpha} a \quad (2.1.10)$$

where

$$\Phi_\alpha = \{a : e^{-\theta a} E(e^{-\theta X}) = \alpha, \theta > 0\} \quad (2.1.11)$$

According to the Chernoff bound for a random variable $-X$,

$$P(-X \geq a) \leq e^{-\theta a} E(e^{-\theta X}) \quad (2.1.12)$$

Therefore, Φ_α is a subset of $\{a : Pr(-X \geq a) \leq \alpha\}$. This illustrates the fact that EVaR is a measure of the tail loss, and its value should be higher than the VaR. In fact, one can prove that under the same confidence level $1 - \alpha$,

$$VaR(X) \leq ES(X) \leq EVaR(X) \quad (2.1.13)$$

Ahmadi-Javid (2012) also generalise the EVaR to a class of entropic risk measures by incorporating the class of generalised relative entropies. The generalised relative entropy is a pseudo-distance or divergence measure from Q to P :(Ahmadi-Javid 2012)

$$H_f(P, Q) := \int f \left(\frac{dQ}{dP} \right) dP \quad (2.1.14)$$

where f is a convex function of the Radon-Nikodym derivative. Accordingly the so-called f -entropic risk measure with divergence level (or relative entropy budget) β is defined by the dual representation:

$$ER_{f,\beta}(X) := \sup_{Q \in \Theta} (-X) \quad (2.1.15)$$

where $\Theta = \{Q : H_f(P, Q) \leq \beta\}$. It is noted that EVaR and ES are both f -entropic risk measures.

Obviously, a f -entropic risk measure obeys the subadditivity property as

$$\begin{aligned} EVaR_{1-\alpha}(X + Y) &= E^{Q_{X+Y}}(-X - Y) \\ &= E^{Q_{X+Y}}(-X) + E^{Q_{X+Y}}(-Y) \\ &\leq \sup_{Q \in \Theta} (E^Q(-X)) + \sup_{Q \in \Theta} (E^Q(-Y)) \\ &= EVaR_{1-\alpha}(X) + EVaR_{1-\alpha}(Y) \end{aligned} \quad (2.1.16)$$

where the probability measure $Q_{X+Y} \in \Theta$ maximises the expected loss of the portfolio combination $X + Y$. This is in consistent with the fact that both EVaR and ES are coherent measures. In general, a f -entropic risk measure can be evaluated by solving the minimising problem:

$$ER_{f,\beta}(X) = \inf_{\theta > 0, \mu \in \mathbb{R}} \left\{ \theta^{-1} [\mu + E^P (f^* (\theta X - \mu + \beta))] \right\} \quad (2.1.17)$$

where f^* is the conjugate (the Legendre-Fenchel transform) of f .

It is noted that Eq. 2.1.14 is more commonly referred as the f -divergence, and is usually denoted by $D_f(Q||P)$. In order for it to be well defined, f needs to be a convex function with $f(1) = 0$ and Q is absolutely continuous with respect to P . Examples of the function $f(x)$ involve $x \ln x$ (KL divergence or relative entropy), $-\ln x$ (reverse KL divergence) and $(x^\alpha - x)/(\alpha(\alpha - 1))$ (α -divergence).

2.1.2 Delta-Gamma Approach

To evaluate the distribution of possible losses and hence the value of the selected risk measure, we need to analyse the exposure of a portfolio to various risk factors. Common risk factors include interest rate risk, stock market risk, credit risk, currency risk, commodity risk, etc. As risks spread across both time (different maturity dates) and instruments (e.g. different stocks), a rigorous risk analysis is difficult and, at a full level of granularity, usually infeasible. This requires the use of risk factor mapping techniques. For example, we map the various systematic risks of stocks to one exposure to the stock index.

Not all exposures are linear with respect to the risk factors. Bond positions are in fact non-linear exposure to the interest rate risks, and derivatives such as options have non-linear exposures as well. A reasonable approximation of dealing with non-linear exposure is by taking the first and second order terms in the Taylor expansion of portfolio value $V(R_1, \dots, R_n)$, (McNeil et al. 2015)

$$\begin{aligned}\delta V(R_1, \dots, R_n) &= \sum_{i=1}^n \tilde{\Delta}_i R_i + \frac{1}{2} \sum_{i,j=1}^n R_i \tilde{\Gamma}_{ij} R_j + O(R_i R_j R_k) \\ &\approx \tilde{\Delta}^T R + \frac{1}{2} R^T \tilde{\Gamma} R\end{aligned}\quad (2.1.18)$$

where R is a column vector of random variables, which represent the log-returns of all mapped risk factors, $R_i = \ln r_i$, where r_i is the arithmetic return of the i -th mapped risk factor. $\tilde{\Delta}$ is the column vector that consists of cash deltas in percentage terms, i.e. portfolio value change due to a hundred percent change of risk a factor. $\tilde{\Gamma}$ is a matrix of cash gamma:

$$\tilde{\Delta}_i = \frac{\partial V}{\partial \ln r_i} = r_i \frac{\partial V}{\partial r_i} \quad (2.1.19)$$

$$\tilde{\Gamma}_{ij} = \frac{\partial^2 V}{\partial \ln r_i \partial \ln r_j} = r_j \frac{\partial \left(r_i \frac{\partial V}{\partial r_i} \right)}{\partial r_j} \quad (2.1.20)$$

In the variance-covariance approach, the target of risk measurement is to calculate the portfolio variance based on the covariance matrix of the returns of risk factors. Following the delta-gamma approximation, the variance of the portfolio value is approximated by

$$\sigma^2(\delta V) \approx \sigma^2(\tilde{\Delta}^T R) + \frac{1}{4} \sigma^2(R^T \tilde{\Gamma} R) + \text{cov}(\tilde{\Delta}^T R, R^T \tilde{\Gamma} R) \quad (2.1.21)$$

If assuming the log-returns of risk factors follow a multivariate Gaussian distribution, i.e. $R \sim N(\mu, \Omega)$, the last term in Eq. 2.1.21 is zero according to the Stein's Lemma. For simplicity, we will only discuss this special case unless otherwise stated.

A natural method of dealing the correlated risk factors is to normalise them into an orthogonal set. This is done by applying a symmetric matrix $\Omega^{-1/2}$ in which $\Omega^{-1/2} \Omega^{-1/2} = \Omega^{-1}$ where Ω is the covariance matrix of the risk factors. The orthogonalised random vector $Y = \Omega^{-1/2} R \sim N(0, I)$. By further diagonalising the matrix $A = \Omega^{1/2} \tilde{\Gamma} \Omega^{1/2}$, the term $R^T \tilde{\Gamma} R$ is thus expressed by the transformed vectors:

$$R^T \tilde{\Gamma} R = Y^T A Y = X^T \Lambda X \quad (2.1.22)$$

where

$$X = O^T Y = O^T \Omega^{-1/2} R \quad (2.1.23)$$

$$\Lambda = O^T A O = O^T \Omega^{1/2} \tilde{\Gamma} \Omega^{1/2} O \quad (2.1.24)$$

Λ is the diagonal matrix of eigenvalues and O is the orthogonal matrix of corresponding eigenvectors. Applying the introduced matrices to the delta term in Eq. 2.1.21, we obtain

$$\tilde{\Delta}^T R = \tilde{\Delta}^T \Omega^{1/2} Y$$

$$\begin{aligned}
&= \tilde{\Delta}^T \Omega^{1/2} O X \\
&= L^T X
\end{aligned} \tag{2.1.25}$$

where the transformed sensitivity vector is defined by $L := O^T \Omega^{1/2} \tilde{\Delta}$. This finally transforms the change of the portfolio value into a Taylor expansion series with respect to a set of uncorrelated risk factors, $X \sim N(0, I)$:

$$\delta V = \sum_i \left(L_i X_i + \frac{1}{2} \lambda_i X_i^2 \right) \tag{2.1.26}$$

where λ_i is the i -th diagonal element of Λ , $X_i \sim N(0, 1)$ and $X_i^2 \sim \chi^2(1)$.

The variance of the portfolio value change can be directly calculated from the covariance matrix of the returns of risk factors:

$$\sigma^2(\delta V) = \tilde{\Delta}^T \Omega \tilde{\Delta} + \frac{1}{2} tr \left(\Omega \tilde{\Gamma}^2 \right) \tag{2.1.27}$$

If assuming the change of portfolio value follows a normal distribution, VaR can be naively calculated by $VaR(c) = -\alpha \sigma(\delta V)$, where α is the percentile of a standard normal distribution. However, this assumption is usually not true (unless $\tilde{\Gamma}$ is a zero matrix). According to Eq. 2.1.26, the change of portfolio value is in fact the sum of a normally distributed random variable and a χ^2 -distributed random variable. A more appropriate approach is therefore required to adjust for the higher moments (e.g. skewness, kurtosis, etc.).

Cornish-Fisher series expansion provides a method of approximating the percentiles of a distribution from its central moments. The normalised percentile is approximated by (Cornish and Fisher 1938)

$$\alpha_{CF} \approx \alpha + \frac{\alpha^2 - 1}{6} \rho_3 + \frac{\alpha^3 - 3\alpha}{24} (\rho_4 - 3) - \frac{2\alpha^3 - 5\alpha}{36} \rho_3^2 \tag{2.1.28}$$

where α is the quantile of a standard normal distribution for the given confidence level. The skewness of the distribution of δV can be calculated from the cash delta vector, the cash gamma matrix, and the covariance matrix of risk factor returns:

$$\rho_3 = \frac{3\tilde{\Delta}^T \Omega \tilde{\Gamma} \Omega \tilde{\Delta} + tr \tilde{\Gamma} \Omega^3}{\left(\tilde{\Delta}^T \Omega \tilde{\Delta} + \frac{1}{2} tr \tilde{\Gamma} \Omega^2 \right)^{\frac{3}{2}}} \tag{2.1.29}$$

The kurtosis is calculated in a similar way,

$$\rho_4 = \frac{12\tilde{\Delta}^T \Omega \tilde{\Gamma} \Omega^2 \tilde{\Delta} + 3tr \tilde{\Gamma} \Omega^4 + 3\mu_2^2}{\left(\tilde{\Delta}^T \Omega \tilde{\Delta} + \frac{1}{2} tr \tilde{\Gamma} \Omega^2 \right)^2} \tag{2.1.30}$$

where μ_2 is the second central moment (i.e. variance) which is calculated from Eq. 2.1.27.

2.1.3 Monte Carlo simulation

Delta-gamma approach is merely an approximation for market risk measurement. Due to the recent advance of computational power, the entire distribution of the change of portfolio value can be evaluated via Monte Carlo simulation. By doing that, we need to simulate a large number of possible paths, according to the modelled dynamics of the risk factors. In this sense, the Monte Carlo simulation can still be regarded as a parametric approach to risk measurement. The advantages of Monte Carlo simulation include fully incorporation of correlations, full evaluation of portfolio values, and little demand for historical data. (McNeil et al. 2015, Dowd 2007)

Because the number of risk factors for some portfolios is too large, we may need to adopt a technique for dimension reduction called the Principal Component Analysis (PCA). It is essentially transforming the risk factors into an orthogonal set, based on the eigen-decomposition of their covariance matrix. If denoting the random vector of risk factors by R , and their covariance matrix by Ω , diagonalisation of Ω gives

$$\Omega = O\Lambda O^T \quad (2.1.31)$$

where Λ is a positive diagonal matrix consisting of all the eigenvalues of Ω , with diagonal elements sorted in a descending order. O is the orthogonal matrix of eigenvectors. The definition of eigenvalue and eigenvectors give

$$\Omega O = O\Lambda \quad (2.1.32)$$

These two equations lead to the orthogonal condition of O , i.e. $O^T O = I$. If among all the eigenvalues, only the first n (principal components) are taken into account, then the random vector R is reconstructed by a smaller amount of uncorrelated random variables

$$R \approx O_{1:n}\epsilon_{1:n} \quad (2.1.33)$$

where $O_{1:n}$ is the matrix of the first n eigenvectors (i.e. a matrix of n columns), and $\epsilon_{1:n}$ is a random vector of n uncorrelated variables. Their variances are given by the n largest eigenvalues of the covariance matrix Ω .

After dimension reduction via PCA, we can start to run the Monte Carlo simulation of risk factors. It starts with the sampling of paths for n independent principal components ϵ_i ($i = 1, 2, \dots, n$). In a simple setting, we may assume that ϵ_i follows a geometric Brownian motion without drift:

$$\frac{d\epsilon_i}{\epsilon_i} = \sqrt{\lambda_i} dW \quad (2.1.34)$$

where λ_i is the variance of ϵ_i , or equivalently the corresponding eigenvalue of Ω . The geometric Brownian motion has a simple analytic solution:

$$\epsilon_i(t) = \epsilon_i(0) \exp\left(\frac{\lambda_i}{2}t + \sqrt{\lambda_i}\sqrt{t}\varepsilon\right) \quad (2.1.35)$$

where ε is sampled according to the standard Gaussian distribution. This allows us to sample a large amount of paths according to Eq. 2.1.35, and calculate the dynamics of actual risk factors (Eq. 2.1.33). Using the Monte Carlo simulation, the portfolio value in each sampled path can be calculated via full valuation. The quantile of the future change of portfolio value can be directly obtained from a large set of sampled paths.

2.1.4 Historical Simulation

If historical data of the mapped risk factors is adequate, we could use a non-parametric approach to simulate the distribution of portfolio value changes. This approach called historical simulation has advantages when the future dynamics of risk factors are hard to model. The time evolution of the risk factors and their dependencies are inferred directly from the historical observations. By applying the historical simulation, we are in fact assuming that the dynamics of the risk factors in the near future can be approximated by its historical path. (McNeil et al. 2015, Dowd 2007)

Historical simulation calculates the values of risk measures (e.g. VaR) from a non-parametric distribution of the portfolio value change, based on the historical dynamics and dependencies of risk factors. It evaluates the portfolio value change using full valuation, thus delta-gamma approximation is not required. Historical simulation applies the current weights to risk factors to a set of historical returns of risk factors:

$$R_{p,t} = \sum_{i=1}^n w_{i,T} R_{i,t}, \quad t = 1, \dots, T \quad (2.1.36)$$

where $R_{p,t}$ is the portfolio return at the historical snapshot t . $w_{i,T}$ is the weight of the i -th risk factor at the current time T , and $R_{i,t}$ is the historical return of the i -th risk factor.

In a standard historical simulation, we put the historical observations of risk factors into a $T \times n$ matrix, where n and T are the numbers of risk factors and historical time periods, respectively: $[S_{i,j}]$, $i = 1, \dots, n$, $j = 1, \dots, T$. This is followed by the calculation of the historical arithmetic returns of risk factors: $\Delta S_{i,j} = (S_{i,j} - S_{i,j-1})/S_{i,j-1}$, which results in a $(T - 1) \times n$ matrix. After that, we need to use the return matrix and the current values of risk factors to calculate the adjusted set of risk factor:

$$S_{i,j}^* = S_{i,T}(1 + \Delta S_{i,j}) \quad (2.1.37)$$

The result is a $T \times n$ matrix of simulated risk factors, which is different from the historically observed values.

Now we can calculate the distribution of profit and loss using the set of simulated risk factors. For each sample of the simulated risk factors (i.e. each row of the $T \times n$ matrix), we calculate the portfolio value via full valuation. This provides us a set of portfolio values, and hence a non-parametric distribution of the change of portfolio value.

It is widely believed by financial practitioners that the recent observations usually have stronger predictive power. In practice, we may not want to weight all the historical observations evenly. This introduces the weighted historical simulation, which assigns exponentially declining weights through the past:

$$\lambda_j = \frac{\lambda^{T-j}(1-\lambda)}{1-\lambda^T} \quad (2.1.38)$$

where λ characterises the rate of exponential decay of the weights. The risk measures are calculated as usual, except that the non-parametric distribution takes into account the weights. Apart from the time weighting, practitioners might also want to weight by volatility. In periods with higher volatility, returns are higher in magnitude. To be consistent with the current volatility level, we would like to normalise all historical returns by their volatilities. To implement the volatility weighting, we simply need to replace Eq. 2.1.37 with

$$S_{i,j}^* = S_{i,T} \left(1 + \frac{\sigma_{i,T}}{\sigma_{i,j}} \Delta S_{i,j} \right) \quad (2.1.39)$$

where $\sigma_{i,j}$ is the historical (or GARCH or EWMA) forecast of the volatility of the return of the i -th risk factor at time j .

2.2 Risk Neutral Density and Option Pricing

2.2.1 Risk Neutral Probability

In the classic asset pricing theory, the existence of a risk neutral measure results from the non-arbitrage hypothesis. Here we introduce this concept by considering a simple finite-state one-period setting (Duffie 2010). Suppose we have a finite set of states $\{1, \dots, S\}$ and N securities. Under a state j , the payoff of a security i is given by D_{ij} . This gives a payoff matrix D . Now denoting the vector of portfolio weights by $\theta \in \mathbb{R}^N$, we calculate the portfolio payoff to be $D^T \theta$. If the vector of market prices of the N securities is denoted by q , then the current portfolio price is $q\theta$. The non-arbitrage condition states that $q\theta > 0$ if $D^T \theta > 0$, and $q\theta \geq 0$ if $D^T \theta \geq 0$.

One can prove that the non-arbitrage condition is equivalent to the existence of a state-price vector $\phi \in \mathbb{R}_{++}^S$, with $q = D\phi$. In fact, let $M = \{(-q\theta, D^T \theta) : \theta \in \mathbb{R}^N\}$ to be a linear subspace of $L = \mathbb{R} \times \mathbb{R}^S$. Another subspace, the arbitrage subspace, is defined on L as well: $K = \mathbb{R}_+ \times \mathbb{R}_+^S$. As both K and M are closed and convex, there is no arbitrage if and only if K and M intersect precisely at 0.

Suppose $K \cap M = \{0\}$. The Separating Hyperplane Theorem implies the existence of a nonzero linear functional $F : L \rightarrow \mathbb{R}$, such that $F(z) < F(x)$ for any $z \in M$ and nonzero $x \in K$ (Boyd and Vandenberghe 2004). Since M is a linear space, the only way of satisfying $F(z) < F(x)$ is to have $F(z) = 0$ everywhere in M . Therefore, $F(x) > 0$ for all nonzero $x \in K$. This means that the linear functional takes the form of $F(v, c) = \alpha v + \phi c$ where $\alpha > 0$ and $\phi \in \mathbb{R}_{++}^S$. This proves that $-\alpha q\theta + \phi(D^T \theta) = 0$ for

any $\theta \in \mathbb{R}^N$, and thus the existence of a state-price vector ϕ where $q = D\phi$. Conversely, a state-price vector ϕ leads to $q\theta = \phi^T D^T \theta$, which obviously follow the non-arbitrage condition.

It is noted that if the number of linearly independent securities N is equal to the number of states S (i.e. complete market), then the state-price vector is uniquely determined by $\phi = D^{-1}q$. If the number of linearly independent securities is less (i.e. incomplete market), then there are an infinite number of state-price vectors. These possible state-price vectors correctly price the securities according to their payoffs, $q = D\phi$. Additional information is required to select the most reasonable state-price vector. Economic context could be one criteria to parametrise the state-price vector, as the security prices reflect economic optimality in terms of the utility.

Suppose the market has a homogenous utility function $U(c)$. The function is maximised at c^* , where c^* is the vector of future cash positions at all possible states. The first condition for optimality states that the marginal utility is zero for any additional portfolio θ with zero price, i.e. for any $\theta \in \mathbb{R}^N$, $\partial U(c^*)^T D^T \theta = 0$ if $q\theta = 0$. This implies that $\partial U(c^*)^T D^T = \mu q$ where μ is a scalar. Assuming the payoff matrix is not a zero matrix (we do not consider the trivial case of zero payoff), then there exists some portfolio θ' with $D^T \theta' > 0$. Due to the lack of arbitrage, the portfolio price $q\theta' > 0$. We thus have $\mu q\theta' = \partial U(c^*)^T D^T \theta' > 0$. This results in the conclusion that the scalar $\mu > 0$ and

$$q = \frac{1}{\mu} D \partial U(c^*) \quad (2.2.1)$$

Therefore, if we know exactly the (high-dimensional) utility function $U(c)$, we are capable of determining the state-price vector uniquely.

It is noted that the risk neutral probability is simply a normalised version of the state-price vector, as the security i can be priced by

$$q_i = \sum_{j=1}^S \phi_j D_{ij} = \left(\sum_{j=1}^S \phi_j \right) \sum_{j=1}^S \frac{\phi_j}{\sum_{j=1}^S \phi_j} D_{ij} \quad (2.2.2)$$

where $\sum_{j=1}^S \phi_j$ is the discount factor (on a riskless asset), and $\frac{\phi_j}{\sum_{j=1}^S \phi_j}$ is the risk-neutral probability of each state j , $j = 1, 2, \dots, S$. The concept of risk-neutral probability can be readily applied to continuous state spaces. For European options, the option prices are completely determined by the risk-neutral density of the underlying asset price at the maturity. In this sense, a risk-neutral density can be regarded as a non-parametric model for pricing European options.

A well calibrated risk-neutral density $p(x)$ should give prices of call options with different strikes the same as their market prices:

$$C(K) = e^{-r\tau} E^P[(S - K)^+] = e^{-r\tau} \int_{x \in \mathbf{X}} p(x)[S(x) - K]^+ dx \quad (2.2.3)$$

In practice, the market prices $C(K)$ are only available for a finite set of strike prices K . The non-parametric model $p(x)$, on the other hand, is a continuous function that

has an infinite dimension. Therefore, there is an infinite number of models that fit the finite number of option prices to their market prices. To calibrate the market to a unique risk-neutral density, additional restrictions are required to interpolate the regimes between the available strikes. In practice, one could use a parametric form of $p(x)$ to fit the option prices. Here we introduce two types of general parametric forms: a moment-based series expansion and an entropic calibration approach.

2.2.2 Derivation of Risk Neutral Density from Option Prices

We first consider the case of a continuous set of strikes. According to the risk neutral pricing formula Eq. 2.2.3, a butterfly (long one for options strike at $K - \Delta K$ and $K + \Delta K$ and short two for option strike at K) is priced by

$$C(K + \Delta K) + C(K - \Delta K) - 2C(K) = e^{-r\tau} \int_{-\Delta K}^{\Delta K} p(K + x) (\Delta K - |x|) dx \quad (2.2.4)$$

Taking the limit ΔK , we link the second derivative of the option price to the risk neutral probability density function, (Breedon and Litzenberger 1978, Schlogl 2013)

$$\frac{\partial^2 C(K)}{\partial K^2} = e^{-r\tau} p(K) \quad (2.2.5)$$

Therefore, with a quasi-continuous set of strike prices, the risk-neutral probability could be calculated by simply taking the second derivative of the option price C with respect to the strike K . If expressing the method by the finite difference approach:

$$p(K_2) = \frac{2e^{rT}}{K_1 - K_3} \left[\frac{C(K_1, T) - C(K_2, T)}{K_1 - K_2} - \frac{C(K_3, T) - C(K_2, T)}{K_3 - K_2} \right] \quad (2.2.6)$$

where K_1 , K_2 and K_3 are three neighbouring strikes in ascending order, and T is the expiry date of these options. The numerical computation of the second-order finite difference is sensitive to pricing errors. Data smoothing and interpolation may be performed before evaluating the finite difference.

2.2.3 Calibration of Risk Neutral Density with Information Criteria

An alternative interpretation on the simplicity of the risk neutral probability measure is on the information that it carries. A simple model should contain as little arbitrary information, in the form of model assumptions, as possible. Therefore, a proper risk neutral density should fit all option prices, and thus contain all information implicit in option prices, but no more.

Under this logic we use a different calibration approach that relies on certain information criteria, such as the Principle of Maximum Entropy (PME) and the Principle of Minimum Relative Entropy (PMRE) (Buchen and Kelly 1996). These entropic criteria are in fact a direct application of the information theory and the efficient market hypothesis (EMH). The combination of these two theories states that the market-calibrated risk-neutral density should minimise the relative entropy with the prior (or nominal) distribution, assuming the market is efficient and no extra information is available.

Principle of Maximum Entropy

Under the PME, the calibration problem is converted to a constrained optimisation problem: maximising the entropy value subject to the condition that options at any strikes are priced correctly (Buchen and Kelly 1996).

$$p^*(x) = \arg \inf_{p(x)} D[p(x)||u(x)] \quad (2.2.7)$$

$$s.t. \int p(x)\mathbf{Z}(x; \mathbf{k})dx = \mathbf{C}(k),$$

where $u(x)$ is the uniform probability distribution function. The component of the vector $\mathbf{Z}(x; \mathbf{k})$, $Z(x; k)$, is the option payoff as a function of the price variable (at maturity) x and the strike price k .

By solving the constrained optimisation problem using Lagrange multipliers, the closed-form expression of the calibrated measure is given by

$$p^*(x) = \frac{\exp(\lambda \mathbf{Z}(x; \mathbf{k}))}{\int \exp(\lambda \mathbf{Z}(x; \mathbf{k}))dx}, \quad (2.2.8)$$

where the Lagrange multipliers

$$\lambda = \arg \sup_{\lambda} \frac{\int e^{\lambda \mathbf{Z}} (\lambda \mathbf{Z} - \ln \int e^{\lambda \mathbf{Z}} dx) dx}{\int e^{\lambda \mathbf{Z}} dx}. \quad (2.2.9)$$

Alternatively, the Lagrange multipliers can be calculated by solving the constraint conditions in Eq. 2.6.4. For European call options, the constraint conditions are expressed in terms of a discrete set of strikes $\{k_i\}$ (in ascending order):

$$\sum_{i=j}^n \left(\int_{k_i}^{k_{i+1}} p_i(x; \lambda) (x - k_j) dx \right) = C(k_j) \quad (2.2.10)$$

where

$$p_j(x; \lambda) \propto \exp \left[\sum_{i=1}^j \lambda_i (x - k_i) \right] \quad (2.2.11)$$

The Lagrange multipliers λ_i are calculated by solving the set of equations Eq. 2.2.10.

Principle of Minimum Relative Entropy

To use the Principle of Minimum Relative Entropy (PMRE), we need to nominate a prior or nominal probability measure. For the purpose of model risk, we could simply use the interpolated measure on the last trading day as the prior measure (Buchen and Kelly 1996). The risk-neutral density calibrated using PMRE is the unique ‘‘satisfying’’ density function, which has the minimum amount of relative entropy from the prior

measure. Here “satisfying” means that all options are priced correctly under the given measure. By formulating this logic as a constrained optimisation problem, we get

$$p^*(x) = \arg \inf_{p(x)} D[p(x)||p_0(x)] \quad (2.2.12)$$

$$s.t. \int p(x)\mathbf{Z}(x; \mathbf{k})dx = \mathbf{C}(\mathbf{k}),$$

where $p_0(x)$ is the prior distribution (the distribution calibrated at the last trading day is commonly used). $\mathbf{Z}(x; \mathbf{k})$ is the payoff as a function of the price variable (at maturity) x and strike \mathbf{k} . By solving the corresponding dual problem, we obtain the Minimum Relative-Entropy Distribution (MRED):

$$p^*(x) = \frac{p_0(x) \exp(\boldsymbol{\lambda}\mathbf{Z}(x; \mathbf{k}))}{E^{P_0}[\exp(\boldsymbol{\lambda}\mathbf{Z}(x; \mathbf{k}))]} \quad (2.2.13)$$

where the vectorised Lagrange multiplier $\boldsymbol{\lambda}$ is the one that maximises the relative entropy:

$$\boldsymbol{\lambda} = \arg \sup_{\boldsymbol{\lambda}} \frac{E^{P_0} [e^{\boldsymbol{\lambda}\mathbf{Z}} (\boldsymbol{\lambda}\mathbf{Z} - \ln E^{P_0}(e^{\boldsymbol{\lambda}\mathbf{Z}}))]}{E^{P_0}[e^{\boldsymbol{\lambda}\mathbf{Z}}]} \quad (2.2.14)$$

Similar to PME, under the PMRE criteria the vectorised Lagrange multiplier can also be calculated by solving a set of equations:

$$\sum_{i=j}^n \left(\int_{k_i}^{k_{i+1}} p_i(x; \boldsymbol{\lambda})(x - k_j)dx \right) = C(k_j) \quad (2.2.15)$$

where

$$p_j(x; \boldsymbol{\lambda}) \propto p_0(x) \exp \left[\sum_{i=1}^j \lambda_i(x - k_i) \right] \quad (2.2.16)$$

2.3 Dynamic Models for Option Pricing

Although options can be priced simply using a parametric or non-parametric risk neutral density, such approaches do not explicitly provide the link between maturities. Therefore, although the risk neutral density approaches are widely used for calibrating the volatility curve, they are not sufficient for calibrating or pricing the entire volatility surface. To obtain the entire volatility surface, a model that describes the dynamics of the underlying asset price is essential.

The simplest dynamic pricing model is based on the Black-Scholes dynamics, which takes the underlying asset price as the only state variable and models it as a (drifted) geometric Brownian motion. A more complicated set of models includes the (instantaneous) volatility as a second state variable, which follows some autoregressive stochastic processes. A Levy process is a combination of continuous Ito processes and Poisson

processes, which attempts to model the jumps on stock prices. Local volatility model maps an arbitrary risk-neutral density to a price-dependent volatility function. The corresponding inter-temporal relation is derived from the evolution of the risk-neutral density via the Fokker-Planck equation.

Despite of the various types of dynamic option pricing models, they share the same properties as they all model the stochastic process of the underlying asset price. In fact, if the underlying asset price is modelled correctly, the volatility surface can be uniquely determined by using risk-neutral pricing or benchmark pricing.

2.3.1 Black-Scholes Model

Under the Black-Scholes dynamics, the underlying asset follows a geometric Brownian motion: (Black and Scholes 1973)

$$dS_t = S_t(\mu + r)dt + S_t\sigma dW_t \quad (2.3.1)$$

Based on the continuous hedging argument, the arbitrage-free condition implies a partial differential equation

$$\frac{\partial C}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0 \quad (2.3.2)$$

where $C(t, S)$ is the fair price of a derivative at time t , given the price of the underlying asset S . σ is the constant volatility of the underlying asset price and r is a constant risk-free rate. The equation above is subject to the terminal condition $C(T, S) = V(S)$ where $V(\cdot)$ is the payoff function of the derivative at its expiration T .

By the Feynmann-Kac equation, the fair price of an European call option, as the solution to the partial differential equation above, can be represented as the discounted expected payoff under the risk-neutral measure Q , i.e. the equivalent measure under which $\mu = 0$:

$$\begin{aligned} C_t(K, T) &= e^{-r(T-t)} E^Q [(S_T - K)^+] \\ &= S_t \Phi(d^*) - K e^{-r(T-t)} \Phi(d^* - \sigma\sqrt{T-t}) \end{aligned} \quad (2.3.3)$$

where

$$d^* = \frac{\ln(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \quad (2.3.4)$$

The objective measure P and the risk-neutral measure Q are linked by the Radon-Nikodym derivative

$$\frac{dQ_t}{dP_t} = \exp\left(-\frac{\mu}{\sigma}W_t - \frac{\sigma^2}{2}t\right) \quad (2.3.5)$$

which transforms the Brownian motion under the objective measure to

$$dW_t^Q = dW_t + \frac{\mu}{\sigma}dt \quad (2.3.6)$$

This change of measure transforms the prices of the underlying asset to their risk-neutral dynamics:

$$dS_t = S_t(rdt + \sigma dW_t^Q) \quad (2.3.7)$$

2.3.2 Stochastic Volatility Model

If expressing the instantaneous volatility as an additional state variable, the option price may be determined by evaluating the stochastic processes of the two state variables. First we consider a class of continuous stochastic volatility models (Platen and Heath 2006),

$$dS_t = S_t(rdt + \sigma_t dW_t) \quad (2.3.8)$$

where volatility state variable follows a factor process $X_t = Y(\sigma_t)$,

$$dX_t = C(X_t)dt + D(X_t)(\rho dW_t + \sqrt{1 - \rho^2} d\tilde{W}_t) \quad (2.3.9)$$

where W_t and \tilde{W}_t are two independent standard Brownian motions under the risk-neutral measure (assuming its existence). ρ is a correlation coefficient, and $C(X_t)$ and $D(X_t)$ are functions of the state variable X_t . The Kolmogorov forward equation can be expressed for the transition density $p(s, x, t, y)$ of the factor process:

$$\frac{\partial p(s, x, t, y)}{\partial t} + \frac{\partial(C(y)p(s, x, t, y))}{\partial y} - \frac{1}{2} \frac{\partial^2(D^2(y)p(s, x, t, y))}{\partial y^2} = 0 \quad (2.3.10)$$

The Kolmogorov forward equation allows us to calculate the stationary density of the factor process:

$$\bar{p}(y) = \lim_{t \rightarrow \infty} p(s, x, t, y) \quad (2.3.11)$$

$$= \frac{A}{D(y)^2} \exp \left\{ 2 \int_{y_0}^y \frac{C(u)}{D(u)^2} du \right\} \quad (2.3.12)$$

where A is the normalisation constant.

Now let us consider one of the most popular stochastic volatility models: the Heston model. Under the Heston dynamics, the price process and the volatility process of the underlying asset are (Heston 1993)

$$dS_t = rS_t dt + \sigma_t S_t dW_1 \quad (2.3.13)$$

$$d(\sigma_t^2) = k[\theta - \sigma_t^2]dt + \eta \sigma_t dW_2 \quad (2.3.14)$$

This model involves two state variables, S_t and σ_t , and five model parameters: r , k , θ , η , ρ where ρ is the correlation coefficient between the two Wiener processes: $\langle W_1, W_2 \rangle = \rho dt$. r is the risk-free rate, k is the rate of mean reversion of the volatility

process, θ is the long-run variance and η is the volatility of volatility. Following the notation of the general class of stochastic volatility models,

$$X_t = Y(\sigma_t) = \sigma_t^2, \quad C(x) = k(\theta - x), \quad D(x) = \eta\sqrt{x} \quad (2.3.15)$$

According to Eq. 2.3.11, the stationary density of the variance σ_t^2 is gamma distributed:

$$\begin{aligned} \bar{p}(\sigma_t^2) &\propto (\eta\sigma_t)^{-2} \exp \left\{ 2 \int_{y_0}^{\sigma_t^2} \left(\frac{k\theta}{\eta^2 u} - \frac{k}{\eta^2} \right) du \right\} \\ &\propto \sigma_t^{4k\theta/\eta^2 - 2} e^{-2\sigma_t^2 k/\eta^2} \end{aligned} \quad (2.3.16)$$

Strictly speaking, for this class of models the risk-neutral measure is not unique, as the payoff involves an additional stochastic process (i.e. volatility process), while the number of underlying assets does not increase. For pricing purpose, an additional asset (i.e. volatility risk) needs to be included. Under the assumption that the market price of volatility risk is zero, the risk-neutral probability is uniquely determined. The forward price of a European call option under the Heston dynamics can be expressed by (Heston 1993)

$$c(x, v, \tau) = E^{\mathbb{Q}}[K \exp(x) - K] = K \exp(x) P_1(x, v, \tau) - K P_0(x, v, \tau), \quad (2.3.17)$$

where

$$x = \ln \left(\frac{S(t)}{B(t, T)K} \right) \quad (2.3.18)$$

is the logarithmic forward moneyness of the option and K is the strike price. P_0 is the probability of exercise under the risk neutral measure, while P_1 is the probability of exercise under the risk neutral measure associated with the numeraire $S(t)$. Following the solution presented by Gatheral (2006), the first probability is solved to be (Gatheral 2006)

$$P_0(x, v, \tau) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left[\frac{\exp\{C(u, \tau)\theta + D(u, \tau)v + iux\}}{iu} \right] du, \quad (2.3.19)$$

where

$$C(u, \tau) = k \left[r_- \tau - \frac{2}{\eta^2} \ln \left(\frac{r_+ - r_- e^{-d\tau}}{r_+ - r_-} \right) \right] \quad (2.3.20)$$

$$D(u, \tau) = r_- \frac{r_+ - r_- e^{-d\tau}}{r_+ - r_-} \quad (2.3.21)$$

$$r_\pm = \frac{\beta \pm d}{\eta^2} \quad d = \sqrt{\beta^2 - 2\alpha\eta^2} \quad (2.3.22)$$

Parameters α and β are functions of u (Fourier transform variable of x):

$$\alpha(u) = -\frac{u^2}{2} - i\frac{u}{2} \quad (2.3.23)$$

$$\beta(u) = k - \rho\eta iu \quad (2.3.24)$$

It is noted that $P_0 = E^{\mathbb{Q}}(\mathbb{1}_{S_T > K})$, since by definition P_0 is the probability of exercise. The probability density function of the risk-neutral measure is therefore obtained:

$$p(S_T) = - \left. \frac{\partial P_0}{\partial K} \right|_{K=S_T} = - \left. \frac{\partial P_0}{\partial x} \frac{\partial x}{\partial K} \right|_{K=S_T} \quad (2.3.25)$$

$$= \frac{1}{2\pi S_T} \int_{-\infty}^{\infty} \exp \left[C(u, \tau)\theta + D(u, \tau)v + iu \ln \left(\frac{S_t}{B(t, T)S_T} \right) \right] du \quad (2.3.26)$$

For simplicity e^y denotes the ratio of the forward price at t to its spot price at maturity T , i.e. $e^y = B(t, T)S_T/S_t$, we derive the risk-neutral probability density with respect to y :

$$p(y) = p(S_T) \left| \frac{\partial y}{\partial S_T} \right|^{-1} \quad (2.3.27)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp [C(u, \tau)\theta + D(u, \tau)v - iuy] du \quad (2.3.28)$$

2.4 Parametric Approach to Model Risk Measurement

2.4.1 Classification of Model Risk

Model risk refers to the financial risk incurred by the use of imperfect model assumptions, incorrect implementation of models or inaccurate model inputs. As any model is a simplified picture of the reality, model risk is unavoidable. However, a quantitative measure is required to understand the impact of this risk. As financial models usually deal with the stochastic nature of the market, their imperfections are related to the modelling of probabilities, or more explicitly, the conditional probabilities of state variables in the future.

The concept of model risk can be further illustrated by considering two option pricing models, the basic Black-Scholes model and the stochastic volatility Heston model. The Black-Scholes model describes the future stock price as a geometric Brownian motion with a constant volatility. When pricing an option using the Black-Scholes model, the volatility is the only parameter that allows adjustment. The lack of free parameters introduces a significant calibration error, for the model is incapable of fitting the prices of options with different strikes and maturities. This calibration error is referred as the Type 1 Model Risk (Schlögl 2015). The Heston model, on the other hand, has more adjustable parameters (long-run variance and volatility of volatility, etc.). This allows for the fitting of option prices with certain skewness and kurtosis, and therefore a reduction of calibration error.

Despite that the Heston model is capable of fitting to the option prices with better accuracy, it involves a larger parameter set. Calibration of these parameters may lead to the issue of "over-fitting". As a result, they are probably even more difficult to predict

from day to day. This means that the parameters calibrated from the current option prices are not necessarily good estimations of their future values. This “recalibration” error is defined as the Type 2 Model Risk.

Another type of model risk is associated with the modelled stochastic processes of the state variables. The Black-Scholes model assumes that the underlying price follows a geometric Brownian motion, and therefore a log-normal (risk-neutral) price distribution. The Heston model, on the other hand, assumes a geometric Brownian price process along with a mean-reverting stochastic volatility process. The pricing error caused by the difference between the actual stochastic processes and the modelled processes is defined as Type 3 Model Risk.

As mentioned before, adding more adjustable parameters (e.g. from the Black-Scholes model to the Heston model) shifts Type 1 model risk to Type 2 model risk. In a similar way, by changing a parameter to a state variable, we effectively shift Type 2 model risk to Type 3 model risk. Still taking the Black-Scholes model and the Heston model as an example, Heston model includes the volatility as a stochastic state variable, introducing an extra state variable and hence an additional level of Type 3 risk. However, compared to the Black-Scholes model, there is no need to recalibrate the volatility parameter to fit the changing market. The Type 2 model risk is thus shifted to the Type 3 risk. (Schlögl 2015)

2.4.2 Model Risk Quantification by Model Comparison

Model risk may rise on both competing model parameters and competing model types. For example, the calibration of option prices using the Black-Scholes model leaves a large amount of calibration error. A stochastic volatility model is a competing model type that performs better than the Black-Scholes model in terms of calibration error. However, in the Heston model there are more model parameters to calibrate and the calibrated values may have less predictive power. Different sets of parameter values provide room for model risk. In this section, we will focus on quantifying the model risk by comparing both different model parameters and different model types. Both approaches are regarded as parametric approaches to model risk measurement. Unless otherwise specified, an alternative model is referred as either a model of the same type but with different parameters or a model of different type.

A simple approach of accounting for model uncertainty is to assign weights to alternative models and then calculate the average market risk (Branger and Schlag 2004). There are two different methods of integrating market and model risk: model integration and risk integration. In model integration, one applies the model-average measure P' instead of the reference measure P by averaging

$$P' = \sum_{Q \in \mathcal{Q}} p(Q)Q$$

where \mathcal{Q} is a set of probability measures defined by possible models. $p(Q)$ is the weight of a particular measure Q . In risk integration, the weighted market risk measure is on

the other hand

$$\rho(X) = \sum_{\mathbb{Q} \in \mathcal{Q}} p(\mathbb{Q}) \phi(\rho_{\mathbb{Q}}(X))$$

where ϕ is an increasing and convex function (for risk-averse investors). Different from the model integration, risk integration calculates the market risk measure for each alternative model followed by the function ϕ adjusting for model uncertainty. The integration is on the risk measure level for a specific portfolio X , instead of the model level.

Perhaps a better way is to separate the model risk component from the market risk component. In addition, from the risk management point of view, one may be more interested in the worst-case scenario instead of the average scenario. A risk-differencing measure was proposed in the following way (Kerkhof, Schumacher and Melenberg 2002)

$$\mu(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \rho_{\mathbb{Q}}(X) - \rho_{\mathbb{P}}(X)$$

where ρ is a market risk measure, \mathbb{Q} is the reference model and \mathcal{Q} is the set of alternative models. Following the worst-case approach, Cont formulated a quantitative framework for measuring the model risk in derivative pricing (Cont 2006). This approach applies to a parametric set of alternative measures which price some benchmark instruments within their respective bid-ask spreads. The probability measures defined by the set of alternative models are not necessarily equivalent measures. For example, the model set can include diffusion processes with different volatilities, and jump and stochastic volatility models as well. In addition, this work proposed a coherent risk measure that provides numbers in price terms, instead of requiring any ad-hoc scaling factors (Artzner et al. 1999, Föllmer and Schied 2002). However, this approach is still parametric and relies on the specification of the set of alternative models. Therefore it is incapable of covering all probability measures that differ from the one defined by the reference model.

Formally, we have a set of arbitrage-free pricing models \mathcal{Q} satisfying

$$\forall \mathbb{Q} \in \mathcal{Q}, \forall i \in I, \mathbb{E}^{\mathbb{Q}}(|H_i|) < \infty, \mathbb{E}^{\mathbb{Q}}(H_i) \in [C_i^{bid}, C_i^{ask}]$$

where $(H_i)_{i \in I}$ are the discounted payoffs of a list of benchmark instruments indexed by i , whose market bid and ask prices are C_i^{bid} and C_i^{ask} . The set of contingent claims can be defined as those with a well defined price in all models

$$\mathcal{C} := \{H \in \mathcal{F}_T, \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}(|H|) < \infty\}$$

A proper risk measure that represents model risk is a mapping $\mu : \mathcal{C} \rightarrow [0, \infty)$ satisfying the following axioms (Cont 2006):

1. for (liquid) benchmark instruments, model uncertainty reduces to the uncertainty of the market values:

$$\forall i \in I \mu(H_i) \leq |C_i^{ask} - C_i^{bid}|$$

2. Convexity

$$\forall X_1, X_2 \in \mathcal{C}, \forall \lambda \in [0, 1] \mu(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda \mu(X_1) + (1 - \lambda)\mu(X_2)$$

3. if an instrument X can be partially hedged by the underlying S_t in a model free way, then the risk measure of X should generate smaller number as its risk can be reduced by constructing a dynamic hedging strategy. Therefore, for every hedging strategy ϕ we have

$$\mu(X) = \mu\left(X + \int_0^T \phi_t dS_t\right)$$

4. Apart from hedging with the underlying, we may also hedge an instrument X using a combination of benchmark options. The reasoning above still applies

$$\forall X \in \mathcal{C}, \forall u \in \mathbb{R}^K \mu\left(X + \sum_{i=1}^k u_i H_i\right) \leq \mu(X) + \sum_{i=1}^k |u_i (C_i^{ask} - C_i^{bid})|$$

A coherent measure of model uncertainty that follows the above axioms can be constructed by the map $X \in \mathcal{C} \mapsto \bar{\pi}(-X)$ where

$$\bar{\pi}(X) := \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}(X)$$

is the upper price bound under all alternative models in the set \mathcal{Q} .

Following Cont's work, Gupta, Reisinger and Whitley (2010) proposed the definition of the spread of a claim $X \in \mathcal{C}$ to be the set of the prices given by all models $\mathbb{Q} \in \mathcal{Q}$. They then imposed monotonicity, subadditivity and homogeneity to the definition of a coherent risk measure (for model risk):

1. Monotonicity: if the spread of prices for Y is greater than that for X , then the risk measure for Y should be greater than that for X
2. Subadditivity: the risk measure of a combined claim X and Y should be less than or equal to the sum of their individual risk measures
3. Homogeneity: the risk measure should scale linearly with the number of claims

An example of such a coherent measure is given by

$$\mu(X) = \sum_{\mathbb{Q} \in \mathcal{Q}} p(\mathbb{Q}) |\mathbb{E}^{\mathbb{Q}}(X) - M(\mathbb{E}^{\mathbb{Q}}(X))|$$

where $p(\mathbb{Q})$ is the weight assigned to the alternative model \mathbb{Q} and M denotes the median value.

Bannör and Scherer (2013) proposed a parametric risk framework that unifies the proposals of Cont (2006), Gupta et al. (2010) and Lindström (2010). This approach

incorporates a distribution of parameter values to capture the risk of parameter uncertainty. They propose to use risk-capturing functionals to determine parameter risk adjusted prices. Such risk adjusted prices could give out the bid-ask spread in instruments that face parameter risk.

The model risk-capturing functional is defined as follows. First we define a normalised risk measure ρ if $\rho(0) = 0$ holds. Let $\mathcal{Q} \in \mathcal{Q}$ be an alternative model in the set of all legitimate models. Let $\mathcal{A} \subset \mathcal{L}^0(\mathcal{Q})$ be a vector space of measurable functions containing the constants and denote $\mathcal{C}^{\mathcal{A}}$ as the vector space of all \mathcal{A} -regular claims being available (i.e. absolutely integrable) for all models in \mathcal{Q} . If $\rho : \mathcal{A} \rightarrow \mathbb{R}$ is a normalised convex risk measure, then the mapping $\gamma : \mathcal{C}^{\mathcal{A}} \rightarrow \mathbb{R}$ defined by

$$\Gamma(X) := \rho(\mathcal{Q} \mapsto \mathbb{E}^{\mathcal{Q}}(X))$$

is called a model risk-capturing functional.

Bannör and Scherer (2013) applied the approach of model risk-capturing functional to explain the bid-ask spread as a safety margin for parameter uncertainty. They define a special type of risk-capturing functional by

$$\rho_{\lambda}^{ent}(X) := \frac{1}{\lambda} \ln(\mathbb{E}^P(e^{\lambda X}))$$

called the entropic-induced risk-capturing functional. For a particular problem, they estimated the model risk on pricing an European exchange option written on two correlated assets. Following a historical distribution of the correlation parameter, they applied the entropic-induced risk-capturing functional to calculate the risk measures for bid and ask prices.

Detering and Packham (2016) approach the problem of model risk measurement based on the residual profit and loss from hedging in the reference model. Given the underlying S and a set of benchmark options with payoffs $(H_i)_{i \in I}$, a trading strategy is defined as a predictable process $\Phi = (\phi^0, \dots, \phi^d, u_1, \dots, u_I)$, where $\phi^j = (\phi_t^j)_{t \geq 0}$ is the holdings in the underlying asset j and u_i is the static holding of the benchmark option i . The value of the trading portfolio is

$$V_t(\Phi) = \sum_{j=0}^d \phi_t^j S_t^j + \sum_{i=1}^I u_i H_t^i$$

This definition of a trading strategy combines dynamic hedging using the underlying assets and static hedging using derivative instruments. The self-financing condition is given by

$$dV_t(\Phi) = \sum_{j=0}^d \phi_t^j dS_t^j + \sum_{i=1}^I u_i dH_t^i$$

For a contingent claim X , the loss at time t (before the terminal time T) is

$$L_t(X, \Phi) = - \left(\mathbb{E}(Y) + \int_0^t \phi(Y) dS - \mathbb{E}(Y | \mathcal{F}_t) \right)$$

where $\Phi = ((\phi(Y))_{t \in [0, T]}, u_1, \dots, u_I)$ and $Y = X - \sum_{i=1}^I u_i H_i$. A model-free or model-independent hedging strategy is a strategy Φ that has $L_t(X, \Phi) = 0$, $\forall t \in [0, T]$, \mathbb{Q} -a.s., for all $\mathbb{Q} \in \mathcal{Q}$. Incorporating the reduction of uncertainty under dynamic and static hedging, the usual risk measures, such as VaR and ES, may be re-defined as

$$\mu_{VaR, \alpha, t}^{\mathbb{Q}}(X) = \inf_{\Phi \in \Pi(X)} VaR_{\alpha}(|L_t(X, \Phi)|) \quad \mu_{ES, \alpha, t}^{\mathbb{Q}}(X) = \inf_{\Phi \in \Pi(X)} ES_{\alpha}(|L_t(X, \Phi)|)$$

To account for model uncertainty, we measure the worst-case risk measure by

$$\mu_{\alpha, t}(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mu_{\alpha, t}^{\mathbb{Q}}(X)$$

Kerkhof, Melenberg and Schumacher (2010) propose a procedure to take model risk into account when computing capital reserves. Instead of formulating model risk in terms of a collection of probability measures, they consider the reality that practitioners may evaluate risk based on models of different natures. For example, models based on a finite number of economic scenarios may be used alongside continuous-time and discrete-time dynamic models. Risk is defined as a mapping Π that assigns each model $\mathbb{Q} \in \mathcal{Q}$ a random variable. The set of all such mappings is denoted by $\mathcal{X}(\mathcal{Q})$. Standard risk measures such as VaR is redefined as a mapping that assigns a model \mathbb{Q} and a random variable X by

$$VaR_p(\mathbb{Q}, X) = -\inf\{x \in \mathbb{R} | \mathbb{Q}(X \leq x) \geq p\}$$

Then they define a multimodel risk measure by a mapping that assigns to each element of $\mathcal{X}(\mathcal{Q})$ a function from \mathcal{Q} to \mathbb{R} , defined by

$$(\rho(\Phi))(\mathbb{Q}) := \rho(\mathbb{Q}, \Pi(\mathbb{Q}))$$

The multimodel risk measure defined above is a function on $\mathcal{X}(\mathcal{Q})$. To get a single number representing the overall risk, we calculate the worst-case risk by

$$RISK_{\rho, \mathcal{Q}} := \sup_{\mathbb{Q} \in \mathcal{Q}} \rho(\mathbb{Q}, \Pi(\mathbb{Q}))$$

This overall risk can be considered as a model risk component added on top of the usual market risk. Given a nominal model \mathbb{P} and a tolerance model set \mathcal{Q} with $\mathbb{P} \in \mathcal{Q}$, we define the model risk by

$$\phi_{\rho}(\Pi, \mathbb{P}, \mathcal{Q}) := RISK_{\rho, \mathcal{Q}}(\Pi) - RISK_{\rho, \mathbb{P}}(\Pi)$$

Kerkhof et al. suggest using past sequential data to assign weights to the possible probability measures in \mathcal{Q} (Kerkhof et al. 2010). They define the mapping from \mathcal{Q} to the set of state processes by $d : \mathbb{Q} \mapsto d(\mathbb{Q})$. The mapping from observable characteristics to state processes is given by $\chi : \xi \rightarrow d(\mathcal{Q})$. The estimation risk (arise from parameter uncertainty) is given by

$$\phi_{\rho}(\Pi, d^{-1}(\chi(e)), d^{-1}(\chi(\hat{E})))$$

where \hat{E} is a confidence interval of parameter values. Model misspecification risk (or calibration risk) is defined similarly by extending the model class $d^{-1}(\chi(\hat{E}))$ to include other types of models. The more interesting part of their work is the inclusion of identification risk which is the same as the Type 3 risk. Past data may not be able to identify specific models as it is not sufficient or some factors have been constant in the data period. In derivative pricing, identical performance may be observed for models with very different dynamics, say a diffusion process and a jump-diffusion process. Risk management is all about the future, so we need to account for the worst case of the models by the measure of identification risk

$$\phi_\rho(\Pi, d_1^{-1}(\chi_1(\hat{E}_1)), d_2^{-1}(\chi_2(\hat{E}_2)))$$

where the estimated set \hat{E}_2 is constructed such that it contains \hat{E}_1 and characteristics that are not distinguishable from those in \hat{E}_1 on the basis of the data. χ_2 and d_2 are the corresponding data characterising mapping and data generating mapping. Typically, \hat{E}_1 refers to backtesting characteristics and \hat{E}_2 arises from stress testing. Therefore this approach combines model with different types together.

From a practical point of view, Boucher, Danielsson, Kouontchou and Maillet (2014) proposed an approach that incorporates model risk into the usual market risk measures. They consider the true VaR as being biased from the VaR estimated from past data:

$$ThVaR(\theta_0, \alpha) = EVaR(\hat{\theta}, \alpha) + bias(\theta_0, \hat{\theta}, \alpha)$$

where $ThVaR$ and $EVaR$ denote the true VaR and the estimated VaR respectively. θ_0 and $\hat{\theta}$ are the true parameters and the estimated parameters. α refers to the confidence level of the VaR. Change of probability law is then taken into account by defining new confidence levels α^* that adjusts for such bias. The authors then adopted a general data generating process to compute sample paths. Under a limited number of sample paths, the estimated VaR differs from the true VaR. The magnitude of such bias is measured by the shift of the confidence level $\alpha^* - \alpha$.

2.4.3 Practical Model Risk Management in Derivative Pricing

Morini (2011) has summarized the model risk involved in choosing an option pricing model. He approached the problem by comparing two general classes of option pricing models, stochastic volatility models (SVM) and local volatility models (LVM). Both classes of models generalise the Black-Scholes model and allow for calibration of the real implied volatility surface.

A stochastic volatility model has two state variables, the underlying asset price and the volatility. They follow two stochastic processes with some degree of correlation. In a stochastic volatility model, volatility clustering is successfully generated due to the autocorrelation of the volatility process. This essentially allows for heavy kurtosis (fat tails). If we set a non-zero correlation between price shock and volatility shock, we are also able to introduce the leverage effect to the implied volatility curve. This means that

the implied volatility is generally lower at higher strikes than at lower strikes, which is a common feature in equity option market.

A local volatility model, on the other hand, describes the volatility as a deterministic function of the price of the underlying asset and the time to maturity. Unlike the stochastic volatility model, the local volatility model has only one state variable, the underlying asset price, which follows

$$dS_t = rS_t dt + \sigma_{LV}(S_t, \tau) S_t dW_t \quad (2.4.1)$$

A general local volatility model represents the local volatility as an arbitrary bivariate function, which allows for calibration of any kind of implied volatility surface. This is a main advantage of LVM over SVM, and therefore LVM is widely adopted among option practitioners.

Despite of the flexibility of a general LVM, it requires additional information in practice due to the discreteness of available maturities and strikes. In practice, arbitrary interpolations and extrapolations both in strike space and in time are widely used to parametrise the local volatility surface. Such interpolations and extrapolations unavoidably introduce a high level of model uncertainty, which potentially has a high level of model risk. A more proper way may be to narrow down the general LVM to a specific LVM with a finite number of parameters.

Another problem with the LVM is that it restricts the type of dynamics that result in a certain continuum of distributions. Supposing we now have a complete set of maturities and strikes, we are thus capable of obtaining the risk-neutral density at any future time. However, there are still an infinite number of dynamics that can result in the specified marginal distributions; the one calculated by the Dupire formula is merely the unique (risk-neutral) diffusive process among all these dynamics (Morini 2011).

A general LVM can fit the market option prices perfectly, which indicates that it is able to reproduce the marginal distributions correctly. However, the diffusive dynamics involved in the LVM can still deviate from the "true" model. Thus model risk is introduced when using the calibrated local volatility surface to price exotic derivatives. SVM, on the other hand, may obtain a good but not perfect fit to the market option prices. However, as SVM introduces an additional state variable, it has dynamics more general than those covered by the local volatility models.

Hull and Suo (2002) tested the pricing powers of well calibrated LVM and SVM. Two exotic derivatives were considered, including a call-on-call compound option and barrier option. In their example, the two models give exactly the same marginal distributions, and the only difference lies in their transition densities. The simulation has shown that in call-on-call compound option the difference between the LVM price and the SVM price is below 2% of the SVM price. However, in the barrier option case the difference can be as large as 50%. The large model risk in the barrier option pricing is due to its higher and more complex dependence on transition densities. This thesis has formulated the entropic approach for derivatives that are dependent on transition densities (see Chapter 4). This allows calculation of the model risk due to different dynamics.

2.4.4 Practical Model Risk Management in Hedging

Because we are not sure about the “true” dynamics of the underlying asset price (and volatility), we are not able to obtain a true perfect hedge. Consider the delta hedge based on the Black-Scholes model, by using the Black-Scholes model we are assuming that the volatility parameter should never be recalibrated. In practice, however, the volatility is constantly changing, and has some dependence with the underlying asset price. If we merely hedge our position based on the Black-Scholes model, we are left with an unhedged residual risk. Therefore, we are required to determine the risk on P&L when hedging the position based on a specified stochastic model.

In option markets with a volatility skew, there is an additional delta termed as shadow-delta or skew-delta, originated by the possible correlation between the implied volatility and the underlying asset price. Hagan, Kumar, Lesniewski and Woodward (2002) showed that the implied volatility curve shifts in the same direction with the movement of the underlying asset price. This appears to be a usual pattern in many markets, and is termed as the comonotonic behaviour. A model that generates this behaviour can have reduced hedging risk and possibly less frequent rebalance.

Hagan et al. (2002) showed that the comonotonic behaviour is not consistent with a local volatility model with the volatility being a univariate function of the underlying asset price. This type of local volatility model has dynamics of

$$dS_t = \sigma_{LV}(S_t)S_t dW_t \quad (2.4.2)$$

which leads to an implied volatility curve of (Hagan et al. 2002)

$$\sigma_{S_t}(K) = \sigma_{LV} \left(\frac{S_t + K}{2} \right) \left[1 + \frac{1}{24} \frac{\sigma_{LV}'' \left(\frac{S_t + K}{2} \right)}{\sigma_{LV} \left(\frac{S_t + K}{2} \right)} (S_t - K)^2 + \dots \right] \quad (2.4.3)$$

If neglecting the higher order terms, we have implied volatility to be approximately equal to the local volatility at the middle point between the strike price and the current underlying asset price:

$$\sigma_{S_t}(K) \approx \sigma_{LV} \left(\frac{S_t + K}{2} \right) \quad (2.4.4)$$

This local volatility model generates exactly the opposite behaviour: when the underlying asset price changes, the implied volatility curve moves towards the opposite direction. This causes undesirable consequences in hedging. In fact, the hedging ratio (delta) calculated from this type of local volatility models is even worse than that calculated from the Black-Scholes model.

To resolve the hedging problem, Hagan et al. (2002) suggested the SABR model, which mixes stochastic volatility and the CEV (constant elasticity of variance) local volatility model. In the SABR model the forward price has the following dynamics:

$$\begin{aligned} dF_t &= V_t F_t^\beta dW_t \\ dV_t &= \epsilon V_t \left(\rho dW + \sqrt{1 - \rho^2} dW'_t \right) \end{aligned} \quad (2.4.5)$$

where β exponent makes the dynamics analogous to the CEV model, while perturbed by the log-normal stochastic volatility process V_t . W_t and W'_t are uncorrelated Brownian motions. The constant ϵ is the volatility of volatility.

Using the singular perturbation method, Hagan et al. (2002) claim that the implied volatility curve can exhibit the comonotonic behaviour. However, Morini (2011) tested the model and concluded that SABR model still generates non-comonotonic behaviours if taking into account the negative correlation between underlying and volatility. Therefore, unless hedging with a model implementation that is not fully consistent with the model assumptions, both models are not capable of reproducing consistent monotonic behaviours. Morini (2011) thus suggests to use empirical ways to determine the hedging ratio, by constantly recalibrating the model. In such an approach the change of the underlying asset price is not treated as an endogenous product of the model, but as an exogenous trigger for model recalibration.

2.5 Non-parametric Approach to Model Risk Measurement

2.5.1 Entropic Model Risk Measurement

Glasserman and Xu (2014) recently developed a theoretical framework to quantify the model risk by measuring the relative entropy between models. By constraining the relative entropy budget, we are able to derive the worst-case model and its associated risk. This approach is not model specific, for it is not merely varying the parameter values of a model. Instead, it evaluates the impact of all possible models, constrained by a permitted level of pseudo-distance (e.g. relative entropy) from the reference model.

The logic underlying this approach originates from information theory. In Bayesian statistics, the relative entropy between the prior and the posterior distributions measures the additional information gained through new observations. Applying this to model risk, the relative entropy measures the amount of information required for the “true” model to be preferable to the reference model. The larger the relative entropy is, the more information is required to justify the deviation from the reference model.

The model risk measured by this approach is regarded as being robust, as it deals with the non-parametric worst-case scenario. In fact, robust optimisation seeks to optimise against the worst-case risks. In this approach, an imaginary adversary is assumed to change the probability law arbitrary constrained by the relative entropy budget. The purpose of robust risk measurement is to estimate the worst possible risk (Glasserman and Xu 2014). Compared to this approach, alternative methods, such as mixing multiple models (see Sec. 2.4), are not as robust, as they heavily depend on the selection of candidate models. It is probable that the worst-case risk is not included in any of the models and the actual risk is significantly underestimated.

It is noted that this approach of measuring model risk is closely linked to the entropic value at risk (see Sec. 2.1.1). Apart from the conceptual difference that EVaR is a measure of market tail loss, the Glasserman approach measures the relative entropy

for the distribution of a set of “state variables”, while EVaR is for the distribution of the profit & loss. Therefore, considering the profit & loss distribution as one of the many branches rooted from a model, the Glasserman approach deals with alternative models directly at the root level while EVaR is at the branch level. For instance, consider measuring the model risk of using the Black-Scholes option pricing model. In the Glasserman approach, the set of alternative models is given based on the relative entropy of the distribution of the underlying asset price. In the EVaR calculation, on the other hand, it is given by the relative entropy of the price distribution of the option we hold. In fact, the alternative models evaluated by EVaR depend not only on the reference model, but also on the portfolio composition. In this sense, the relative entropy calculated for state variables is a consistent and more appropriate measure of model risk.

Technically, the approach is based on the maximising of risks under alternative probability measures. Each alternative measure is characterised by its likelihood ratio (or Radon-Nikodym derivative) with respect to the reference measure. Our purpose is to find the one Radon-Nikodym derivative that maximises the risk, under the constraint of the relative entropy budget. By its nature of constrained optimisation, this problem is solved with a Lagrange multiplier approach. The worst-case risk subject to a relative entropy budget of η is

$$\sup_{m \in \mathcal{P}_\eta} E^{\tilde{P}} [V(X)] \quad (2.5.1)$$

where m is the Radon-Nikodym derivative of an alternative measure \tilde{P} , with respect to the reference model P :

$$m = \frac{d\tilde{P}}{dP} \quad (2.5.2)$$

\mathcal{P}_η is the set of Radon-Nikodym derivatives, whose corresponding probability measures are constrained by the relative entropy budget η with respect to the reference measure:

$$\mathcal{P}_\eta = \left\{ \frac{d\tilde{P}}{dP} : E^{\tilde{P}} \left(\ln \frac{d\tilde{P}}{dP} \right) \leq \eta \right\} \quad (2.5.3)$$

Applying a Lagrange multiplier $1/\theta$, it is transformed to a dual problem:

$$\inf_{\theta > 0} \sup_m E^P \left[mV(X) - \frac{1}{\theta} (m \ln m - \eta) \right] \quad (2.5.4)$$

where the inner supremum problem has a close-form solution:

$$m_\theta^* = \frac{\exp[\theta V(X)]}{E^P[\exp[\theta V(X)]]} \quad (2.5.5)$$

The Lagrange multiplier, θ , can be expressed by a function of the market model risk (measured by relative entropy) η :

$$\theta(\eta) = \arg \inf_{\theta > 0} \frac{\ln E(e^{\theta V})}{\theta} + \frac{\eta}{\theta}. \quad (2.5.6)$$

Alternatively, the relative entropy budget η is solved to be a function of the multiplier θ ,

$$\begin{aligned}\eta(\theta) &= \frac{E(\theta V e^{\theta V})}{E(e^{\theta V})} - \ln E(e^{\theta V}) \\ &= \theta \kappa'(\theta) - \kappa(\theta)\end{aligned}\tag{2.5.7}$$

where $\kappa(\theta) = \ln E(\exp(\theta V))$ is the cumulant generating function of $V(X)$. In practice, the function $\eta(\theta)$ is first evaluated numerically according to Eq. 2.5.7. This is followed by searching for a θ value where $\eta(\theta)$ gives the market level of relative entropy budget.

2.5.2 Minimisation of Model Risk under Constraints

The entropic model risk measure naturally leads to model risk management. By making investment decisions with considerations on the model risk, we are capable of minimising the risk under alternative models. We will illustrate this point by taking portfolio optimisation as an example.

In portfolio optimisation, we have the freedom of choosing weights allocated to different assets. To minimise the model risk, we need to solve the following optimisation problem with respect to the asset weight a :

$$\inf_a \sup_{m \in \mathcal{P}_\eta} E[mV_a(X)]\tag{2.5.8}$$

By introducing a Lagrange multiplier θ^{-1} , the constraint optimisation problem is converted to

$$\inf_{\theta > 0} \inf_a \sup_m E[mV_a(X) - \theta^{-1}(m \ln m - \eta)]\tag{2.5.9}$$

By solving the inner problem, we immediately obtain the optimal decision $a^*(\theta)$:

$$a^*(\theta) = \arg \inf_a \theta^{-1} \ln E[\exp(\theta V_a(X))]\tag{2.5.10}$$

and under the optimal decision, the worst-case change of measure is

$$m_\theta^* = \frac{\exp(\theta V_{a^*(\theta)}(X))}{E[\exp(\theta V_{a^*(\theta)}(X))]\tag{2.5.11}$$

In addition to the constraint of the relative entropy budget, we could further constrain the set of alternative measures by incorporating additional information. For example, consider measuring the model risk that affects the price of an option. We could calculate the worst-case risk-neutral measure using the entropic approach, assuming all measures satisfying the relative entropy constraint are feasible. However, in practice we might want to consider the market prices of other related instruments, such as the price of the underlying asset. If we believe the underlying asset is fairly priced, to avoid arbitrage we would like to find the worst-case measure constrained by

the fair price of the underlying asset. Therefore we need to study the worst-case risk under certain expectation constraints.

Here we consider the general case with additional constraints taking the form of $E[mh_i(a, X)] \leq \eta_i$ for some functions h_i and constants η_i . We can convert the constraints to penalty functions by introducing extra Lagrange multipliers, λ_i per constraint (h_i, η_i) :

$$\inf_{\theta>0, \lambda_i>0} \inf_a \sup_m E \left[mV_a(x) - \frac{1}{\theta}(m \ln m - \eta) - \sum_{i=1}^{n_M} \lambda_i [mh_i(a, X) - \eta_i] \right] \quad (2.5.12)$$

Due to the extra constraints, instead of Eq. 2.5.11 the worst-case measure m_θ^* is now expressed by

$$m_\theta^*(a, \theta, \lambda_i) \propto \exp \left(\theta \left[V_a(X) - \sum_{i=1}^{n_M} \lambda_i h_i(a, X) \right] \right) \quad (2.5.13)$$

This is followed by calculating the optimal decision a^* :

$$a^*(\theta, \lambda_i) = \arg \inf_a \frac{1}{\theta} \ln E \left[\exp \left(\theta \left[V_a(X) - \sum_{i=1}^{n_M} \lambda_i h_i(a, X) \right] \right) \right] \quad (2.5.14)$$

The Lagrange multipliers θ and λ_i are solved by optimising

$$\inf_{\theta>0, \lambda_i>0} \frac{1}{\theta} \ln E \left[\exp \left(\theta \left[V_{a^*(\theta, \lambda_i)}(X) - \sum_{i=1}^{n_M} \lambda_i h_i(a^*(\theta, \lambda_i), X) \right] \right) \right] + \frac{\eta}{\theta} + \sum_{i=1}^{n_M} \eta_i \lambda_i \quad (2.5.15)$$

2.5.3 α -Divergence as a Generalised Relative Entropy Measure

Glasserman and Xu (2014) pointed out that the standard relative entropy measure only works when the tail of the risk distribution $V(X)$ is exponentially bounded. To deal with heavy-tailed distributions, a more general set of relative entropy called α -divergence is introduced:

$$D_\alpha(m) = \frac{1 - E(m^\alpha)}{\alpha(1 - \alpha)} \quad (2.5.16)$$

where m is the Radon-Nikodym derivative and $\alpha \in (0, 1)$. The standard relative entropy (Kullback-Leibler divergence) $D_{KL}(m) = \lim_{\alpha \rightarrow 1^+} D_\alpha(m)$. Clearly, α -divergence is a set of generalised relative entropy measures defined in Eq. 2.1.14.

In the general setting of α -divergence measure, the constrained optimisation problem Eq. 2.5.9 is generalised to

$$\inf_{\theta>0} \inf_a \sup_m E \left[mV_a(X) - \theta^{-1}(D_\alpha(m) - \eta) \right] \quad (2.5.17)$$

The pair $(m^*(\theta, \alpha, a), c(\theta, \alpha, a))$ that solves the following equation with probability 1 is an optimal solution to the inner problem of Eq. 2.5.17:

$$m^*(\theta, \alpha, a) = [\theta(\alpha - 1)V_a(X) + c(\theta, \alpha, a)]^{\frac{1}{\alpha-1}} \quad (2.5.18)$$

for some constant $c(\theta, \alpha, a)$ such as $m^*(\theta, \alpha, a) \leq 0$ and $E(m^*(\theta, \alpha, a)) = 1$. Eq. 2.5.18 can be derived by taking the functional derivative of (the inner function of) Eq. 2.5.17 with respect to m :

$$\delta E \left[mV_a(X) - \frac{1}{\theta}(D_\alpha(m) - \eta) - c(m - 1) \right] = E \left[\left(V_a(X) + \frac{\theta^{-1}}{1 - \alpha} m^{\alpha-1} - c \right) \delta m \right] \quad (2.5.19)$$

It is noted that c is the Lagrange multiplier introduced so that m is restricted to be a feasible Radon-Nikodym derivative, i.e. $E(m) = 1$. Eq. 2.5.18 solves Eq. 2.5.19 and thus is the optimal Radon-Nikodym derivative. The optimal decision under the α -divergence measure is

$$a^*(\theta) = \arg \min_a \left\{ \frac{\alpha - 1}{\alpha} E[m^*(\theta, \alpha, a)V_a(X)] + \frac{c(\theta, \alpha, a)}{\theta\alpha(1 - \alpha)} \right\} \quad (2.5.20)$$

It is noted that α -divergence risk measure is a special case of the f -entropic risk measure defined in Eq. 2.1.15. As a result, it is also a measure of the tail loss. Clearly, the worst-case measure with m^* is the one that maximise the tail loss under the relative entropy constraint. If we assume the risk function has a polynomial form with respect to the state variable x , i.e. $V_a(x) \approx x^k$, then the worst-case measure gives a density function of

$$\tilde{f}_X(x) \approx cx^{\frac{k}{\alpha-1}} f_X(x), \quad x \gg 0 \quad (2.5.21)$$

for $\alpha > 1$, where $f_X(x)$ is the probability density function under the reference measure. This illustrates the point that the worst-case measure makes the tail heavier, asymptotically, by a factor of $x^{k/(\alpha-1)}$. This incurs a shift of the mean of the state variable, and hence a higher expected value of the risk function. This is consistent with EVaR, which measures the tail loss essentially by measuring the risk under the worst-case measure.

2.5.4 Applications of Entropic Model Risk Measure

Portfolio Variance

The return of a portfolio can be expressed by aX , where a is the vector of portfolio weights and X denotes the vector of asset returns. The portfolio variance is expressed by

$$E[a^T(X - \mu)(X - \mu)^T a] := E[V(X)] \quad (2.5.22)$$

where μ denotes the vector of the means of the asset returns. The worst-case variance under the change of measure m can be expressed as

$$\sup_{m \in \mathcal{P}_\eta} E[mV(x)] = \sup_{m \in \mathcal{P}_\eta} E [ma^T(X - \mu)(X - \mu)^T a] \quad (2.5.23)$$

If the reference model says the asset return follows a multivariate normal distribution $X \sim \mathcal{N}(\mu, \Sigma)$, then the worst-case measure has a density proportional to

$$\exp [\theta a^T(x - \mu)(x - \mu)^T a] \times \exp \left[-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \right] \quad (2.5.24)$$

This density still belongs to a multivariate normal distribution, $\mathcal{N}(\mu, \tilde{\Sigma})$, but with a different covariance matrix $\tilde{\Sigma}$: (Glasserman and Xu 2014)

$$\tilde{\Sigma} = (\Sigma^{-1} - 2\theta aa^T)^{-1} \quad (2.5.25)$$

The worst-case portfolio variance is thus equal to

$$a^T \tilde{\Sigma} a = a^T \Sigma a + 2\theta (a^T \Sigma a)^2 + O(\theta^2) \quad (2.5.26)$$

The worst-case variance is increased by a quadratic term of the nominal portfolio variance, $a^T \Sigma a$.

Model Risk on Market Risk Measures

The market risk measures (VaR or ES) are tail loss measures that model the future change of portfolio value via some parametric or non-parametric approaches. There is an extra level of risk associated with the model adopted in the calculation. Here we quantify the model risk associated with the VaR and ES risk measures (Rockafellar and Uryasev 2002):

$$\begin{aligned} VaR_\alpha &= \arg \inf_b \frac{1}{1 - \alpha} E[(X - b)^+] + b \\ ES_\alpha &= \inf_b \frac{1}{1 - \alpha} E[(X - b)^+] + b \end{aligned} \quad (2.5.27)$$

where α is the confidence level of the risk measures, and X is the random variable of possible losses.

Eq. 2.5.27 has a simple explanation, as the derivative of $-E[(X - a)^+]$ with respect to a gives the cumulative distribution function (cdf) of X . Here we do not intend to prove this rigorously. One can imagine $E[(X - a)^+]$ as the call option price at a strike a (assuming zero interest rate). It is well known that the second derivative of the option price with respect to strike gives the probability density function (pdf). As a result, the first derivative of $-E[(X - a)^+]$ with respect to a provides the cdf. Therefore, a valued at the minimum of Eq. 2.5.27 is the quantile that gives a cdf value of $1 - \alpha$. This is

consistent with the original definition of VaR_α . The expression for ES, with a replaced by VaR , is a direct result from the definition of ES (Eq. 2.1.4).

Suppose X follows a double exponential distribution with a location parameter μ and a scale parameter b , i.e.

$$f(x) \propto \exp\left(-\frac{|x - \mu|}{b}\right) \quad (2.5.28)$$

Under this reference measure, VaR and ES are calculated as

$$\begin{aligned} VaR_\alpha &= \mu - b \ln[2(1 - \alpha)] \\ ES_\alpha &= VaR_\alpha + b \end{aligned} \quad (2.5.29)$$

The worst-case measure (which maximises ES) is calculated by multiplying $f(x)$ with m^* , which is given by Eq. 2.5.5,

$$\tilde{f}(x) = m^*(x)f(x) \propto \exp\left(-\frac{|x - \mu|}{b} + \frac{\theta}{1 - \alpha}(x - a)^+\right) \quad (2.5.30)$$

The VaR and ES under the worst-case measure are calculated according to Eq. 2.5.10:

$$\begin{aligned} VaR_{\alpha,\theta} &= \mu - b \ln\left[\frac{2(1 - \alpha - \theta b)}{1 - \theta b}\right] \\ CVaR_{\alpha,\theta} &= VaR_{\alpha,\theta} + \left(\frac{1}{b} + \frac{\theta}{1 - \alpha}\right)^{-1} \end{aligned} \quad (2.5.31)$$

Comparing Eq. 2.5.31 with Eq. 2.5.27, we can see that the worst-case VaR and ES are larger than the risk measures evaluated under the reference measure.

Portfolio Credit Risk

We apply the entropic model risk measure to the portfolio credit risk measurement. In this problem, we deal with the total loss that is the sum of individual defaults:

$$L = \sum_{i=1}^n c_i Y_i \quad (2.5.32)$$

where Y_i takes the value of 0 (if not default) or 1 (if default). Our aim is to measure the tail probability $P(L > x)$ for a given loss threshold x .

In the Gaussian copula model, each default indicator Y_i is represented by the event $\{X_i > x_i\}$, where X_i has a standard normal distribution. x_i is chosen to match the marginal distribution, i.e. $P(X_i > x_i) = p_i$, where p_i is the estimated probability of default. Dependence between default indicators are given by the correlations between normally distributed X_i . To simplify the complete correlation matrix to a single-factor problem, we consider a homogeneous model:

$$X_i = \rho Z + \sqrt{1 - \rho^2} \epsilon_i \quad (2.5.33)$$

where $Z \sim N(0, 1)$ indicates the systematic risk, and $\epsilon_i \sim N(0, 1)$ is the idiosyncratic risk associated with the i -th obligator. They are independent of each other.

The risk measure is

$$P(L > x) = E[I_{L>x}] \quad (2.5.34)$$

which results in the worst-case change of measure:

$$m_\theta^* = \frac{\exp(\theta I_{L>x})}{E[\exp(\theta I_{L>x})]} \quad (2.5.35)$$

This means that under the worst-case measure, the probability of any outcome is increased by $\exp(\theta)/C$ if it yields a loss greater than x . Otherwise, the probability is lowered by a factor of C , where C is a normalisation constant.

Delta Hedging Error

Compared to the previous applications, measuring the model risk in delta hedging is more challenging as it involves the entire dynamics of the underlying asset, rather than a simple marginal distribution. This requires a high-dimensional evaluation of the worst-case measure. Glasserman and Xu (2014) dealt with the problem using Monte Carlo simulation.

In the Monte Carlo framework, a large number of paths are simulated based on the reference model. For each simulated path, the delta hedging error is calculated with a discrete-time implementation. The initial values of cash and stock in the portfolio are given by

$$\begin{aligned} \text{cash}(0) &= C(0, T, S_0) - S_0 \delta(0, S_0) \\ \text{stock}(0) &= S_0 \delta(0, S_0) \end{aligned} \quad (2.5.36)$$

where $\delta(t, S_t)$ denotes the number of shares of stock held at time t , determined by the delta value given by the reference model. After rebalancing at time $k\Delta t$, the values are given by

$$\begin{aligned} \text{cash}(k\Delta t) &= e^{r_n \Delta t} \text{cash}(k-1) - S_{k\Delta t} [\delta(k\Delta t, S_{k\Delta t}) - \delta((k-1)\Delta t, S_{(k-1)\Delta t})] \\ \text{stock}(k\Delta t) &= S_{k\Delta t} \delta(k\Delta t, S_{k\Delta t}) \end{aligned} \quad (2.5.37)$$

The total hedging error until the maturity of the option is calculated for each path by

$$V(X) = |(S_T - K)^+ - \text{cash}(T) - \text{stock}(T)| \quad (2.5.38)$$

After calculating the hedging error for each path, the worst-case change of measure is obtained by assigning the following weight to each path:

$$m_\theta^*(\text{path}_i) = \frac{\exp[\theta V(\text{path}_i)]}{\sum_{j=1}^N \exp[\theta V(\text{path}_j)]/N} \quad (2.5.39)$$

where $V(\text{path}_i)$ is the hedging error of a sampled path i , $i = 1, 2, \dots, N$. The expected hedging error under the worst-case measure can then be calculated as a weighted average. Alternatively, one can use an acceptance-rejection scheme, that is to accept a sampled path according to the probability

$$P_\theta(\text{path}_i) = \frac{m_\theta^*(\text{path}_i)}{M} \quad (2.5.40)$$

where M is a constant that is sufficiently large to ensure that $P_\theta(\text{path}_i) \leq 1$ for $i = 1, 2, \dots, N$.

Glasserman and Xu (2014) calculated the worst-case measure with respect to the nominal Black-Scholes dynamics. Alternative dynamics, including the Merton's jump-diffusion model and the Heston model, are simulated as well to compare against the worst-case dynamics. The Merton's jump-diffusion model reads

$$\frac{dS_t}{S_{t-}} = (r_n - \lambda E[e^{Y_i} - 1]) dt + \sigma_n dW_t + dJ_t \quad (2.5.41)$$

where J is a compound Poisson process, i.e. $J_t = \sum_{i=1}^{N_t} \exp(Y_i)$, with N_t a Poisson process with intensity λ , and $Y_i \sim N(0, 1)$ i.i.d. This is to say, the price of the underlying asset follows the combination of a geometric Brownian motion and a zero-drift jump process. The simulation results demonstrated significantly lower hedging errors produced by the two alternative models, compared to the error given by the worst-case measure (Glasserman and Xu 2014). This illustrates the importance of using a model risk measure that is not model specific. The model risk could be significantly underestimated by simply comparing different models.

2.5.5 Distribution Model Risk Measurement

Breuer and Csiszár (2016) suggest to measure model risk based on alternative distributions constrained using some divergence form. The distribution model risk is defined by

$$MR := - \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}(X)$$

Instead of being a set of parametric models, \mathcal{Q} here refers to a set of legitimate distributions of relevant risk factors. The set of distributions is defined by a constraint, as proposed in this paper:

$$\mathcal{Q} := \{\mathbb{Q} : D(\mathbb{Q}||\mathbb{P}) \leq \eta\}$$

where \mathbb{P} is a reference distribution, η is a constant (some sort of budget). D is the divergence form, such as the relative entropy (i.e. Kullback-Leibler divergence).

Specifically, the authors applied the Bregman distance to constrain the alternative distributions. Given a probability space $(\Omega, \mathcal{F}, \mu)$, The Bregman distance of distributions $\mathbb{Q} \ll \mu$, $\mathbb{P} \ll \mu$ is defined by

$$B_{f,\mu}(\mathbb{Q}, \mathbb{P}) := \int_{\Omega} f\left(\frac{d\mathbb{Q}}{d\mu}(r)\right) - f\left(\frac{d\mathbb{P}}{d\mu}(r)\right) - f'\left(\frac{d\mathbb{P}}{d\mu}(r)\right) \left(\frac{d\mathbb{Q}}{d\mu}(r) - \frac{d\mathbb{P}}{d\mu}(r)\right) \mu(dr)$$

where f is a strictly convex function in $(0, \infty)$. For $f(s) = s \ln s - s + 1$, the Bregman distance is the relative entropy. The authors also propose a more general framework of measuring model risk by defining the set of legitimate probability measures

$$\mathcal{Q} := \{\mathbf{Q} \ll \mu : H\left(\frac{d\mathbf{Q}}{d\mu}\right) \leq k\}$$

for a given measure μ on Ω . H is a convex functional defined by

$$H(q) := \int_{\Omega} \beta(r, q(r)) \mu(dr)$$

for some measurable, non-negative function p on Ω . $\beta : \Omega \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ is a given map such that $\beta(r, s)$ is a measurable function of r for each $s \in \mathbb{R}_{++}$, and a strictly convex function of s for each $r \in \Omega$.

Under the general formulation, given a random variable X one may calculate the convex conjugate of the function β with respect to the second variable:

$$\beta^*(r, x) := \sup_{s \in \mathbb{R}} (xs - \beta(r, s)), \quad x \in \mathbb{R}$$

and define the probability density function by

$$p_{\theta}(r) := (\beta^*)'(r, \theta_1 + \theta_2 X(r)), \quad \theta = (\theta_1, \theta_2) \in \Theta$$

where

$$\Theta := \{\theta : K(\theta_1, \theta_2) < \infty, \theta_1 + \theta_2 X(r) < \beta'(r, \infty) \mu - a.s.\}$$

The authors consider the general problem

$$J(a, b) := \inf_{p: \int p d\mu = a, \int X p d\mu = b} H(p), \quad (a, b) \in \mathbb{R}^2 \tag{2.5.42}$$

and define the convex conjugate of J

$$K(\theta_1, \theta_2) := \sup_{(a, b) \in \mathbb{R}^2} (\theta_1 a + \theta_2 b - J(a, b))$$

The authors prove that assuming

$$H(q) \geq H(p) = 0, \forall p \in \{p \geq 0 \mid \int p d\mu = 1\}$$

and

$$m := \mu - \text{ess inf}(X) < \int X p_0 d\mu < \mu - \text{ess sup}(X)$$

and

$$0 < k < \lim_{b \downarrow m} \inf_{p: \int p d\mu = 1, \int X p d\mu = b} H(p)$$

if some $\theta = (\theta_1, \theta_2) \in \Theta$ satisfies

$$\theta_2 < 0, \int p_{\theta} d\mu = 1, \theta_1 + \theta_2 \int X p_{\theta} d\mu - K(\theta) = k$$

then the problem formulated in Eq. 2.5.42 has the solution $p = p_{\theta}$.

2.5.6 Wasserstein Metric and Model Uncertainty

Part of the thesis is about formulating the problem of model risk quantification using the Wasserstein metric. In addition, there is work developed simultaneously by applying the Wasserstein metric to tackle model uncertainty. In this section, we will introduce the Wasserstein metric and the work published in parallel to this thesis.

Starting from this point, we will always assume a continuous-state space unless otherwise stated. The approach for discrete-state spaces follows the same routine and therefore is omitted. Now let the state space (Ω, d) be a Polish metric space, we may define the transportation cost $c : \Omega \times \Omega \rightarrow \mathbb{R}_+$ by the n -th power of the metric, i.e. $c(x, y) = d(x, y)^n$, where $n \in [1, \infty)$. Given two probability measures P and Q on (Ω, d) , we may formulate the optimal transportation problem using either a transportation map or a transportation plan. For the former approach, we aim to find the transportation map $T : \Omega \rightarrow \Omega$ that realizes the infimum

$$\begin{aligned} \inf_T \int_{\Omega} p(x)c(x, T(x)) dx & \quad (2.5.43) \\ \text{s.t. } |J_T(x)|q(T(x)) &= p(x), \forall x \in \Omega \end{aligned}$$

where $p(x)$ and $q(x)$ are the probability density functions of the two measures P and Q , respectively. J_T is the Jacobian of the map T . It is part of the constraint that enforces the map T to be measure-preserving. Eq. 2.5.43 is referred as the **Monge's formulation** of the optimal transportation problem (Villani (2008), p.18 and p.22).

The problem of Monge's formulation is that the existence of a measure-preserving map T is not guaranteed. In a continuous-state space, a measure-preserving map sends a Dirac measure to another Dirac measure. Therefore, measure-preserving map does not exist if P is a Dirac measure while Q is not. The ill-posed Monge's formulation can be improved by adopting a transportation plan $\gamma : \Omega \times \Omega \rightarrow \mathbb{R}_+$:

$$\begin{aligned} \inf_{\gamma} \int_{\Omega \times \Omega} \gamma(x, y)c(x, y) dx dy & \quad (2.5.44) \\ \text{s.t. } \int_{\Omega} \gamma(x, y) dy &= p(x) \\ \int_{\Omega} \gamma(x, y) dx &= q(y) \end{aligned}$$

Eq. 2.5.44 is referred as the **Kantorovich's formulation** of the optimal transportation problem. It is clear that every transportation map T can be given by a transportation plan

$$\gamma(x, y) = |J_T(x)|q(y)\delta(y - T(x)) \quad (2.5.45)$$

where $\delta(\cdot)$ is the Dirac- δ function. In addition, the existence of a transportation plan is guaranteed as $\gamma(x, y) = p(x)q(y)$ always satisfies the constraints in Eq. 2.5.44. According to these observations, the Kantorovich's formulation is preferred over the

Monge's formulation. Remember that the transportation cost $c(x, y)$ is the n -th power of the metric $d(x, y)$. The n -th **Wasserstein metric**, denoted by W_n , is defined by

$$W_n = \left(\inf_{\gamma} \int_{\Omega \times \Omega} \gamma(x, y) c(x, y) dx dy \right)^{\frac{1}{n}} \quad (2.5.46)$$

The transportation cost function $c(x, y)$ will be regarded as a generic non-negative function, without reference to its specific form or the power n .

Bartl, Drapeau and Tangpi (2018) introduce the Wasserstein metric to coherent risk measures and option pricing to account for model uncertainty. In particular, they apply the dual formulation of the Wasserstein metric:

$$d_c(p, q) = \sup_{f, g \in C(X)} \left(\int_X f dp + \int_X g dq \right) \text{ s.t. } f(x) + g(y) \leq c(x, y) \quad \forall x, y \in X$$

where $c : X \times X \rightarrow \mathbb{R}^+$ is the cost function. $C(X)$ is the set of all bounded continuous functions $X \rightarrow \mathbb{R}$. They further defined the λc -transform of a function $f : X \rightarrow \mathbb{R}$ by

$$f^{\lambda c}(x) := \sup_{y \in X} (f(y) - \lambda c(x, y))$$

This is similar to the convex conjugate of a function ϕ , defined by

$$\phi^*(x) := \sup_{y \in X} (xy - \phi(y))$$

Bartl et al. provide the following theorem.

Theorem 2.5.1. *Given a closed set $X \subset \mathbb{R}^d$, a cost function c and a penalisation function ϕ , then the following holds ($\mathcal{M}(X)$ is the set of all probability measures on the Borel σ -algebra of X)*

$$\sup_{q \in \mathcal{M}(X)} \left(\int_X f dq - \phi(d_c(p, q)) \right) = \inf_{\lambda \geq 0} \left(\int_X f^{\lambda c} dp + \phi^*(\lambda) \right)$$

for every measurable function f bounded from below. Moreover, the infimum over λ is attained.

Bartl et al. apply the theorem to the optimised certainty equivalent risk measure, defined with respect to a loss function l ,

$$OCE(l) := \inf_{m \in \mathbb{R}} \left(\int l(x - m) p(dx) + m \right)$$

The robust version of this risk measure is given by

$$\begin{aligned} & \inf_{m \in \mathbb{R}} \left(\sup_{q \in \mathcal{M}(\mathbb{R})} \left(\int l(x - m) q(dx) - \phi(d_c(p, q)) \right) + m \right) \\ &= \inf_{\lambda \geq 0} (OCE(l^{\lambda c}) + \phi^*(\lambda)) \end{aligned}$$

according to the theorem above.

In a similar way, one may provide such robustness to pricing a European option with payoff $h(S)$:

Proposition 2.5.2. For every measurable payoff $h : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\sup_{x \in \mathbb{R}^d} |h(x)|/(1+|x|) < \infty$ one has the robust option price given by

$$\inf_{\alpha \in \mathbb{R}^d} \inf_{\lambda \geq 0} \left(\int h^{\lambda c, \alpha} dp + \phi^*(\lambda) \right)$$

where

$$h^{\lambda c, \alpha}(x) := \sup_{y \in \mathbb{R}^d} (h(y) + \alpha(y - s) - \lambda c(x, y))$$

α here is a Lagrange multiplier that constrains the expectation of the payoff function for it is given by the underlying asset price and is assumed to be fair. It is noted that the robust option price given above, however, includes the penalty term, $\phi(d_c(p, q))$. We may instead approach the problem as a constrained optimisation problem with its dual formulation with penalty. The final result, here the robust option price, should be converted back to exclude the penalty term. This would be more consistent with the primal problem with a constrained formulation.

A parallel work by Blanchet, Chen and Zhou (2018) introduces Wasserstein metric to the problem of portfolio optimisation. It provides a robust mean-variance optimality by accounting for model uncertainty. The mean-variance optimisation problem under the reference measure P is formulated by

$$\min_{\phi \in \mathcal{F}_{\infty, \alpha}} \phi^T (E^P(R^2) - E^P(R)^2) \phi \quad (2.5.47)$$

where R is the vector of asset returns and ϕ is the portfolio weight.

$$\mathcal{F}_{\delta, \alpha} = \{ \phi : \phi^T 1 = 1, \min_{D_c(Q, P) \leq \delta} E^Q(\phi^T R) \geq \alpha \}$$

is the feasible set of portfolio weights. It is required that the feasible weights must provide an expected return no less than α , under all alternative probability measures Q such that the Wasserstein distance $D_c(Q, P)$ does not exceeds δ . We further define $\mathcal{M}(\delta) := \{Q : D_c(Q, P) \leq \delta\}$ as the set of alternative measures with constrained Wasserstein distance. The robust mean-variance optimisation problem optimises the portfolio weight under the worst-case measure surrounding P :

$$\min_{\phi \in \mathcal{F}_{\delta, \alpha}} \max_{P \in \mathcal{M}(\delta)} \phi^T (E^P(R^2) - E^P(R)^2) \phi \quad (2.5.48)$$

Blanchet et al. provide characterisation and simplification to this problem.

Proposition 2.5.3. For cost function $c(x, y) = \|x - y\|_q^2, q \geq 1$, we have

$$\min_{Q \in \mathcal{M}(\delta)} E^Q(\phi^T R) = E^P(\phi^T R) - \sqrt{\delta} \|\phi\|_p$$

with $1/p + 1/q = 1$.

The proposition above characterises the mean part of the problem. To characterise the variance part, we have

Proposition 2.5.4. *For arbitrary cost function c that is lower semicontinuous and non-negative, the optimal value function of*

$$\max_{Q \in \mathcal{M}(\delta), E^Q(\phi^T R) = \alpha} \phi^T E^P(RR^T) \phi$$

is given by

$$\inf_{\lambda_1 \geq 0, \lambda_2} \left(\frac{1}{2} \sum_{i=1}^n \Phi(R_i) + \lambda_1 \delta + \lambda_2 \alpha \right)$$

where

$$\Phi(R_i) := \sup_u \left((\phi^T u)^2 - \lambda_1 c(u, R_i) - \lambda_2 \phi^T u \right)$$

If we take a quadratic cost function, we may solve the infimum problem with respect to the two Lagrange multipliers λ_1 and λ_2 ,

Proposition 2.5.5. *Let $c(x, y) = \|x - y\|_q^2$ with $q \geq 1$ and $1/p + 1/q = 1$. If $(\alpha - \phi^T E^P(R))^2 - \delta \|\phi\|_p^2 \leq 0$, then the value of*

$$\max_{Q \in \mathcal{M}(\delta), E^Q(\phi^T R) = \alpha} \phi^T E^P(RR^T) \phi$$

is given by

$$\begin{aligned} h(\alpha, \phi) := & E^P((\phi^T R)^2) + 2(\alpha - \phi^T E^P(R)) \phi^T E^P(R) + \delta \|\phi\|_p^2 \\ & + 2\sqrt{\delta \|\phi\|_p^2 - (\alpha - \phi^T E^P(R))^2} \sqrt{\phi^T \text{Var}^P(R) \phi} \end{aligned}$$

The problem of robust portfolio optimisation, Eq. 2.5.48, is converted to an optimisation problem with respect to the portfolio weight. The converted problem differs from Eq. 2.5.47 by an extra regularization term:

Theorem 2.5.6. *The primal formulation given in Eq. 2.5.48 is equivalent to the dual problem*

$$\min_{\phi \in \mathcal{F}_{\delta, \alpha}} \left(\sqrt{\phi^T (E^P(R^2) - E^P(R)^2) \phi} + \sqrt{\delta} \|\phi\|_p \right)^2$$

2.6 Towards Dynamic Theory of Model Risk

2.6.1 Robust Control and Model Uncertainty

Hansen et al. developed the theory of robust control under model ambiguity (Hansen, Sargent, Turmuhambetova and Williams 2006, Hansen and Sargent 2001). Their work

in principle is quite similar to the quantification of financial model risk. Starting from a reference model, we look for alternative models surrounding the reference model and find out the worst-case scenario, seeking a solution to the control problem which is robust with respect to such worst-case scenarios. Technically, there are two categories of approaches. The simpler one is to consider alternative models in a parametric way, by varying parameter values or adding some parametric terms (e.g. risk-sensitivity). The other category is non-parametric, aiming to cover an entire neighborhood of the reference model.

Hansen et al. studied the control (optimization) problem in which the decision maker solves the following problem under the reference model:

$$J(x_0) = \max_{c \in C} \mathbb{E} \left(\int_0^\infty e^{-\delta t} U(c_t, x_t) dt \right) \quad (2.6.1)$$

where x_t is the state variable at time t and c_t is the control variable. $c = \{c_t\}_{t \in \mathbb{R}^+}$ is taken from C , the set of admissible control policies. The reference model is an Ito diffusion process $dx_t = \mu(c_t, x_t)dt + \sigma(c_t, x_t)dW_t$. We have the following observations with regards to the problem formulation:

1. The reference model is Markov, i.e. information of the entire history is “absorbed” by the current state (c_t, x_t) .
2. It only considers diffusion process (i.e. continuous sample path a.s.)
3. The objective is a cumulative function over time, where the accumulation rate, $U(c_t, x_t)$ is Markov.

These three points will be generalised in this thesis, using the powerful tool of functional Ito calculus.

Due to its cumulative nature, Eq. 2.6.1 may be solved by iterating backwards with the help of dynamic programming. The HJB equation for this problem is

$$\delta J(x) = \max_{c \in \hat{C}} U(c, x) + \mu(c, x)J_x(x) + \frac{1}{2} \text{Tr} (\sigma(c, x)^T J_{xx}(x)\sigma(c, x)) \quad (2.6.2)$$

where $J(x)$ is the value function, and $\delta J(x)$ is the accumulation rate of the value function. Note that here c and x denote potentially realized value at time t . \hat{C} is the set of admissible values of the control variable c . Subscripts on value functions denote the respective derivatives.

Model uncertainty may be introduced to the control problem in a non-parametric way. First we need to introduce the notations that formulate the problem. The reference model (Ito diffusion process) induces a reference probability measure P on the measurable space (Ω, \mathcal{F}) . For any alternative measure Q on (Ω, \mathcal{F}) , we let Q_t denote the restriction to \mathcal{F}_t . Hansen et al. weaken the condition of absolute continuity to deal with the infinite-time setting of the stochastic process.

Definition 2.6.1. A probability measure Q is absolutely continuous over finite intervals with respect to P if Q_t is absolutely continuous with respect to P_t for all $t < \infty$.

To measure the discrepancy between the alternative measure Q and the reference measure P , Hansen et al. proposed a discounted measure of relative entropy:

$$\mathcal{R}(Q) := \delta \int_0^\infty e^{-\delta t} \left(\int \ln \left(\frac{dQ_t}{dP_t} \right) dQ_t \right) dt$$

After defining the discounted relative entropy, Hansen et al. proposed two versions of robust control problems,

$$V(\theta) = \sup_{c \in \mathcal{C}} \inf_{Q \in \mathcal{M}} \int_0^\infty e^{-\delta t} \left(\int U(c_t, x_t) dQ_t \right) dt + \theta \mathcal{R}(Q) \quad (2.6.3)$$

$$K(\eta) = \sup_{c \in \mathcal{C}} \inf_{Q \in \mathcal{M}_\eta} \int_0^\infty e^{-\delta t} \left(\int U(c_t, x_t) dQ_t \right) dt \quad (2.6.4)$$

where $\mathcal{M}(\eta) := \{Q \in \mathcal{M} : \mathcal{R}(Q) \leq \eta\}$. Eq. 2.6.3 is referred as a penalty robust control problem and Eq. 2.6.4 is a constrained robust control problem. Hansen et al. established the relation between the two formulations of robust control problems. By virtue of the convex nature of the problem, their work results in the following proposition.

Proposition 2.6.2. *Fix $\eta > 0$. Suppose that c and Q solve the constrained robust control problem for $K(\eta) > -\infty$. Then there exists $\theta > 0$ such that the corresponding penalty robust control problem has the same solution. Moreover,*

$$K(\eta) = \max_{\theta \geq 0} V(\theta) - \theta \eta$$

The proposition above proves the existence of a θ that corresponds to a given η . To establish a closed-form relation between the two parameters, we need to introduce a min-max condition. It is clear that Eq. 2.6.3 describes a two-party zero-sum game, in which one party tries to maximise while the other tries to minimise the value of a function. Under certain convexity conditions, the following assumption can indeed be established:

Assumption 2.6.3. The inf's and sup's are attained in the control problems, and

$$\max_{c \in \mathcal{C}} \min_{Q \in \mathcal{M}_\eta} \int_0^\infty e^{-\delta t} \left(\int U(c_t, x_t) dQ_t \right) dt = \min_{Q \in \mathcal{M}_\eta} \max_{c \in \mathcal{C}} \int_0^\infty e^{-\delta t} \left(\int U(c_t, x_t) dQ_t \right) dt$$

Under such assumption, the following proposition can be proved.

Proposition 2.6.4. *Suppose that Assumption 2.6.3 holds, and (c, Q) solves the penalty robust control problem Eq. 2.6.3. Then c also solves the constrained robust control problem Eq. 2.6.4 for $\eta = \mathcal{R}(Q)$, and*

$$V(\theta) = \min_{\eta \geq 0} K(\eta) + \theta \eta$$

Hansen et al. also define two preference orderings associated with the two versions of robust control problems.

Definition 2.6.5. For any two progressively measurable c and c^* , $c^* \succeq_\eta c$ (constrained preference ordering) if

$$\hat{K}(c^*, \eta) \geq \hat{K}(c, \eta)$$

where

$$\hat{K}(c, \eta) := \inf_{Q \in \mathcal{M}_\eta} \int_0^\infty e^{-\delta t} \left(\int U(c_t, x_t) dQ_t \right) dt \quad (2.6.5)$$

Definition 2.6.6. For any two progressively measurable c and c^* , $c^* \succeq_\theta c$ (penalty preference ordering) if

$$\hat{V}(c^*, \eta) \geq \hat{K}(c, \eta)$$

where

$$\hat{V}(c, \eta) := \inf_{Q \in \mathcal{M}} \int_0^\infty e^{-\delta t} \left(\int U(c_t, x_t) dQ_t \right) dt + \theta \mathcal{R}(Q)$$

The two preference orderings may differ. However, they do agree locally in the sense that the indifference curve at a given point c^* in the consumption set are tangent for both preference orderings. Applying the Lagrange multiplier theorem we convert Eq. 2.6.5 into its dual form:

$$\hat{K}(c, \eta) = \max_{\theta \geq 0} \inf_{Q \in \mathcal{M}} \int_0^\infty e^{-\delta t} \left(\int U(c_t, x_t) dQ_t \right) dt + \theta (\mathcal{R}(Q) - \eta)$$

Let θ^* denote the maximizing value of θ . If $c^* \succeq_\eta c$, then

$$\hat{V}(c, \theta^*) - \theta^* \eta \leq \hat{K}(c, \eta) \leq \hat{K}(c^*, \eta) = \hat{V}(c^*, \theta^*) - \theta^* \eta$$

thus $c^* \succeq_\theta c$. This shows that the indifference curves of the two preference orderings touch at c^* but do not cross.

Hansen et al. represent the change of measure by a nonnegative martingale defined by $Z(t) = \frac{dQ_t}{dP_t}$.

Lemma 2.6.7. Suppose that for all $t \geq 0$, Q_t is absolutely continuous with respect to P_t . The process $\{Z(t)\}_{t \geq 0}$ on (Ω, \mathcal{F}, P) is a nonnegative martingale adapted to $\{\mathcal{F}_t\}_{t \geq 0}$ with $E(Z(t)) = 1$. Moreover,

$$\int \phi_t dQ_t = E(Z(t)\phi_t) \quad (2.6.6)$$

for any bounded and \mathcal{F}_t -measurable function ϕ_t . Conversely, if $\{Z(t)\}_{t \geq 0}$ is a nonnegative progressively measurable martingale with $E(Z(t)) = 1$, then the probability measure Q defined by Eq. 2.6.6 is absolutely continuous with respect to P over finite intervals.

Since the reference model is provided by an Ito diffusion process, we may exploit the Brownian motion $\{W_t\}_{t \geq 0}$ and represent the martingale density $\{Z(t)\}_{t \geq 0}$ by

$$Z(t) = 1 + \int_0^t k_s dW_s$$

where k is a progressively measurable process satisfying

$$P\left(\int_0^t |k_s|^2 ds < \infty\right) = 1$$

for any $t < \infty$. Define $h_t := k_t/Z(t)$ if $Z(t) > 0$ otherwise $h(t) := 0$. Then $\{Z(t)\}_{t \geq 0}$ is the solution to the SDE

$$dZ(t) = Z(t)h_t dW_t \tag{2.6.7}$$

subject to the initial condition $Z(0) = 1$.

Therefore, we may reformulate the penalty robust control problem, Eq. 2.6.3, with respect to the process $\{h_t\}_{t \geq 0}$ instead of the alternative measure Q . First we need to evaluate the discounted version of relative entropy:

Lemma 2.6.8. For $Q \in \mathcal{M}$, let Z be the density martingale associated with Q and let h be the progressively measurable process satisfying Eq. 2.6.7, then

$$\mathcal{R}(Q) = \frac{1}{2} E\left(\int_0^\infty e^{-\delta t} Z(t) |h_t|^2 dt\right)$$

Using the Lemma above, we reformulate the penalty robust control problem:

Definition 2.6.9. The martingale version of the penalty robust control problem is

$$\max_{c \in \hat{C}} \min_{h \in H} \int_0^\infty e^{-\delta t} \left(\int Z(t) \left(U(c_t, x_t) + \frac{\theta}{2} |h_t|^2 \right) dQ_t \right) dt \tag{2.6.8}$$

subject to

$$\begin{aligned} dx_t &= \mu(c_t, x_t) dt + \sigma(c_t, x_t) dW_t \\ dZ(t) &= Z(t) h_t dW_t \end{aligned}$$

Fleming and Souganidis (1989) show that a Bellman-Isaacs condition provides equilibrium outcomes identical with outcomes of a Markov perfect equilibrium. This allows us to formulate a HJB equation to solve the penalty control problem Eq. 2.6.8,

$$\delta J(x) = \max_{c \in \hat{C}} \min_h U(c, x) + \frac{\theta}{2} h^2 + (\mu(c, x) + \sigma(c, x)h) J_x(x) + \frac{1}{2} \text{Tr}(\sigma(c, x)^T J_{xx}(x) \sigma(c, x)) \tag{2.6.9}$$

where $J(x)$ is now the value function of the penalty robust control problem. Compared to Eq. 2.6.2, the HJB equation above introduces robustness by imposing a minimization with respect to the change-of-measure characteristics, h , and including extra terms of h . Solving the inner minimization problem results in a minimizing value of h :

$$h^* = -\frac{\sigma(c, x)J_x(x)}{\theta}$$

Substituting this into Eq. 2.6.9 leads to

$$\delta J(x) = \max_{c \in \hat{C}} +\mu(c, x)J_x(x) - \frac{1}{2\theta^2}\sigma^2(c, x)J_x^2(x) + \frac{1}{2}\text{Tr}(\sigma(c, x)^T J_x x(x)\sigma(c, x)) \quad (2.6.10)$$

Again, since the primary concern of this thesis is on the quantification of model risk, the controlling part $\max_{c \in \hat{C}}$ is not particularly relevant here. Ignoring $\max_{c \in \hat{C}}$, the form of Eq. 2.6.10 will be given as a special case of a general theory in Chapter. 4.

2.6.2 Functional Ito Calculus

Cont et al. developed the functional Ito calculus as path-dependent generalization of the renowned Ito calculus (Cont 2016, Bally, Caramellino and Cont 2016). Building on the work of Hansen et al. reviewed above, in Chapter 4 we will make use of functional Ito calculus to develop a new approach to dynamic model risk measurement. In this section we will discuss the ideas and concepts of the theory.

We assume a continuous-time setting with finite interval. A (path-dependent) model is essentially a probability measure on the path space. Formally, fix $T \in (0, \infty)$ and $d \in \mathbb{N}$, and let $\Omega := D([0, T], \mathbb{R}^d)$ denote the set of càdlàg paths $\omega : [0, T] \rightarrow \mathbb{R}^d$. Let $[0, T] \ni t \mapsto X(t)$ be the canonical process on Ω , which means to say that $X(t)(\omega) := \omega(t)$, for all $(t, \omega) \in [0, T] \times \Omega$. Let $\mathfrak{F}^0 = (\mathcal{F}_t^0)_{t \in [0, T]}$ denote the filtration on Ω generated by X , which is to say that

$$\mathcal{F}_t^0 := \bigvee_{s \in [0, t]} \{X(s)^{-1}(U) \mid U \in \mathcal{B}(\mathbb{R}^d)\} = \bigvee_{s \in [0, t]} \bigcup_{U \in \mathcal{B}(\mathbb{R}^d)} \{\omega \in \Omega \mid \omega(s) \in U\},$$

for all $t \in [0, T]$. In particular,

$$\mathcal{F}_0^0 := \{X(0)^{-1}(U) \mid U \in \mathcal{B}(\mathbb{R}^d)\} = \bigcup_{U \in \mathcal{B}(\mathbb{R}^d)} \{\omega \in \Omega \mid \omega(0) \in U\}.$$

Fix a reference probability measure \mathbb{P} on $(\Omega, \mathcal{F}_T^0)$, subject to the condition

$$\mathbb{P}(X(0)^{-1}(U)) = \mathbb{P}(\{\omega \in \Omega \mid \omega(0) \in U\}) = \begin{cases} 1 & \text{if } 0 \in U; \\ 0 & \text{if } 0 \notin U, \end{cases}$$

for all $U \in \mathcal{B}(\mathbb{R}^d)$, which is to say that almost all paths start at zero under \mathbb{P} . Note that this condition ensures that $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, for all $A \in \mathcal{F}_0^0$.

To be consistent with the notation in Cont (2016), we shall write $\omega_t := \omega(t \wedge \cdot) \in \Omega$ to denote the path $\omega \in \Omega$ stopped at time $t \in [0, T]$. We impose an equivalence relation \sim on $[0, T] \times \Omega$, by specifying that

$$(t, \omega) \sim (t', \omega') \quad \text{if and only if} \quad t = t' \quad \text{and} \quad \omega_t = \omega'_{t'},$$

for all $(t, \omega), (t', \omega') \in [0, T] \times \Omega$. That is to say, two pairs, each consisting of a time and a path, are equivalent if the times are equal and the corresponding stopped paths are the same. The quotient set $\Lambda_T^d := [0, T] \times \Omega / \sim$ forms a complete metric space, when endowed with the metric $d_\infty : (\Lambda_T^d)^2 \rightarrow \mathbb{R}_+$, defined by

$$d_\infty((t, \omega), (t', \omega')) := \sup_{s \in [0, T]} |\omega(s \wedge t) - \omega'(s \wedge t')| + |t - t'| = \|\omega_t - \omega'_{t'}\|_\infty + |t - t'|,$$

for all $(t, \omega), (t', \omega') \in \Lambda_T^d$. We refer to (Λ_T^d, d_∞) as the *space of stopped paths*.

A measurable function $F : \Lambda_T^d \rightarrow \mathbb{R}$ is called a *non-anticipative functional*, where Λ_T^d is endowed with the Borel sigma-algebra generated by d_∞ and \mathbb{R} is endowed with the Borel sigma-algebra generated by the usual Euclidean metric. Since $(t, \omega) \sim (t, \omega_t)$, for all $(t, \omega) \in [0, T] \times \Omega$, we may regard a non-anticipative functional $F : \Lambda_T^d \rightarrow \mathbb{R}$ as an appropriately measurable function $F : [0, T] \times \Omega \rightarrow \mathbb{R}$ that satisfies the condition $F(t, \omega) = F(t, \omega_t)$. That is to say, the value of a non-anticipative functional, when applied to a particular time and path, depends only on the behaviour of the path up to the time. Note that $(F(t, \cdot))_{t \in [0, T]}$ is a progressively measurable process, adapted to the filtration \mathfrak{F}^0 .

Cont et al. developed the functional Ito formula for general semimartingales.(Cont 2016, Bally et al. 2016). Here we only introduce the version for continuous semimartingales. First of all, just like standard Ito calculus, we need (second-order) derivatives. This requires formal definition of such derivatives in the context of non-anticipative functionals. Two types of derivatives are present for a non-anticipative functional $F : \Lambda_T^d \rightarrow \mathbb{R}$, namely the horizontal derivative and the vertical derivative. The horizontal derivative at $(t, \omega) \in \Lambda_T^d$ is defined by the limit

$$\mathcal{D}F(t, \omega) := \lim_{h \rightarrow 0^+} \frac{F(t+h, \omega) - F(t, \omega)}{h}$$

if it exists. Intuitively, it describes the rate of change w.r.t time, assuming no change of the state variable from t onwards, and conditional to its history up to t given by the stopped path ω_t . On the other hand, the vertical derivative describes the rate of change w.r.t the state variable from t onwards. Formally, the vertical derivative at $(t, \omega) \in \Lambda_T^d$, denoted by $\nabla_\omega F(t, \omega)$, is defined as the gradient of the function $\mathbb{R}^d \ni x \mapsto F(t, \omega_t + x \mathbf{1}_{[t, T]})$ at 0, assuming its existence. The horizontal and vertical derivatives of a non-anticipative functional are also non-anticipative functionals.

We define the left-continuous non-anticipative functionals by noticing that the space of stopped paths, Λ_T^d , is endowed with a metric d_∞ . Suppose $F : \Lambda_T^d \rightarrow \mathbb{R}$ is a non-anticipative functional. F is left-continuous if for every $(t, \omega) \in \Lambda_T^d$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $|F(t, \omega) - F(t', \omega')| < \varepsilon$ for all $(t', \omega') \in \Lambda_T^d$

satisfying $t' < t$ and $d_\infty((t, \omega), (t', \omega')) < \delta$. We may further impose a boundedness condition to a non-anticipative functional F . It states that for any compact $K \subset \mathbb{R}^d$ and $t_0 < T$, there exists a $C > 0$ such that $|F(t, \omega)| \leq C$ for all $t \leq t_0$ and $\omega \in \Omega$. Suppose a non-anticipative functional F is horizontally differentiable and vertically twice-differentiable for all $(t, \omega) \in \Lambda_T^d$, and $\mathcal{D}F$, $\nabla_\omega F$ and $\nabla_\omega^2 F$ satisfy the boundedness condition above. In addition, F , $\nabla_\omega F$ and $\nabla_\omega^2 F$ are left-continuous, and $\mathcal{D}F$ is continuous for all $(t, \omega) \in \Lambda_T^d$. Then we call F a regular functional.

Suppose the canonical process X on Ω is a continuous semimartingale and $F : \Lambda_T^d \rightarrow \mathbb{R}$ is a regular functional. The \mathbb{R} -valued process $(Y(t))_{t \in [0, T]}$, defined by $Y(t) = F(t, \cdot)$ for all $t \in [0, T]$, follows the functional Ito formula P-a.s. (Bally et al. 2016, pp. 190–191)

$$Y(t) - Y(0) = \int_0^t \mathcal{D}F(u, \cdot) du + \int_0^t \nabla_\omega F(u, \cdot) dX(u) + \frac{1}{2} \int_0^t \text{Tr}(\nabla_\omega^2 F(u, \cdot) d[X](u))$$

If we further impose the constraint that $\int_0^T \xi(t) dX(t) = 0$ for all bounded predictable processes ξ satisfying $\int_0^T \xi(t) dt = 0$, then the canonical process X is a strong solution to the SDE (Revuz and Yor 2013)

$$dX(t) = \mu(t)dt + \sigma(t)dW(t) \tag{2.6.11}$$

where $(W(t))_{t \in [0, T]}$ is a \mathbb{R}^d -valued standard Wiener process on the underlying filtered probability space (assuming its existence). $(\mu(t))_{t \in [0, T]}$ is a \mathbb{R}^d -valued predictable process, and $(\sigma(t))_{t \in [0, T]}$ is a \mathbb{R}^{d^2} -valued predictable process. We may identify their elements, say $(\mu_i(t))_{t \in [0, T]}$ and $(\sigma_{ij}(t))_{t \in [0, T]}$, with non-anticipative functionals. The SDE Eq. 5.3.1 may be regarded as a path-dependent generalisation of the renowned Ito diffusion process. The existence and uniqueness of its solutions have been given in the literature by imposing various conditions (e.g. boundedness and Lipschitz properties, see (Bally et al. 2016)). Now if X satisfies Eq. 5.3.1 P-a.s., then it follows from the functional Ito formula that the process Y is a strong solution to the SDE

$$dY(t) = \left(\mathcal{D}F(t, \cdot) + \mu(t)\nabla_\omega F(t, \cdot) + \frac{\text{Tr}(\sigma(t)^2 \nabla_\omega^2 F(t, \cdot))}{2} \right) dt + \sigma(t)\nabla_\omega F(t, \cdot) dW(t)$$

Note that the square of $\sigma(t)$ is in the sense of matrix multiplication, i.e. $\sigma(t)^2 = \sigma(t)\sigma(t)^T$.

Chapter 3

Quantifying Model Risks in Option Pricing Models

The renowned statistician George E. P. Box wrote that “essentially, all models are wrong, but some are useful.”¹ This is certainly true in finance, where many models and techniques that have been extensively empirically invalidated remain in widespread use, not just in academia, but also (perhaps especially) among practitioners. At times, the way models are used directly contradicts the model assumptions: As observed market prices change, parameters in option pricing models, which are assumed to be time-invariant, are recalibrated, often on a daily basis. Incorrect models, and model misuse, represent a source of risk that is being increasingly recognised — this is called “model risk.”

In this chapter, we focus on two particular aspects of model risk: the inability of a chosen model to fit observed market prices at a given point in time (calibration error) and the model risk due to recalibration of model parameters (in contradiction to the model assumptions). In this context, we follow the approach of Glasserman and Xu (2014) to use relative entropy as a consistent pre-metric in order to quantify these two sources of model risk in a common framework, and consider the trade-offs between them when choosing a model and the frequency with which to recalibrate to the market. We illustrate this approach applied to the models of Black and Scholes (1973) and Heston (1993), using option data for Apple (AAPL) and Google (GOOG).

3.1 Introduction

As a paper by the Board of Governors of the Federal Reserve System put it in 2011,² “The use of models invariably presents model risk, which is the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports.” However, there has been surprisingly little research to date on quantifying

¹See Box and Draper (1987).

²See Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency (2011).

this risk, or putting the analysis of this risk on a more rigorous footing. One notable exception is a paper by Glasserman and Xu (2014), who propose relative entropy as a consistent pre-metric by which to measure model risk from different sources. This allows one to optimise the choice of a model and the choice of the way a model is implemented in day-to-day practice under a unifying criterion, while also quantifying the remaining model risk.

In broad terms, one could identify four general classes of model risk inherent to the way mathematical models are used in finance, for example in (but not limited to) option pricing applications:

- ❖ Parameter uncertainty (and sensitivity to parameters) – let’s call this **“Type 0”** model risk for short. If model parameters need to be statistically estimated, they will only be known up to some level of statistical confidence, and this parameter uncertainty induces uncertainty about the correctness of the model outputs.³
- ❖ Inability to fit a model to a full set of simultaneous market observations – this is “calibration error,” let’s call this **“Type 1”** model risk for short. To the extent that a model cannot match observed prices on a given day, single-day (a.k.a. “cross-sectional”) market data already contradicts the model assumptions. The classical example of this is the Black/Scholes implied volatility smile.
- ❖ Change in parameters due to recalibration – let’s call this **“Type 2”** model risk for short. Once one moves from one day to the next, this aspect of model risk becomes apparent: In order to again fit the market as closely as possible, it is common practice in the industry to recalibrate models. This recalibration results in model parameters (which the models assume to be fixed) changing from day to day, contradicting the model assumptions.
- ❖ The “true” dynamics of state variables don’t match model dynamics⁴ – let’s call this violation of model assumptions **“Type 3”** model risk.⁵ The classical example of this is the econometric rejection of the hypothesis that asset prices follow geometric Brownian motion, thus invalidating the key assumption in the seminal model of Black and Scholes (1973). This type of model risk would impact in particular the effectiveness of hedging strategies based on a model.⁶

Note that there is a gradual transition between the different types of model risk, and depending on one’s modelling choices, to a certain extent one can trade off one type of model risk against another. For example,

³Examples of where this type of risk is considered explicitly in the literature include Löffler (2003), Bannör and Scherer (2013) and Kerkhof et al. (2010).

⁴This type of model risk is considered for example in Kerkhof et al. (2010), who also relate this to *identification risk*, which they define as risk which “arises when observationally indistinguishable models have different consequences for capital reserves.”

⁵Boucher et al. (2014) present a method for making value-at-risk more robust with respect to this source of model risk by “learning” from the results of model backtesting.

⁶Detering and Packham (2016) take the approach of measuring model risk based on the residual profit/loss from hedging in a misspecified model.

- ❖ Less stringent requirements of an exact fit to market observations (Type 1) allows less frequent recalibration (Type 2).
- ❖ Instead of different model dynamics (Type 3), one could consider a parameterised family of models (Type 2).
- ❖ Regime-switching models “legalise” changes in parameters, so Type 2 becomes more like Type 3.
- ❖ Adding parameters shifts model risk from Type 1 to Type 2 (or, to a certain extent, to Type 0).
- ❖ Adding state variables shifts model risk from Type 2 to Type 3.

Glasserman and Xu (2014) propose relative entropy as a consistent pre-metric by which to measure model risk from different sources.⁷ What matters in the application of mathematical models in finance is the probability distributions which the models imply,⁸ either under a “risk-neutral” probability measure (for applications to relative pricing of financial instruments) or the “physical” (a.k.a. “real-world”) probability measure (for risk management applications such as the calculation of expected shortfall). Each type of model risk manifests itself as some form of ambiguity about the “true” probability measures which should be used for these purposes, and being able to quantify different types of model risk in a unified setting using a pre-metric for the divergence between distributions (like relative entropy) allows one to make an informed choice about the trade-offs between different sources of model risk. Glasserman and Xu (2014) postulate a “relative entropy budget” defining a set of models sufficiently close (in the sense of relative entropy) to a nominal reference model to be considered in an evaluation of model risk expressed as a “worst case” expectation – i.e., a worst-case price or a worst-case risk measure. However, they say little as to how one typically would obtain a specific number for this “relative entropy budget”. In a sense, we invert this problem by noting that higher relative entropy between model distributions indicates higher model risk, and propose a method to jointly evaluate model risk of two types, based on how this model risk manifests itself when option pricing models are calibrated and recalibrated to liquid market instruments.

We focus on the model risk inherent in the calibration and recalibration (i.e., in the above terminology, Types 1 and 2) of option pricing models, and to illustrate our approach we consider the models of Black and Scholes (1973) and Heston (1993), thus comparing the most classical option pricing model with its popular extension incorporating stochastic volatility. Clearly, if (as is often the case in practice) one focuses solely on calibration error, Heston (1993) will always be preferred to Black and Scholes (1973), and more frequent recalibration preferred to less. We quantify

⁷Instead of using a relative entropy pre-metric, one could approach quantifying model risk in terms of optimal-transport distance, using for example Wasserstein distance, which most recently has become popular for this purpose (see Bartl et al. (2018), Blanchet et al. (2018) and Feng and Schlögl (2018)). In the present paper, we follow the more established approach using relative entropy, which has its roots in the seminal work of Hansen and Sargent (see e.g. Hansen and Sargent (2006)).

⁸Breuer and Csizsár (2016) call this *distribution model risk*.

calibration and recalibration risk in both models applied to equity option data, and also explore the trade-off between these two types of model risk, finding that there is no longer a trivial answer to the question which model and which recalibration frequency should be preferred when these two sources of model risk are considered in a unified framework.

The rest of the chapter is organised as follows. Section 3.2 introduces a framework for the joint evaluation of model risk due to calibration error and due to model recalibration. The numerical implementation of the method is discussed in Section 3.3. Section 3.4 presents the results obtained by applying this method to option price data, and Section 3.5 concludes.

3.2 Calibration error, model risk due to recalibration, and treatment of latent state variables

As noted above, model risk is reflected in the ambiguity with regard to the “correct” probability distribution to use for relative pricing or risk assessment. Following Glasserman and Xu (2014), we quantify this ambiguity using the divergence between probability measures. In the present context, these can be classified as divergence measures defined as a function $D(\cdot||\cdot) : S \times S \rightarrow \mathbb{R}$ satisfying

$$D(Q||P) \geq 0 \quad \forall P, Q \in S \quad (3.2.1)$$

$$D(Q||P) = 0 \Leftrightarrow P = Q \quad (3.2.2)$$

where S is a space of all probability measures with a common support. More specifically, most divergence measures belong to the class of f -divergence, which gives the divergence between two equivalent measures as:⁹

$$D(Q||P) = \int_{\Omega} f\left(\frac{dQ}{dP}\right) dP \quad (3.2.3)$$

where f is a convex function of the Radon-Nikodym derivative satisfying $f(1) = 0$. Kullback–Leibler divergence (a.k.a. relative entropy) is the most common f -divergence, which assigns $f(m) = m \ln m$. It is noted that the methodology of this paper applies to all types of statistical distances in principle, though in the empirical study the Kullback–Leibler divergence is adopted due to its simplicity and widespread use.

If we wish to quantify calibration error (Type 1 model risk) in this fashion, then in equations (3.2.1)–(3.2.3), the probability measure P corresponds to the calibrated model and thus is parametric in some form. The probability measure Q , on the other hand, serves as a reference measure exactly matching observed market prices at a given point in time, unrestricted by the assumptions of the model under consideration. On calibrating an option pricing model, we may regard the measure Q as some non-parametric risk-neutral measure that explains the market in full assuming absence of

⁹See e.g. Ali and Silvey (1966), Csiszár (1967) or Ahmadi-Javid (2012).

arbitrage. In practice, however, the measure Q is not unique as the market is usually incomplete. We therefore define the space of all probability measures that explains the market in full by $S_Q \subset S$.

We may further define the space of probability measures given by all possible choices of parameter values for the target model by $S_P \subset S$. The new calibration methodology proposed here aims to minimise the calibration error as quantified by the divergence between the two measures P and Q , taken from their respective spaces, i.e.

$$(Q^*, P^*) = \arg \min_{Q \in S_Q, P \in S_P} D(Q||P) \quad (3.2.4)$$

This is to say, the new approach attempts to calibrate a model measure P^* (i.e., a set of model parameters θ^*) and non-parametric perfect fit to the market (at a given point in time) Q^* , in a fashion which minimises the calibration error expressed by

$$\eta_1 = D(Q^*||P^*) \quad (3.2.5)$$

This is not an end in itself — it is required in order to compare model risk due to calibration error and model risk due to recalibration (as specified below) in a unified framework.

The classical approaches of model calibration, such as minimising the mean-squared error between model and market prices for options, would be inappropriate in this context, as they would lead to unnecessarily high model risk quantities. It is the choice of divergence measure which informs the calibration procedure, resulting in a pair of probability measures, (Q^*, P^*) , one of which corresponds to the calibrated model while the other provides a consistent reference measure fitting the market exactly.

To quantify the model risk due to recalibration, let us consider the more specific case where the model is Markovian in a vector of observable state variables X , the model is characterised by a vector of model parameters θ , and market prices are given for European option prices of a single maturity T .¹⁰ Suppose we solved (3.2.4) yesterday (at time t_{i-1}) to obtain a P^* — to be as explicit as possible, denote this by

$$P_{t_{i-1}, X_{t_{i-1}}(\omega), \theta_{t_{i-1}}^*}(A) \quad \forall A \in \mathcal{F}_T \quad (3.2.6)$$

I.e., this is a (conditional) probability measure defined on all \mathcal{F}_T -measurable events, where the conditioning is on the state variables at time t_{i-1} , $X_{t_{i-1}}$, and we write $X_{t_{i-1}}(\omega)$ to express that the time t_{i-1} realisations of the state variables are known at the time that these probabilities are evaluated. We write the subscript $\theta_{t_{i-1}}^*$ to express that these probabilities are evaluated in a model with parameters calibrated by solving (3.2.4) at time t_{i-1} . Furthermore, denote the non-parametric measure Q^* resulting from solving (3.2.4) at time t_{i-1} by $Q_{t_{i-1}}$.

¹⁰This last assumption of a single maturity T avoids the need to constrain the choice of Q to ensure the absence of calendar spread arbitrage between non-parametric risk-neutral measures for different time horizons — parametric models typically ensure this by construction. If we appropriately constrain Q , this assumption can be lifted.

Now, if we recalibrate today (at time t_i) by solving (3.2.4), we obtain Q_{t_i} and

$$P_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}(A) \quad \forall A \in \mathcal{F}_T \quad (3.2.7)$$

We can then define the model risk quantity due to recalibration as

$$\eta_2 = D(P_{t_i, X_{t_i}(\omega), \theta_{t_i}^*} \| P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}) \quad (3.2.8)$$

which is the divergence between the (conditional) probability measures evaluated at time t_i , where one measure is based on the recalibrated parameters $\theta_{t_i}^*$ and the other is based on the previously calibrated parameters $\theta_{t_{i-1}}^*$ (thus expressing, in terms of divergence, the inconsistency with the model assumptions due to the fact that we are going “outside of the model” to change parameters in recalibration). The aggregate of calibration error and model risk due to recalibration is then

$$\eta_3 = D(Q_{t_i} \| P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}) \quad (3.2.9)$$

i.e., the divergence between the non-parametric probability measure Q_{t_i} obtained by solving (3.2.4) at time t_i , and the non-recalibrated parametric probability measure, consisting of probabilities conditional on the state at time t_i , but based on model parameters obtained by solving (3.2.4) at time t_{i-1} . However, this approach minimises the divergence between the reference distribution and the recalibrated distribution, thus arguably overstating the divergence to the non-recalibrated (i.e. model-consistent) distribution, and therefore overstating the aggregate model risk η_3 .

Alternatively, we may choose as the non-parametric reference distribution at time t_i :

$$\hat{Q}_{t_i} = \arg \min_{Q \in S_Q} D(Q \| P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}) \quad (3.2.10)$$

resulting in a lower aggregate model risk of

$$\eta_3 = D(\hat{Q}_{t_i} \| P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}) \quad (3.2.11)$$

Note that $\theta_{t_i}^*$ is still obtained by solving (3.2.4), because both Q_{t_i} and \hat{Q}_{t_i} represent non-parametric probability measures fitting observed market prices exactly, so $\theta_{t_i}^*$ remains the best available parametric fit to the market at time t_i (\hat{Q}_{t_i} is only used to determine minimum divergence of the non-recalibrated model to a measure giving a perfect fit).

In the heuristic schematic of Figure 1(a),¹¹ point A represents $P_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}$, being the parametric probability measure “closest” to the set of non-parametric probability measures fitting the market exactly, where point C represents Q_{t_i} . If we do not recalibrate at time t_i , we end up with the parametric probability measure $P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}$ (point B), to which \hat{Q}_{t_i} (point D) is the “closest” non-parametric probability measure fitting the market exactly.

¹¹Note that these graphs are for the purpose of heuristic illustration only – in particular, we are not requiring that the two sets of probability measures are convex.

In the case of Kullback–Leibler divergence, note that if Type 1 (calibration error) and 2 (recalibration) model risk involve independent Radon–Nikodym derivatives, then, in the first case considered above, aggregate model risk equals the sum of the two components. In fact, the Radon–Nikodym derivatives, as random variables, take the key role in evaluating the two types of model risk. At the time the model is recalibrated, we again consider the optimisation (3.2.4), with S_Q now changed to reflect the change in observed market prices, so we have the following Radon–Nikodym derivatives:

$$\text{For calibration error:} \quad m_1 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}} \quad (3.2.12)$$

$$\text{For model risk due to recalibration:} \quad m_2 = \frac{dP_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}}{dP_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}} \quad (3.2.13)$$

$$\text{For aggregate model risk:} \quad m_1 m_2 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}} \quad (3.2.14)$$

Abbreviating $dP_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}$ as P and $dP_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}$ as P^* , the aggregate risk can be expressed in terms of m_1 and m_2 as:

$$\eta_3 = E^P[m_1 m_2 \ln(m_1 m_2)] \quad (3.2.15)$$

$$= E^{P^*}[m_1 \ln(m_1)] + E[m_1 m_2 \ln(m_2)] \quad (3.2.16)$$

$$= \eta_1 + [1 - \text{cov}^P(m_1, m_2)]\eta_2 + \text{cov}^P[m_1, m_2 \ln(m_2)] \quad (3.2.17)$$

If m_1 and m_2 are independent, $\text{cov}^P(m_2, m_1) = \text{cov}^P[m_2, m_1 \ln(m_1)] = 0$. The total model risk is equal to the sum of the calibration risk and the recalibration risk. Surprisingly, in our empirical exploration below we found that this equality is followed quite well by the Black–Scholes model. However, it typically does not hold in the Heston model, suggesting substantial dependence (of Radon–Nikodym derivatives) between the calibration error and model risk due to recalibration.

We also consider models which involve one or more latent state variables. An example of that is the class of stochastic volatility models where the volatility is taken as a latent state variable rather than a model parameter (in the empirical examples below, we specifically consider the model of Heston (1993), which falls into this category). Under the framework of a single stochastic volatility state variable, a model specified by a given set of parameters forms a one-dimensional manifold (Fig. 3.2.1(b)) for possible realisations of the state variable, rather than a point in the Black–Scholes world (Fig. 3.2.1(a)).

Thus, the model which we are now considering is Markovian in a vector of state variables (X, V) , where the state variables X are observable and the state variables V are latent (unobservable). Then, the initial calibration problem (3.2.4) becomes

$$(Q^*, v^*, \theta^*) = \arg \min_{Q \in S_Q, v \in S_v, \theta \in S_\theta} D(Q || P(v; \theta)) \quad (3.2.18)$$

where S_v and S_θ are the sets of legitimate values of the state variables and the parameters, respectively. θ^* is the set of model parameters calibrated to the market, and

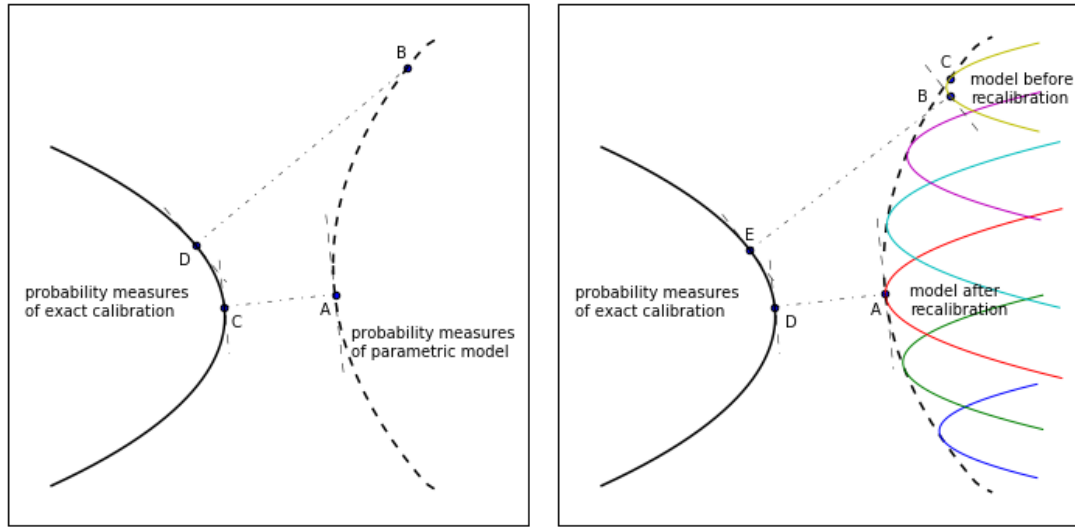


Figure 3.2.1: Graphic illustration of the mathematical definitions of model risks for (a) observable state variable, (b) latent state variable.

v^* is the best estimate of the latent state variables under the calibrated model.¹² The notation in (3.2.6) is amended to

$$P_{t_{i-1}, X_{t_{i-1}}(\omega), v_{t_{i-1}}^*, \theta_{t_{i-1}}^*}(A) \quad \forall A \in \mathcal{F}_T \quad (3.2.19)$$

At time t_i , we have for the calibration error

$$\eta_1 = D(Q_{t_i} || P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*}) \quad (3.2.20)$$

The model risk due to recalibration is

$$\eta_2 = D(P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*} || P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_{i-1}}^*}) \quad (3.2.21)$$

The aggregate model risk using Q_{t_i} from Q^* and $v_{t_i}^*$ from v^* in (3.2.18) is

$$\eta_3 = D(Q_{t_i} || P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_{i-1}}^*}) \quad (3.2.22)$$

or alternatively, using \hat{Q}_{t_i} and \hat{v}_{t_i} determined analogously to (3.2.10), i.e.,

$$(\hat{Q}_{t_i}, \hat{v}_{t_i}) = \arg \min_{Q \in S_Q, v \in S_v} D(Q || P_{t_i, X_{t_i}(\omega), v, \theta_{t_{i-1}}^*}) \quad (3.2.23)$$

¹²This effectively treats the latent (unobserved) state variable as an additional parameter to be calibrated, but the recalibration of which does not contribute to (Type 2) model risk due to recalibration, because it is consistent with the model assumptions for this latent state variable to evolve stochastically. This does shift Type 2 model risk to Type 3, the risk that the state variable dynamics are not (econometrically) consistent with the dynamics assumed in the model. However, in the present paper we deliberately set aside Type 3 model risk for the purposes of our analysis, leaving the integration of all four types of model risk for future research.

which results in

$$\hat{\eta}_3 = D(\hat{Q}_{t_i} || P_{t_i, X_{t_i}(\omega), \hat{v}_{t_i}, \theta_{t_i-1}^*}) \quad (3.2.24)$$

We then have the following Radon–Nikodym derivatives:

$$\text{For calibration error:} \quad m_1 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*}} \quad (3.2.25)$$

$$\text{For model risk due to recalibration:} \quad m_2 = \frac{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*}}{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i-1}^*}} \quad (3.2.26)$$

$$\text{For aggregate model risk:} \quad m_1 m_2 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i-1}^*}} \quad (3.2.27)$$

Note that the key difference between (3.2.12)–(3.2.14) and (3.2.25)–(3.2.27) is that the change in v , being permitted by the model assumptions, does not contribute to the model risk quantities. In (3.2.4) and (3.2.18), we are deliberately prioritising the minimisation of calibration error, as this is congruent to the (often exclusive) focus of practitioners on calibration error (with little or no regard to model risk due to recalibration). If desired, one could reformulate this approach to prioritise the minimisation of aggregate model risk, or of model risk due to recalibration.

3.3 Numerical implementation

In this section, we outline the numerical scheme for solving the minimisation problems arising when taking into account calibration error and model risk due to recalibration in the manner described in the previous section, including problems of the type (3.2.4) involving the optimal choice of two probability measures. In this case, an iterative process is required, optimising two probability measures Q and P in turn until convergence, in the following manner:

- 1) Produce $P^{(0)}$ from a parametric model based on an initial guess of the model parameters (and latent state variables, where required).
- 2) Solve for $Q^{(0)}$ via Lagrange multipliers for the constrained problem that minimises $D(Q^{(0)} || P^{(0)})$.
- 3) Solve for $P^{(1)}$ to obtain model parameters for the $P^{(1)}$ that minimises $D(Q^{(0)} || P^{(1)})$.
- 4) Iterate steps 2 and 3: $P^{(n)} \rightarrow Q^{(n)} \rightarrow P^{(n+1)}$ until convergence.

In Step 1, the initial guess may be obtained in several different ways. A common way is to minimise the mean-squared error between model and market option prices at all available strikes. We opted for the Broyden/Fletcher/Goldfarb/Shanno (BFGS) algorithm for conducting this initial calibration of the model parameters and (where required) latent state variables.

In Step 2, we solve the following constrained minimisation problem using Lagrange multipliers:

$$Q^{(0)} = \arg \min_{Q \in \mathcal{S}_Q} D(Q || P^{(0)}) \quad (3.3.1)$$

$$s.t. \mathbf{B} \leq E^Q(\mathbf{Z}) \leq \mathbf{A} \quad (3.3.2)$$

Note that here we specify the constraints in the form of expectations under the measure Q , where these expectation are the model prices for our calibration instruments for the model based on the non-parametric reference distribution Q . In general, \mathbf{B} , \mathbf{Z} and \mathbf{A} are vectors; thus (3.3.2) is a “stack” of inequality constraints representing observed market prices. Also notice that for generality we “relax” each equality constraint into two inequality constraints. This is in order to account for the bid-ask spread of each option traded on the market. The vector \mathbf{B} denotes a list of bid prices while the vector \mathbf{A} contains ask prices. In a simplified scenario, where exact option prices are given, we may set $\mathbf{B} = \mathbf{A}$. \mathbf{Z} denotes the vector of discounted option payoffs. By introducing vectors of Lagrange multipliers λ_B and λ_A , we convert the constrained problem to an unconstrained dual problem,

$$\max_{\lambda_B \geq 0, \lambda_A \leq 0} \min_{Q \in \mathcal{S}_Q} D(Q || P^{(0)}) - \lambda_B [E^Q(\mathbf{Z}) - \mathbf{B}] - \lambda_A [E^Q(\mathbf{Z}) - \mathbf{A}] \quad (3.3.3)$$

In the case of Kullback-Leibler divergence, solving the inner problem gives the probability density function $q^{(0)}$ of $Q^{(0)}$ in terms of the density $p^{(0)}$ of $P^{(0)}$,

$$q^{(0)} = \frac{p^{(0)} e^{\lambda \mathbf{Z}}}{E^{P^{(0)}}(e^{\lambda \mathbf{Z}})} \quad (3.3.4)$$

Substituting (3.3.4) into (3.3.3), we get a maximisation problem,

$$\max_{\lambda_B \geq 0, \lambda_A \leq 0} -\ln E^{P^{(0)}}(e^{(\lambda_B + \lambda_A)\mathbf{Z}}) + (\lambda_B + \lambda_A) \frac{\mathbf{B} + \mathbf{A}}{2} - (\lambda_B - \lambda_A) \frac{\mathbf{A} - \mathbf{B}}{2} \quad (3.3.5)$$

If $\mathbf{B} = \mathbf{A}$, then the last term vanishes, representing the problem with exact market prices. If (component-wise) $\mathbf{B} > \mathbf{A}$, then the last term reflects a penalty on the objective function that is proportional to the difference of the two Lagrange multipliers. We may therefore transform the Lagrange multipliers by

$$\lambda_+ = \lambda_B + \lambda_A \quad (3.3.6)$$

$$\lambda_- = \lambda_B - \lambda_A \quad (3.3.7)$$

and the objective function becomes

$$\max_{\lambda_+ \in \mathbb{R}^d, \lambda_- \geq |\lambda_+|} -\ln E^{P^{(0)}}(e^{\lambda_+ \mathbf{Z}}) + \lambda_+ \frac{\mathbf{B} + \mathbf{A}}{2} - \lambda_- \frac{\mathbf{A} - \mathbf{B}}{2} \quad (3.3.8)$$

$$= \max_{\lambda_+ \in \mathbb{R}^d} -\ln E^{P^{(0)}}(e^{\lambda_+ \mathbf{Z}}) + \lambda_+ \frac{\mathbf{B} + \mathbf{A}}{2} - |\lambda_+| \frac{\mathbf{A} - \mathbf{B}}{2} \quad (3.3.9)$$

We may numerically solve the maximization problem by taking its gradient with respect to λ_+ ,

$$E^{P^{(0)}}(\mathbf{Z}e^{\lambda_+\mathbf{Z}}) = \left[\frac{B+A}{2} - \text{sgn}(\lambda_+) \frac{A-B}{2} \right] E^{P^{(0)}}(e^{\lambda_+\mathbf{Z}}) \quad (3.3.10)$$

where the element-wise sign function $\text{sgn}(\lambda_+)$ assigns 1, -1 or 0 to each element of λ_+ . However, due to the discontinuity of the sign function (3.3.10) cannot be solved directly in a stable way. To bypass this problem, we approximate the sign function with a continuous step function:

$$\text{sgn}(x) = 1 - \frac{2}{1 + \exp\left(\frac{x}{\delta}\right)} \quad (3.3.11)$$

We use Powell's hybrid method to solve the multidimensional equations (3.3.10), where δ controls the steepness of the function and carefully choosing this value is critical for a fast and stable convergence of the method.

In Step 3, we use L-BFGS-B algorithm to minimise the divergence with respect to model parameters (or latent variables or both). Step 2 and Step 3 are repeated until convergence. The convergence criterion adopted here is that all the percentage changes of parameters after one iteration do not exceed a certain threshold, say 0.1%.

3.4 Examining the trade-off between calibration error and model risk due to recalibration

As an application example of the method described in the previous two sections, we consider historical data consisting of daily market prices for call options on AAPL and GOOG stock over a period from 6 January 2004 to 19 December 2008 for AAPL and 4 January 2005 to 19 December 2008 for GOOG. This gives us a reasonably straightforward application example free of extraneous complications,¹³ while still covering reasonably liquid options and including a period of "interesting" market volatility (2007/8). From this data, we remove options very far away from the money, restricting the range of strikes from delta 2.5% to delta 97.5%. Furthermore, we remove prices of options which had zero trading volume on a given day, in order to avoid using prices which are likely to be stale.

On this data we consider two parametric models, Black and Scholes (1973) and Heston (1993) – arguably the two most popular option pricing models available, where the latter introduces a latent variable for stochastic volatility. The unified methodology, quantifying calibration error, model risk due to recalibration, and the aggregate of the two, allows us to explore the trade-off between calibration error (which is, unsurprisingly, reduced by moving from Black and Scholes (1973) to Heston (1993)) and model

¹³Although these options are of the American type, i.e. permitting early exercise, AAPL and GOOG did not pay any dividends during this period. Thus the possibility of early exercise may be ignored (see Merton (1973)).

risk due to recalibration (which has hitherto been largely ignored) when moving from one parametric model to another as well as when changing the frequency with which the model is recalibrated.

We start by evaluating the calibration, recalibration and aggregated model risks under a Black/Scholes model, i.e. where the underlying asset price is assumed to follow geometric Brownian motion, with dynamics under the risk-neutral measure given by

$$dS(t) = rS(t)dt + \sigma S(t)dW(t) \quad (3.4.1)$$

where r is the continuously compounded risk-free rate of interest and σ is a constant volatility parameter. We note that in the Black/Scholes model we obtain a simple closed form expression for the recalibration risk defined in (3.2.8):

$$\eta_2 = \left(\frac{\sigma^{*2}}{\sigma^2} - 1 \right) \left[\frac{1}{2} + \frac{T}{8}(\sigma^{*2} - \sigma^2) \right] + \ln \left(\frac{\sigma}{\sigma^*} \right) \quad (3.4.2)$$

where σ^* is the correctly recalibrated Black/Scholes volatility parameter and σ is the parameter value obtained in a previous calibration. This formula is a consequence of the log-normal distribution of returns assumed in the Black/Scholes model.

We can express the aggregate model risk as the sum of the calibration error, the recalibration risk and a residual. As noted in equation (3.2.15), the residual is zero if the likelihood ratios involved in the calibration and recalibration risks are two independent random variables. In practice, the residual usually takes a small non-zero value. In Figure 3.4.1 we demonstrate the decomposition of the total model risk into the three components.¹⁴ Unsurprisingly (as it is well documented that the Black/Scholes model cannot fit the implied volatility “smile” observed in most options markets), we see that calibration error typically predominates.

In the Heston model, the dynamics (3.4.1) are extended to allow for stochastic volatility, i.e.

$$dS(t) = rS(t)dt + \sigma S(t)dW_1(t) \quad (3.4.3)$$

$$d(\sigma^2(t)) = \kappa(\theta - \sigma^2(t))dt + \eta\sigma(t)dW_2(t) \quad (3.4.4)$$

This model involves two state variables, the underlying asset price $S(t)$ and the volatility $\sigma(t)$, and five model parameters: r , κ , θ , η , ρ where ρ is the correlation coefficient between the two Wiener processes:

$$d\langle W_1, W_2 \rangle_t = \rho dt$$

r is the risk-free rate,¹⁵ and κ , θ and η relate to the volatility process, κ being the rate of mean reversion, θ the long-run mean and η the volatility of this process.

¹⁴The vertical axis denotes the numerical value of the relative entropy.

¹⁵In our empirical application examples, we take the risk-free rate as one of the financial variables observed in the market, but we do not explicitly take into account interest rate risk in our empirical analysis. For the short-dated options considered here, interest rate risk is known to be of relatively little importance — for a discussion of this issue, see e.g. Cheng, Nikitopoulos and Schlögl (2017) and the literature cited therein.

Following Gatheral (2006), the risk-neutral probability of exercise of a European call option with strike K in the Heston model is given by

$$P_0(x, v, \tau) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left[\frac{\exp\{C(u, \tau)\theta + D(u, \tau)v + iux\}}{iu} \right] du, \quad (3.4.5)$$

where v is the current value of the volatility state variable $\sigma(t)$, $\tau = T - t$ is the time to maturity, and x is the logarithmic forward moneyness of the option, i.e.

$$x : = \ln \left(\frac{F_{t,T}}{K} \right)$$

$$F_{t,T} = \frac{S(t)}{B(t, T)}$$

with $B(t, T)$ the time t price of a zero bond maturing in T . Furthermore,

$$C(u, \tau) = \kappa \left[r_- \tau - \frac{2}{\eta^2} \ln \left(\frac{r_+ - r_- e^{-d\tau}}{r_+ - r_-} \right) \right] \quad (3.4.6)$$

$$D(u, \tau) = r_- \frac{r_+ - r_+ e^{-d\tau}}{r_+ - r_- e^{-d\tau}} \quad (3.4.7)$$

$$r_\pm = \frac{\beta \pm d}{\eta^2} \quad d = \sqrt{\beta^2 - 2\alpha\eta^2} \quad (3.4.8)$$

Parameters α and β are functions of u (Fourier transform variable of x):

$$\alpha(u) = -\frac{u^2}{2} - i\frac{u}{2} \quad (3.4.9)$$

$$\beta(u) = \kappa - \rho\eta iu \quad (3.4.10)$$

It is noted that $P_0 = E^{\mathbb{Q}}(\mathbb{1}_{S_T > K})$ since by definition P_0 is the probability of exercise. The probability density function of the risk-neutral measure is therefore obtained:

$$p(S_T) = - \frac{\partial P_0}{\partial K} \Big|_{K=S_T} = - \frac{\partial P_0}{\partial x} \frac{\partial x}{\partial K} \Big|_{K=S_T} \quad (3.4.11)$$

$$= \frac{1}{2\pi S_T} \int_{-\infty}^\infty \exp \left[C(u, \tau)\theta + D(u, \tau)v + iu \ln \left(\frac{S_t}{B(t, T)S_T} \right) \right] du \quad (3.4.12)$$

For simplicity e^y denotes the ratio of the forward price at t to its spot price at maturity T , i.e. $e^y = B(t, T)S_T/S_t$, we derive the risk-neutral probability with respect to y :

$$p(y) = p(S_T) \left| \frac{\partial y}{\partial S_T} \right|^{-1} \quad (3.4.13)$$

$$= \frac{1}{2\pi} \int_{-\infty}^\infty \exp [C(u, \tau)\theta + D(u, \tau)v - iuy] du \quad (3.4.14)$$

$p(y)$ can be calculated by fast Fourier transform (FFT).

Aggregate Model Risk

Risk measure	Black-Scholes					Heston					
	1 day	3 days	1 week	2 weeks	1 quarter	1 day	3 days	1 week	2 weeks	1 quarter	
mean	0.070	0.071	0.073	0.073	0.085	0.037	0.035	0.038	0.039	0.046	
median	0.045	0.047	0.047	0.051	0.057	0.005	0.005	0.005	0.006	0.010	
Quantile	99%	0.474	0.427	0.471	0.455	0.462	0.508	0.462	0.512	0.503	0.649
	95%	0.212	0.221	0.221	0.212	0.251	0.173	0.169	0.177	0.177	0.185
	90%	0.158	0.160	0.165	0.160	0.192	0.096	0.097	0.098	0.105	0.121
	75%	0.092	0.094	0.095	0.099	0.112	0.027	0.027	0.029	0.031	0.034

Calibration Error

mean	0.055	0.066	0.069	0.072	0.084	0.008	0.026	0.032	0.036	0.045	
median	0.038	0.043	0.045	0.049	0.057	0.001	0.003	0.004	0.005	0.009	
Quantile	99%	0.239	0.397	0.433	0.443	0.455	0.163	0.416	0.496	0.495	0.648
	95%	0.171	0.207	0.212	0.210	0.249	0.015	0.130	0.158	0.168	0.182
	90%	0.135	0.151	0.158	0.158	0.191	0.008	0.065	0.083	0.097	0.119
	75%	0.075	0.087	0.091	0.096	0.111	0.003	0.012	0.019	0.026	0.034

Model Risk due to Recalibration

mean	0.024	0.009	0.005	0.002	0.001	0.057	0.019	0.011	0.006	0.001	
median	0.002	0.001	0.001	0.000	0.000	0.004	0.001	0.001	0.001	0.000	
Quantile	99%	0.458	0.196	0.057	0.030	0.012	0.630	0.226	0.117	0.067	0.011
	95%	0.103	0.034	0.021	0.010	0.005	0.315	0.108	0.062	0.032	0.006
	90%	0.056	0.020	0.011	0.006	0.002	0.164	0.052	0.031	0.016	0.004
	75%	0.013	0.006	0.004	0.002	0.001	0.055	0.016	0.011	0.005	0.001

Table 3.4.1: Model risks (in terms of relative entropy) under different models and recalibration frequencies for AAPL.

The decomposition of the total model risk into the three components (the components due calibration and recalibration, and the positive or negative residual measuring the departure from independence between the first two components) when using the Heston model as the baseline is given in Figure 3.4.2. Again, since it is well documented that the Heston model can fit observed option prices better than Black/Scholes, it is unsurprising that in this case the relative entropy measuring calibration error is much lower – however, already in this set of example days it is evident that this comes at a price of increased model risk due to recalibration.

These observations are reinforced when we consider aggregate model risk, calibration error and model risk due to recalibration over the entire sample period, as presented in Tables 3.4.1 and 3.4.2. Note that the absolute numbers refer to relative entropy and thus lack direct financial interpretation – what matters are the relative values when comparing the model across models and different recalibration frequencies, in particular when considering the aggregate model risk. Here, we consider recalibrating the Black/Scholes and Heston models either daily, every three days, every week, every two weeks, or every quarter year. We see that recalibrating more frequently has little effect on the aggregate model risk, neither when using the Black/Scholes model nor when using the Heston model. Essentially, recalibrating more frequently simply shifts calibration error into model risk due to recalibration,¹⁶ highlighting the dan-

¹⁶Note that on days on which we do not recalibrate, the model risk due to recalibration is zero, because

Aggregate Model Risk

Risk measure	Black-Scholes					Heston					
	1 day	3 days	1 week	2 weeks	1 quarter	1 day	3 days	1 week	2 weeks	1 quarter	
mean	0.165	0.168	0.165	0.165	0.188	0.115	0.105	0.113	0.114	0.127	
median	0.109	0.111	0.112	0.115	0.138	0.053	0.050	0.055	0.053	0.059	
Quantile	99%	0.705	0.722	0.722	0.699	0.728	0.726	0.668	0.711	0.699	0.744
	95%	0.519	0.530	0.521	0.496	0.549	0.475	0.438	0.455	0.467	0.560
	90%	0.393	0.391	0.387	0.385	0.408	0.330	0.287	0.320	0.324	0.380
	75%	0.219	0.223	0.214	0.215	0.256	0.145	0.137	0.142	0.153	0.157

Calibration Error

mean	0.125	0.155	0.158	0.162	0.187	0.055	0.083	0.103	0.107	0.126	
median	0.082	0.102	0.106	0.111	0.137	0.006	0.022	0.043	0.046	0.058	
Quantile	99%	0.655	0.701	0.709	0.696	0.728	0.626	0.651	0.706	0.688	0.744
	95%	0.400	0.482	0.507	0.494	0.548	0.329	0.381	0.438	0.445	0.557
	90%	0.295	0.359	0.362	0.379	0.408	0.172	0.250	0.298	0.306	0.375
	75%	0.171	0.208	0.207	0.212	0.255	0.034	0.103	0.127	0.141	0.155

Model Risk due to Recalibration

mean	0.036	0.014	0.008	0.004	0.001	0.119	0.044	0.012	0.012	0.003	
median	0.004	0.002	0.001	0.001	0.000	0.043	0.016	0.003	0.003	0.001	
Quantile	99%	0.405	0.172	0.091	0.050	0.010	0.757	0.255	0.087	0.077	0.012
	95%	0.173	0.066	0.035	0.018	0.004	0.543	0.187	0.056	0.064	0.011
	90%	0.109	0.038	0.022	0.012	0.003	0.411	0.151	0.034	0.045	0.009
	75%	0.036	0.013	0.008	0.004	0.001	0.142	0.052	0.016	0.013	0.003

Table 3.4.2: Model risks (in terms of relative entropy) under different models and recalibration frequencies for GOOG.

gers in the common practice of focusing solely on the calibration of derivative pricing models, at the expense of all other sources of model risk.

In addition, we observe that if we are interested in “robustness” at a high level of confidence (looking at, say, the 99% quantile of aggregate model risk), moving from Black/Scholes to Heston also does not appear to deliver any advantage (it does yield some improvement at lower quantiles, or average or median, aggregate model risk). This means that when high levels of confidence are required, any gain in calibration accuracy delivered by the Heston model is offset by higher model risk due to recalibration. One should note that this last point holds even *before* considering Type 3 model risk, which may well be worse when additional state variables are introduced (as in the Heston model). For these results, in Tables 3.4.1 and 3.4.2 we calculated means, medians and quantiles across all available option maturities. If we consider only particular maturity “buckets”, the same qualitative conclusions are evident – Tables 3.4.3 and 3.4.4 illustrate this in the case of daily recalibration.

(consistent with the model assumptions) we are keeping previously calibrated parameters unchanged – so on those days aggregate model risk is entirely due to calibration error (which increases because the fit to market prices deteriorates when we do not recalibrate).

Aggregate Model Risk									
Risk measure	Black-Scholes				Heston				
	all	0-0.2 year	0.2-0.7 year	>0.7 year	all	0-0.2 year	0.2-0.7 year	>0.7 year	
mean	0.070	0.041	0.066	0.109	0.037	0.039	0.024	0.046	
median	0.045	0.021	0.052	0.081	0.005	0.004	0.004	0.008	
Quantile	99%	0.474	0.471	0.266	0.567	0.508	0.590	0.302	0.472
	95%	0.212	0.120	0.185	0.282	0.173	0.243	0.091	0.191
	90%	0.158	0.081	0.143	0.213	0.096	0.074	0.066	0.142
	75%	0.092	0.047	0.090	0.141	0.027	0.021	0.015	0.054
Calibration Error									
mean	0.055	0.026	0.057	0.087	0.008	0.012	0.004	0.007	
median	0.038	0.014	0.048	0.068	0.001	0.001	0.001	0.001	
Quantile	99%	0.239	0.157	0.230	0.289	0.163	0.072	0.052	0.116
	95%	0.171	0.091	0.157	0.212	0.015	0.021	0.011	0.017
	90%	0.135	0.062	0.126	0.181	0.008	0.006	0.007	0.010
	75%	0.075	0.036	0.077	0.128	0.003	0.002	0.003	0.003
Model Risk due to Recalibration									
mean	0.024	0.020	0.013	0.038	0.057	0.062	0.048	0.060	
median	0.002	0.002	0.001	0.003	0.004	0.004	0.003	0.009	
Quantile	99%	0.458	0.506	0.116	0.662	0.630	0.702	0.536	0.512
	95%	0.103	0.066	0.066	0.169	0.315	0.443	0.274	0.247
	90%	0.056	0.030	0.042	0.102	0.164	0.223	0.127	0.160
	75%	0.013	0.009	0.010	0.031	0.055	0.030	0.048	0.091

Table 3.4.3: Model risks (in terms of relative entropy) under different models, by maturity buckets, for AAPL (daily recalibration frequency).

3.5 Conclusion

Under our approach, less relative entropy implies less model risk, and we are able to evaluate two hitherto disparate sources of model risk (calibration error and model risk due to recalibration) in a unified fashion, and examine the potential trade-off between the two. We have considered a simple choice between two models, and between different recalibration frequencies. “Putting a number on model risk” by calculating quantiles for the maximum model risk (quantified by relative entropy) over a time series of market data allows one to assess the added value (if any) of more complicated stochastic models of financial markets.

In our application, we are deliberately prioritising the minimisation of calibration error, as this is congruent to the (often exclusive) focus of practitioners on calibration error (with little or no regard to model risk due to recalibration).¹⁷ Even in this case, we see that by including recalibration as one of the sources of aggregate model risk, recalibrating a model frequently to a changing market simply interchanges one source of model risk for another, and more complicated stochastic models may well underperform when aggregate model risk is taken into account.

The approach is based on the path independence of option pricing, thus allowing

¹⁷If desired, one could reformulate this approach to prioritise the minimisation of aggregate model risk, or of model risk due to recalibration.

Aggregate Model Risk

Risk measure	Black-Scholes				Heston				
	all	0-0.2 year	0.2-0.7 year	>0.7 year	all	0-0.2 year	0.2-0.7 year	>0.7 year	
mean	0.165	0.167	0.162	0.164	0.115	0.115	0.120	0.109	
median	0.109	0.104	0.099	0.120	0.053	0.052	0.057	0.048	
Quantile	99%	0.705	0.736	0.709	0.661	0.726	0.714	0.707	0.736
	95%	0.519	0.527	0.551	0.459	0.475	0.468	0.500	0.476
	90%	0.393	0.417	0.382	0.371	0.330	0.351	0.322	0.304
	75%	0.219	0.221	0.216	0.218	0.145	0.149	0.155	0.136

Calibration Error

mean	0.125	0.121	0.123	0.132	0.055	0.055	0.052	0.057	
median	0.082	0.074	0.075	0.099	0.006	0.005	0.006	0.006	
Quantile	99%	0.655	0.662	0.650	0.639	0.626	0.658	0.531	0.665
	95%	0.400	0.403	0.423	0.391	0.329	0.349	0.311	0.330
	90%	0.295	0.298	0.301	0.285	0.172	0.155	0.185	0.171
	75%	0.171	0.162	0.160	0.176	0.034	0.031	0.034	0.034

Model Risk due to Recalibration

mean	0.036	0.038	0.039	0.033	0.119	0.117	0.117	0.124	
median	0.004	0.005	0.003	0.005	0.043	0.043	0.037	0.049	
Quantile	99%	0.405	0.419	0.479	0.284	0.757	0.746	0.767	0.754
	95%	0.173	0.176	0.188	0.155	0.543	0.510	0.558	0.554
	90%	0.109	0.111	0.115	0.098	0.411	0.390	0.418	0.418
	75%	0.036	0.038	0.032	0.036	0.142	0.144	0.137	0.146

Table 3.4.4: Model risks (in terms of relative entropy) under different models, by maturity buckets, for GOOG (daily recalibration frequency).

for the comparison of terminal price distributions. This makes it possible to compare between different types of models, e.g. Black-Scholes model and stochastic volatility models. For other model applications such as hedging, path dependency needs to be incorporated. Therefore different types of models, such as stochastic volatility models, local volatility models and jump-diffusion models, can have very different model risks.

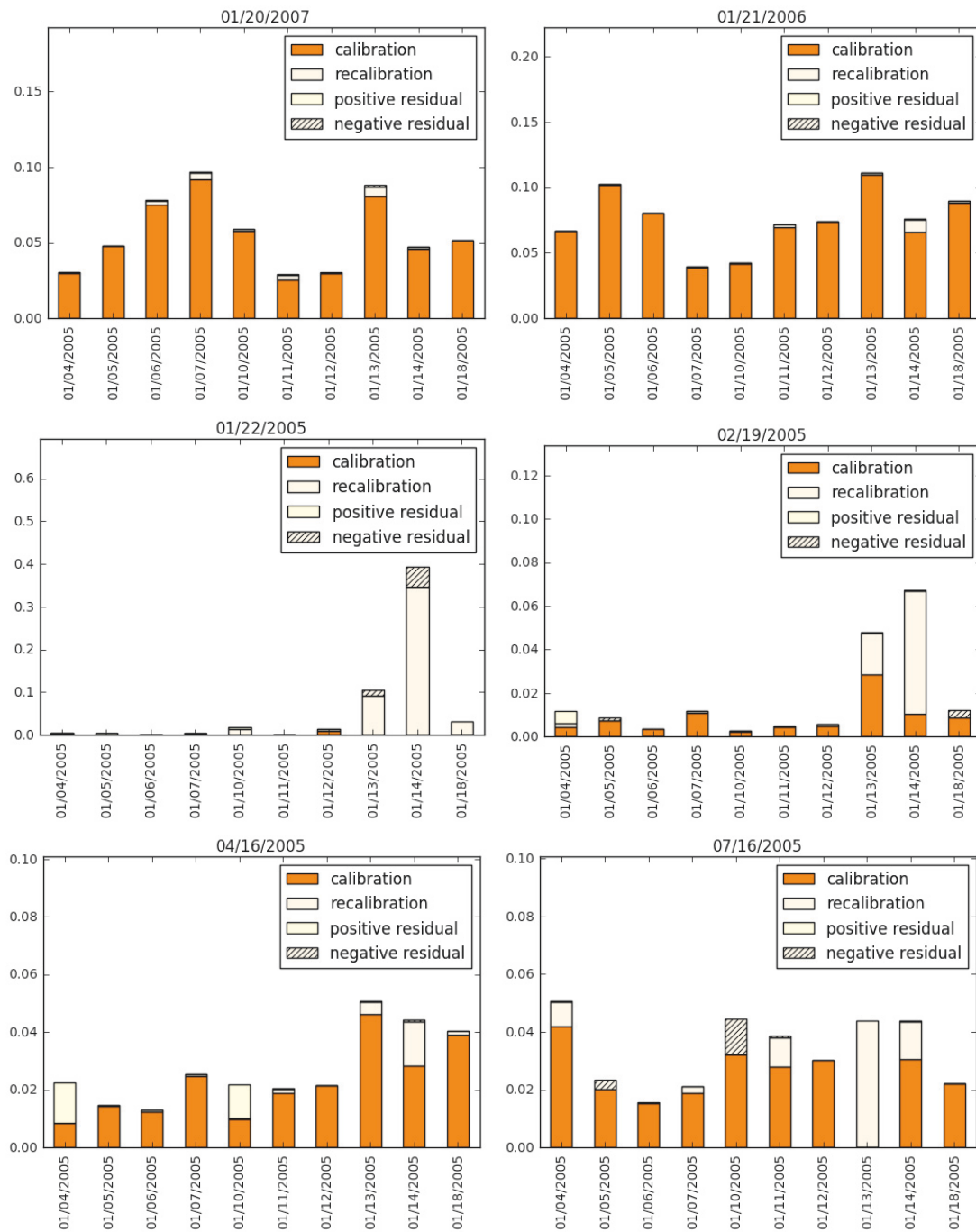


Figure 3.4.1: Decomposition of model risks of AAPL options using the Black/Scholes model as the nominal model, for a selection of option maturity dates (given as the title of each graph).

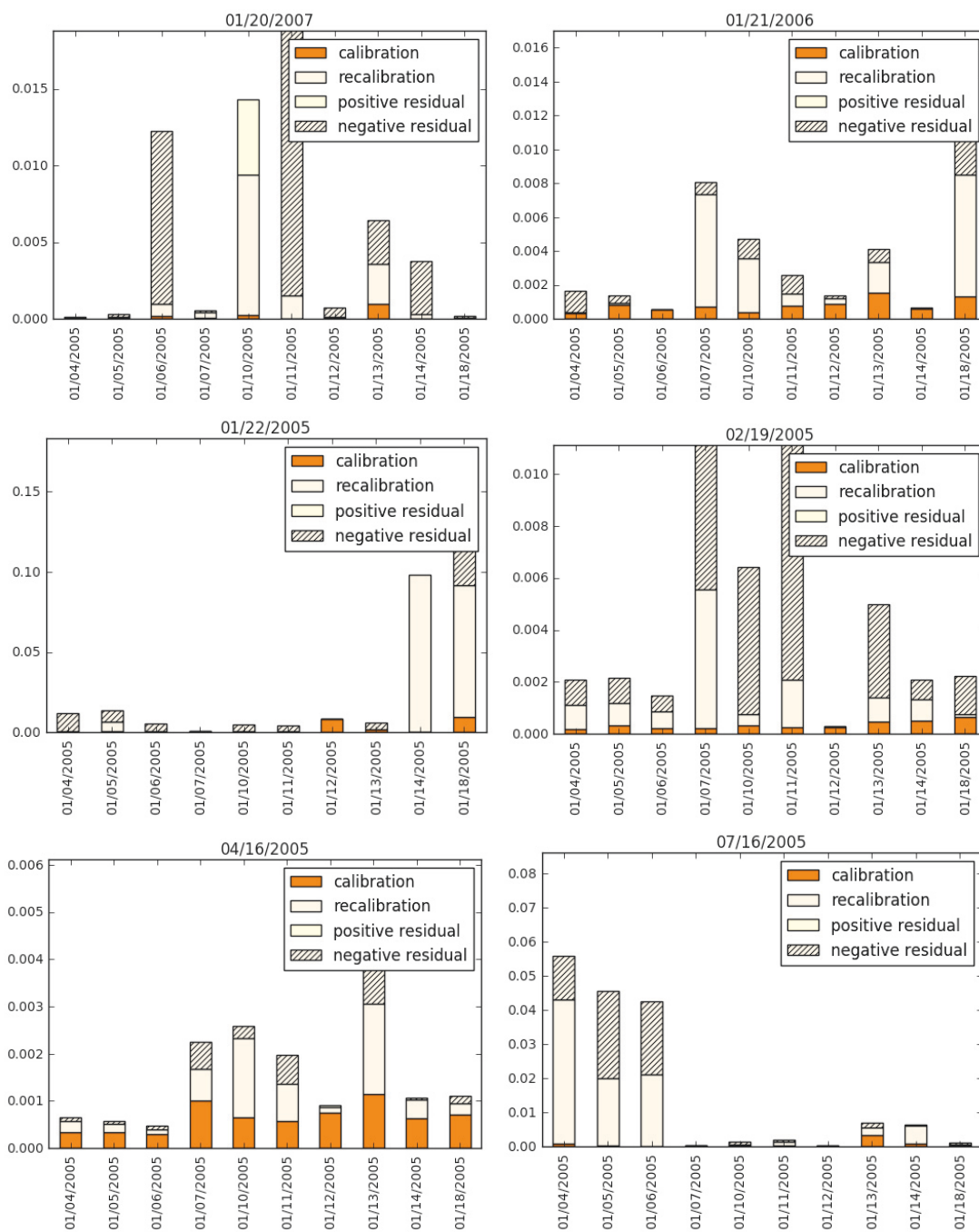


Figure 3.4.2: Decomposition of model risks of GOOG options using the Heston model as the nominal model, for a selection of option maturity dates (given as the title of each graph).

Chapter 4

Theory of Dynamic Model Risk Measurement

As a working definition, model risk refers to the quantification of unanticipated losses resulting from the use of inappropriate models to value and manage financial securities, including widely traded securities like stocks and bonds, for which market prices are readily available, and less traded derivatives written on such securities. Unlike other financial risks, which are concerned with the impact of randomness within the paradigm of a chosen model, model risk is concerned with the possibility that the wrong modelling paradigm was chosen in the first place. This makes it a much more challenging proposition, both conceptually and in terms of implementation. It is thus unsurprising that model risk continues to languish behind its more traditional counterparts, such as price risk, interest rate risk and credit risk, both in terms of identifying an appropriate theoretical methodology and in the development of specific metrics.

Most approaches to model risk measurement are parametric in the sense that they consider alternative models parametrised by a finite set of parameters. To go beyond that, (Glasserman and Xu 2014) proposed a non-parametric approach. Under this framework, a worst-case model is found among alternative models in a neighborhood of a reference model. Glasserman and Xu adopted the relative entropy (or the Kullback-Leibler divergence) to measure the distance between the probability measure given by the reference model and an (equivalent) alternative measure. By imposing a constraint on the relative entropy budget, the set of legitimate alternative models is defined in a non-parametric fashion, and the worst-case scenario can then be solved analytically within a finite distance to the reference model. This approach is formulated w.r.t the distribution of a state variable, thus less applicable when the state variable evolves dynamically.

In this paper, we apply the entropic approach to the problem of measuring model risk w.r.t a state process. Starting from a nominal process and a loss functional, we characterise the worst-case change of measure that maximises the expected loss. Technically, we solve the problem in a dual formulation and handle its path-dependency with the help of the functional Itô calculus (Cont 2016). The constraint that defines the legitimate alternative models is w.r.t the f -divergence, a more general choice than

the Kullback-Leibler divergence. As a result, the worst-case risk is characterised by a martingale approach or a path-dependent partial differential equation (for continuous semimartingales only).

4.1 Problem Formulation

Fix $T \in (0, \infty)$ and $d \in \mathbb{N}$, and let $\Omega := D([0, T], \mathbb{R}^d)$ denote the set of càdlàg paths $\omega : [0, T] \rightarrow \mathbb{R}^d$. Let $[0, T] \ni t \mapsto X(t)$ be the canonical process on Ω , which means to say that $X(t)(\omega) := \omega(t)$, for all $(t, \omega) \in [0, T] \times \Omega$. Let $\mathfrak{F}^0 = (\mathcal{F}_t^0)_{t \in [0, T]}$ denote the filtration on Ω generated by X , which is to say that

$$\mathcal{F}_t^0 := \bigvee_{s \in [0, t]} \{X(s)^{-1}(U) \mid U \in \mathcal{B}(\mathbb{R}^d)\} = \bigvee_{s \in [0, t]} \bigcup_{U \in \mathcal{B}(\mathbb{R}^d)} \{\omega \in \Omega \mid \omega(s) \in U\},$$

for all $t \in [0, T]$. In particular,

$$\mathcal{F}_0^0 := \{X(0)^{-1}(U) \mid U \in \mathcal{B}(\mathbb{R}^d)\} = \bigcup_{U \in \mathcal{B}(\mathbb{R}^d)} \{\omega \in \Omega \mid \omega(0) \in U\}.$$

Fix a reference probability measure \mathbb{P} on $(\Omega, \mathcal{F}_T^0)$, subject to the condition

$$\mathbb{P}(X(0)^{-1}(U)) = \mathbb{P}(\{\omega \in \Omega \mid \omega(0) \in U\}) = \begin{cases} 1 & \text{if } 0 \in U; \\ 0 & \text{if } 0 \notin U, \end{cases}$$

for all $U \in \mathcal{B}(\mathbb{R}^d)$, which is to say that almost all paths start at zero under \mathbb{P} . Note that this condition ensures that $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, for all $A \in \mathcal{F}_0^0$.

To be consistent with the notation in Cont (2016), we shall write $\omega_t := \omega(t \wedge \cdot) \in \Omega$ to denote the path $\omega \in \Omega$ stopped at time $t \in [0, T]$. We impose an equivalence relation \sim on $[0, T] \times \Omega$, by specifying that

$$(t, \omega) \sim (t', \omega') \quad \text{if and only if} \quad t = t' \quad \text{and} \quad \omega_t = \omega'_{t'},$$

for all $(t, \omega), (t', \omega') \in [0, T] \times \Omega$. That is to say, two pairs, each consisting of a time and a path, are equivalent if the times are equal and the corresponding stopped paths are the same. The quotient set $\Lambda_T^d := [0, T] \times \Omega / \sim$ forms a complete metric space, when endowed with the metric $d_\infty : (\Lambda_T^d)^2 \rightarrow \mathbb{R}_+$, defined by

$$d_\infty((t, \omega), (t', \omega')) := \sup_{s \in [0, T]} |\omega(s \wedge t) - \omega'(s \wedge t')| + |t - t'| = \|\omega_t - \omega'_{t'}\|_\infty + |t - t'|,$$

for all $(t, \omega), (t', \omega') \in \Lambda_T^d$. We refer to (Λ_T^d, d_∞) as the *space of stopped paths*.

A measurable function $F : \Lambda_T^d \rightarrow \mathbb{R}$ is called a *non-anticipative functional*, where Λ_T^d is endowed with the Borel sigma-algebra generated by d_∞ and \mathbb{R} is endowed with the Borel sigma-algebra generated by the usual Euclidean metric. Since $(t, \omega) \sim (t, \omega_t)$, for all $(t, \omega) \in [0, T] \times \Omega$, we may regard a non-anticipative functional $F : \Lambda_T^d \rightarrow \mathbb{R}$ as

an appropriately measurable function $F : [0, T] \times \Omega \rightarrow \mathbb{R}$ that satisfies the condition $F(t, \omega) = F(t, \omega_t)$. That is to say, the value of a non-anticipative functional, when applied to a particular time and path, depends only on the behaviour of the path up to the time. Note that $(F(t, \cdot))_{t \in [0, T]}$ is a progressively measurable process, adapted to the filtration \mathfrak{F}^0 .

Let \mathcal{M} denote the family of (right-continuous versions of) martingales on the filtered probability space $(\Omega, \mathcal{F}_T^0, \mathfrak{F}^0, \mathbb{P})$, over the compact time-interval $[0, T]$, and let

$$\mathcal{M}_+(1) := \{Z \in \mathcal{M} \mid Z \geq 0 \text{ and } Z(0) = 1\}$$

denote the sub-family of non-negative martingales starting at one. Each $Z \in \mathcal{M}_+(1)$ defines a probability measure \mathbb{Q}_Z on $(\Omega, \mathcal{F}_T^0)$ satisfying $\mathbb{Q}_Z \ll \mathbb{P}$ (i.e. \mathbb{Q}_Z is absolutely continuous w.r.t \mathbb{P}), according to the recipe $\mathbb{Q}_Z(A) := \mathbb{E}(\mathbf{1}_A Z(T))$, for all $A \in \mathcal{F}_T^0$. Conversely, each probability measure \mathbb{Q} on $(\Omega, \mathcal{F}_T^0)$ satisfying $\mathbb{Q} \ll \mathbb{P}$ can be written as $\mathbb{Q} = \mathbb{Q}_Z$, where $Z \in \mathcal{M}_+(1)$ is determined by (Meyer 1966)

$$Z(t) := \mathbb{E}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \middle| \mathcal{F}_t^0\right),$$

for all $t \in [0, T]$.

Consider a twice-differentiable strictly convex function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ satisfying $f(1) = 0$. For any probability measure \mathbb{Q} on $(\Omega, \mathcal{F}_T^0)$ satisfying $\mathbb{Q} \ll \mathbb{P}$, the f -divergence of \mathbb{Q} with respect to \mathbb{P} is defined by

$$D_f(\mathbb{Q} \parallel \mathbb{P}) := \mathbb{E}\left(f\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right) \quad (4.1.1)$$

(see Basseville (2013) Section 2). Intuitively, f -divergence provides a measure of the distance between two probability measures. Hence, the set

$$\mathcal{Z}_\eta := \{Z \in \mathcal{M}_+(1) \mid D_f(\mathbb{Q}_Z \parallel \mathbb{P}) \leq \eta\}, \quad (4.1.2)$$

where $\eta \geq 0$, corresponds to the family of absolutely continuous probability measures that are close to the reference probability measure \mathbb{P} .

Finally, fix a non-anticipative functional $\ell : \Lambda_T^d \rightarrow \mathbb{R}$ satisfying $\ell(0, 0) = 0$. We shall interpret $\ell(t, \omega)$ as the cumulative realized loss up to time t , incurred by a portfolio of financial securities. The state of the portfolio is completely determined by the path $\omega \in \Omega$. The condition of the reference probability measure guarantees

$$\mathbb{P}(X(0)^{-1}\{0\}) = \mathbb{P}(\{\omega \in \Omega \mid \omega(0) = 0\}) = 1,$$

It follows that $\ell(0, \cdot) = 0$ P-a.s. That is to say, the initial realized loss incurred by the portfolio is zero under the reference probability measure. If we interpret \mathbb{P} as the probability measure associated with a nominal model for the dynamics of the portfolio, then $\mathbb{E}(\ell(T, \cdot))$ gives the expected total loss under the nominal model. In financial applications, we usually set the terminal time T as the point when the entire portfolio gets liquidated, thus realizing the cumulative loss.

Suppose, now, that there is some uncertainty about which model best describes the portfolio. In particular, suppose that each probability measure determined by a member of \mathcal{Z}_η , for some $\eta \geq 0$, corresponds to a plausible model for the dynamics of the portfolio.¹ In that case, a risk manager would be interested in the following quantities:

$$\sup_{Z \in \mathcal{Z}_\eta} \mathbb{E}^{\mathbb{Q}^Z}(\ell(T, \cdot)) \quad \text{and} \quad \sup_{Z \in \mathcal{Z}_\eta} \mathbb{E}^{\mathbb{Q}^Z}(\ell(T, \cdot)) - \mathbb{E}(\ell(T, \cdot)). \quad (4.1.3)$$

The former expression may be regarded as the worst-case expected loss suffered by the portfolio under all plausible models, while the latter expression quantifies the difference between the worst-case expected loss and the expected loss under the default model. As such, it serves as a measure of model risk.

The problem defined in (4.1.3) may be formulated in a dual form (Glasserman and Xu 2014). We first define the Lagrangian $\mathcal{L} : \mathcal{M}_+(1) \times (0, \infty) \times (0, \infty) \rightarrow \mathbb{R}$ by

$$\mathcal{L}(Z, \vartheta, \eta) := \mathbb{E}^{\mathbb{Q}^Z}(\ell(T, \cdot)) - \frac{D_f(\mathbb{Q}^Z \| \mathbb{P}) - \eta}{\vartheta} = \mathbb{E}^{\mathbb{Q}^Z} \left(\ell(T, \cdot) - \frac{f(Z(T))}{\vartheta Z(T)} \right) + \frac{\eta}{\vartheta},$$

The Lagrangian leads to a dual function defined by

$$d(\vartheta, \eta) := \sup_{Z \in \mathcal{M}_+(1)} \mathcal{L}(Z, \vartheta, \eta) = \sup_{Z \in \mathcal{M}_+(1)} \mathbb{E}^{\mathbb{Q}^Z}(\widehat{\ell}_\vartheta(T, Z)) + \frac{\eta}{\vartheta},$$

where

$$\widehat{\ell}_\vartheta(t, Z) := \ell(t, \cdot) - \frac{f(Z(t))}{\vartheta Z(t)}, \quad (4.1.4)$$

for all $t \in [0, T]$ and all $Z \in \mathcal{M}_+(1)$. As with ℓ , $\widehat{\ell}_\vartheta(\cdot, Z)$ is a non-anticipative functional.

If the primal problem is convex and the constraint (Eq. 4.1.2) satisfies Slater's condition (Slater 2014), then strong duality holds, giving

$$\sup_{Z \in \mathcal{Z}_\eta} \mathbb{E}^{\mathbb{Q}^Z}(\ell(T, \cdot)) = \inf_{\vartheta \in (0, \infty)} d(\vartheta, \eta) \quad (4.1.5)$$

This is proved in the following lemma.

Lemma 4.1.1. *The following statements are true:*

1. The set \mathcal{Z}_η is convex.
2. The function $\mathcal{Z}_\eta \ni Z \mapsto \mathbb{E}^{\mathbb{Q}^Z}(\ell(T, \cdot))$ is convex.
3. Strong duality Eq. 4.1.5 holds.
4. Given $\vartheta^* \in (0, \infty)$, suppose that $Z^* \in \mathcal{M}_+(1)$ satisfies

$$Z^* = \arg \max_{Z \in \mathcal{M}_+(1)} \mathbb{E}^{\mathbb{Q}^Z}(\widehat{\ell}_{\vartheta^*}(T, Z)).$$

¹The idea here is that all absolutely continuous probability measures close enough to the reference measure (in the sense of f -divergence) correspond with models that are plausibly close to the reference model.

Then

$$Z^* = \arg \max_{Z \in \mathcal{Z}_\eta} \mathbf{E}^{\mathbf{Q}_Z}(\ell(T, \cdot)),$$

with $\eta := \mathbf{E}(f(Z^*(T)))$.²

Proof. (1) Given $Z_1, Z_2 \in \mathcal{Z}_\eta$, observe that

$$\begin{aligned} D_f(\mathbf{Q}_{\lambda Z_1 + (1-\lambda)Z_2} \| \mathbf{P}) &= D_f(\lambda \mathbf{Q}_{Z_1} + (1-\lambda) \mathbf{Q}_{Z_2} \| \mathbf{P}) = \mathbf{E}(f(\lambda Z_1(T) + (1-\lambda)Z_2(T))) \\ &\leq \mathbf{E}(\lambda f(Z_1(T)) + (1-\lambda)f(Z_2(T))) \\ &= \lambda \mathbf{E}(f(Z_1(T))) + (1-\lambda) \mathbf{E}(f(Z_2(T))) \\ &= \lambda D_f(\mathbf{Q}_{Z_1} \| \mathbf{P}) + (1-\lambda) D_f(\mathbf{Q}_{Z_2} \| \mathbf{P}), \end{aligned}$$

for all $\lambda \in [0, 1]$, by virtue of the convexity of f and Jensen's inequality. Since $D_f(\mathbf{Q}_{Z_1} \| \mathbf{P}) \leq \eta$ and $D_f(\mathbf{Q}_{Z_2} \| \mathbf{P}) \leq \eta$, the inequality above leads to $D_f(\mathbf{Q}_{\lambda Z_1 + (1-\lambda)Z_2} \| \mathbf{P}) \leq \eta$. This implies that $\lambda Z_1 + (1-\lambda)Z_2 \in \mathcal{Z}_\eta$, by virtue of the fact that $\lambda Z_1 + (1-\lambda)Z_2 \in \mathcal{M}_+(1)$.

(2) Given $Z_1, Z_2 \in \mathcal{Z}_\eta$, observe that

$$\begin{aligned} \mathbf{E}^{\mathbf{Q}_{\lambda Z_1 + (1-\lambda)Z_2}}(\ell(T, \cdot)) &= \mathbf{E}((\lambda Z_1(T) + (1-\lambda)Z_2(T))\ell(T, \cdot)) \\ &= \lambda \mathbf{E}(Z_1(T)\ell(T, \cdot)) + (1-\lambda) \mathbf{E}(Z_2(T)\ell(T, \cdot)) \\ &= \lambda \mathbf{E}^{\mathbf{Q}_{Z_1}}(\ell(T, \cdot)) + (1-\lambda) \mathbf{E}^{\mathbf{Q}_{Z_2}}(\ell(T, \cdot)), \end{aligned}$$

for all $\lambda \in [0, 1]$. Hence, the function $\mathcal{Z}_\eta \ni Z \mapsto \mathbf{E}^{\mathbf{Q}_Z}(\ell(T, \cdot))$ is linear and therefore also convex.

(3) For a given $\eta \in (0, \infty)$, the constant process $Z = 1$ satisfies $D_f(\mathbf{Q}_Z \| \mathbf{P}) = D_f(\mathbf{P} \| \mathbf{P}) = 0 < \eta$. It is also an interior point of the subset $\mathcal{Z}_\eta \subseteq \mathcal{M}_+(1)$.³ According to Slater's condition (Slater 2014), strong duality holds.

(4) Let $\eta := \mathbf{E}(f(Z^*(T)))$, and observe that

$$\begin{aligned} \inf_{\vartheta \in (0, \infty)} d(\vartheta, \eta) &\leq d(\vartheta^*, \eta) \\ &= \sup_{Z \in \mathcal{M}_+(1)} \mathbf{E}^{\mathbf{Q}_Z}(\widehat{\ell}_{\vartheta^*}(T, Z)) + \frac{\eta}{\vartheta^*} \\ &= \mathbf{E}^{\mathbf{Q}_{Z^*}}(\widehat{\ell}_{\vartheta^*}(T, Z^*)) + \frac{\eta}{\vartheta^*} \\ &= \mathbf{E}^{\mathbf{Q}_{Z^*}}(\ell(T, \cdot)) - \frac{1}{\vartheta^*} \mathbf{E}^{\mathbf{Q}_{Z^*}}\left(\frac{f(Z^*(T))}{Z^*(T)}\right) + \frac{\eta}{\vartheta^*} \end{aligned}$$

²Jensen's inequality ensures that $\eta \geq f(\mathbf{E}(Z^*(T))) = f(1) = 0$

³To see this point, consider the continuous function $\mathcal{H} : \mathcal{M}_+(1) \rightarrow \mathbb{R}$ defined by $\mathcal{H}(Z) = \mathbf{E}(f(Z(T)))$ (we endow $\mathcal{M}_+(1)$ with the topology induced by the metric $d(Z_1, Z_2) = \mathbf{E}(|f(Z_1(T)) - f(Z_2(T))|)$). The continuity ensures that $S := \mathcal{H}^{-1}((-\eta, \eta))$ is an open subset of $\mathcal{M}_+(1)$. Furthermore,

$$S \subseteq \{Z \in \mathcal{M}_+(1) | D_f(\mathbf{Q}_Z \| \mathbf{P}) < \eta\} \subseteq \mathcal{Z}_\eta$$

suggesting that $S \subseteq \text{int}(\mathcal{Z}_\eta)$. As an element in S , the constant process $Z = 1$ is an interior point of \mathcal{Z}_η .

$$\begin{aligned}
 &= \mathbb{E}^{\mathbb{Q}_{Z^*}}(\ell(T, \cdot)) - \frac{1}{\vartheta^*} \mathbb{E}(f(Z^*(T))) + \frac{\eta}{\vartheta^*} \\
 &= \mathbb{E}^{\mathbb{Q}_{Z^*}}(\ell(T, \cdot)) \\
 &\leq \sup_{Z \in \mathcal{Z}_\eta} \mathbb{E}^{\mathbb{Q}_Z}(\ell(T, \cdot))
 \end{aligned}$$

Lemma 4.1.1(3) then ensures that

$$\inf_{\vartheta \in (0, \infty)} d(\vartheta, \eta) = \mathbb{E}^{\mathbb{Q}_{Z^*}}(\ell(T, \cdot)) = \sup_{Z \in \mathcal{Z}_\eta} \mathbb{E}^{\mathbb{Q}_Z}(\ell(T, \cdot))$$

and the result follows. \square

For the primal problem formulated in Eq. 4.1.3, Lemma. 4.1.1(4) implies the existence of a solution Z^* that lies on the boundary of \mathcal{Z}_η given $\eta > 0$ (i.e. $\mathbb{E}(f(Z^*(T))) = \eta$), as long as Z^* solves

$$\max_{Z \in \mathcal{M}_+(1)} \mathbb{E}^{\mathbb{Q}_Z}(\widehat{\ell}_\vartheta(T, Z)) \tag{4.1.6}$$

for some $\vartheta \in (0, \infty)$. In the following context, we will consider the dual problem formulated in Eq. 4.1.6 instead of the primal problem. For simplicity, we will regard $\theta > 0$ as given and express $\widehat{\ell}_\vartheta$ by $\widehat{\ell}$.

4.2 Characterising the Worst-Case Expected Loss

This section provides implicit characterisation of the solution to the worst-case expected loss problem formulated in (4.1.3).

Given $t \in [0, T]$ and $\bar{Z} \in \mathcal{M}_+(1)$, define the family of \bar{Z} -consistent martingale densities up to time t by

$$\mathcal{Z}(t, \bar{Z}) := \{Z \in \mathcal{M}_+(1) \mid Z(t) = \bar{Z}(t)\}.$$

Note that $\mathcal{Z}(0, \bar{Z}) = \mathcal{M}_+(1)$, since $Z(0) = 1 = \bar{Z}(0)$ for all $Z \in \mathcal{M}_+(1)$. Note that the martingale property of the members of $\mathcal{Z}(t, \bar{Z})$ ensures that

$$Z(s) = \mathbb{E}(Z(t) \mid \mathcal{F}_s^0) = \mathbb{E}(\bar{Z}(t) \mid \mathcal{F}_s^0) = \bar{Z}(s),$$

for all $Z \in \mathcal{Z}(t, \bar{Z})$ and all $s \in [0, t]$. In other words, $\mathcal{Z}(t, \bar{Z})$ is the set of processes in $\mathcal{M}_+(1)$ that are consistent with \bar{Z} over the interval $[0, t]$. Moreover, we observe that

$$\begin{aligned}
 \mathbb{Q}_Z(A) &= \mathbb{E}(\mathbf{1}_A Z(T)) = \mathbb{E}(\mathbb{E}(\mathbf{1}_A Z(T) \mid \mathcal{F}_t^0)) = \mathbb{E}(\mathbf{1}_A Z(t)) = \mathbb{E}(\mathbf{1}_A \bar{Z}(t)) \\
 &= \mathbb{E}(\mathbb{E}(\mathbf{1}_A \bar{Z}(T) \mid \mathcal{F}_t^0)) \\
 &= \mathbb{E}(\mathbf{1}_A \bar{Z}(T)) \\
 &= \mathbb{Q}_{\bar{Z}}(A),
 \end{aligned}$$

for all $Z \in \mathcal{Z}(t, \bar{Z})$ and all $A \in \mathcal{F}_t^0$. That is to say, the probability measures associated with members of $\mathcal{Z}(t, \bar{Z})$ agree with each other on all \mathcal{F}_t^0 -measurable events. This is the set of feasible alternative measures by looking forward (from time t).

Given $\bar{Z} \in \mathcal{M}_+(1)$, we now define the \mathfrak{F}^0 -adapted process $(\widehat{L}(t, \bar{Z}))_{t \in [0, T]}$ by

$$\widehat{L}(t, \bar{Z}) := \max_{Z \in \mathcal{Z}(t, \bar{Z})} \mathbb{E}^{\mathbb{Q}_Z}(\widehat{\ell}(T, Z) - \widehat{\ell}(t, Z) \mid \mathcal{F}_t^0) \quad (4.2.1)$$

for all $t \in [0, T]$, assuming the maximum always exists. Since $\widehat{\ell}(\cdot, Z)$ is a non-anticipative functional satisfying $\widehat{\ell}(0, Z) = 0$ P-a.s. and $Z(0) = 1$ implies that $\mathbb{Q}_Z|_{\mathcal{F}_0^0} = \mathbb{P}|_{\mathcal{F}_0^0}$, it follows that $\widehat{\ell}(0, Z) = 0$ \mathbb{Q}_Z -a.s. as well. Consequently,

$$\widehat{L}(0, \bar{Z}) = \max_{Z \in \mathcal{M}_+(1)} \mathbb{E}^{\mathbb{Q}_Z}(\widehat{\ell}(T, Z) \mid \mathcal{F}_0^0) = \max_{Z \in \mathcal{M}_+(1)} \mathbb{E}^{\mathbb{Q}_Z}(\widehat{\ell}(T, Z)), \quad (4.2.2)$$

where the second equality follows from the fact that \mathcal{F}_0^0 and \mathcal{F}_T^0 are independent sigma-algebras, with respect to \mathbb{Q}_Z .⁴ This is simply the problem given in Eq. 4.1.6.

Definition 4.2.1. A worst-case density process is some $Z^* \in \mathcal{M}_+(1)$ that solves the maximisation problem (4.2.1) w.r.t the family of Z^* -consistent martingale densities:

$$\mathbb{E}^{\mathbb{Q}_{Z^*}}(\widehat{\ell}(T, Z^*) - \widehat{\ell}(t, Z^*) \mid \mathcal{F}_t^0) = \widehat{L}(t, Z^*) \quad (4.2.3)$$

for each $t \in [0, T]$.

Suppose $Z^* \in \mathcal{M}_+(1)$ is a worst-case martingale density according to the definition above, then Z^* solves the problem formulated in Eq. 4.1.6. This is confirmed by substituting Eq. 4.2.2 into Eq. 4.2.3 which leads to $\mathbb{E}^{\mathbb{Q}_{Z^*}}(\widehat{\ell}(T, Z^*)) = \max_{Z \in \mathcal{M}_+(1)} \mathbb{E}^{\mathbb{Q}_Z}(\widehat{\ell}(T, Z))$. In the proposition below, we characterizes such worst-case density by its martingale property.

Proposition 4.2.2. Fix $\bar{Z} \in \mathcal{M}_+(1)$ and suppose the maximum in (4.2.1) exists for each $t \in [0, T]$. Then the process $[0, T] \ni t \mapsto \widehat{L}(t, \bar{Z}) + \widehat{\ell}(t, \bar{Z})$ is a $\mathbb{Q}_{\bar{Z}}$ -supermartingale. It is a $\mathbb{Q}_{\bar{Z}}$ -martingale iff \bar{Z} is a worst-case density process.

Proof. Given an arbitrary $t \in [0, T]$, we suppose $Z' \in \mathcal{Z}(t, \bar{Z})$ solves the maximisation problem (Eq. 4.2.1). Applying the law of iterated expectation, we have

$$\mathbb{E}^{\mathbb{Q}_{Z'}}(\widehat{\ell}(T, Z') - \widehat{\ell}(t, Z') \mid \mathcal{F}_s^0) = \mathbb{E}^{\mathbb{Q}_{Z'}}\left(\mathbb{E}^{\mathbb{Q}_{Z'}}(\widehat{\ell}(T, Z') - \widehat{\ell}(t, Z') \mid \mathcal{F}_t^0) \mid \mathcal{F}_s^0\right)$$

⁴First observe that $Z(0) = 1$ implies that $\mathbb{Q}_Z(A) = \mathbb{P}(A) = 0$ or $\mathbb{Q}_Z(A) = \mathbb{P}(A) = 1$, for all $A \in \mathcal{F}_0^0$. Consequently, given $A \in \mathcal{F}_0^0$ and $B \in \mathcal{F}_T^0$, we obtain

$$0 \leq \mathbb{Q}_Z(A \cap B) \leq \mathbb{Q}_Z(A) = 0 = \mathbb{Q}_Z(A)\mathbb{Q}_Z(B),$$

in the case when $\mathbb{Q}_Z(A) = 0$, while

$$\begin{aligned} \mathbb{Q}_Z(A)\mathbb{Q}_Z(B) &= \mathbb{Q}_Z(B) \geq \mathbb{Q}_Z(A \cap B) = \mathbb{Q}_Z((A^c \cup B^c)^c) = 1 - \mathbb{Q}_Z(A^c \cup B^c) \geq 1 - (\mathbb{Q}_Z(A^c) + \mathbb{Q}_Z(B^c)) \\ &= 1 - \mathbb{Q}_Z(B^c) \\ &= \mathbb{Q}_Z(B) \\ &= \mathbb{Q}_Z(A)\mathbb{Q}_Z(B), \end{aligned}$$

in the case when $\mathbb{Q}_Z(A) = 1$.

$$= \mathbb{E}^{\mathbb{Q}_{Z'}} \left(\widehat{L}(t, \bar{Z}) \mid \mathcal{F}_s^0 \right) \quad (4.2.4)$$

for all $s \in [0, t]$. By virtue of $Z'(s) = \bar{Z}(s)$, $\widehat{\ell}(s, Z') = \widehat{\ell}(s, \bar{Z})$ for all $s \in [0, t]$. The same condition also leads to $Z' \in \mathcal{Z}(s, \bar{Z})$. According to the definition of \widehat{L} (Eq. 4.2.1), we have the following inequality

$$\begin{aligned} \widehat{L}(s, \bar{Z}) &\geq \mathbb{E}^{\mathbb{Q}_{Z'}} \left(\widehat{\ell}(T, Z') - \widehat{\ell}(s, Z') \mid \mathcal{F}_s^0 \right) \\ &= \mathbb{E}^{\mathbb{Q}_{Z'}} \left(\widehat{\ell}(t, Z') - \widehat{\ell}(s, Z') \mid \mathcal{F}_s^0 \right) + \mathbb{E}^{\mathbb{Q}_{Z'}} \left(\widehat{\ell}(T, Z') - \widehat{\ell}(t, Z') \mid \mathcal{F}_s^0 \right) \\ &= \mathbb{E}^{\mathbb{Q}_{\bar{Z}}} \left(\widehat{\ell}(t, \bar{Z}) - \widehat{\ell}(s, \bar{Z}) + \widehat{L}(t, \bar{Z}) \mid \mathcal{F}_s^0 \right) \end{aligned} \quad (4.2.5)$$

for all $s \in [0, t]$. In the last equality, we replace $\mathbb{Q}_{Z'}$ by $\mathbb{Q}_{\bar{Z}}$ because $\widehat{\ell}(t, \bar{Z})$, $\widehat{\ell}(s, \bar{Z})$ and $\widehat{L}(t, \bar{Z})$ are all \mathcal{F}_t^0 -measurable.⁵ Since $t \in [0, T]$ is chosen arbitrarily, Eq. 4.2.5 holds for any s and t that satisfies $0 \leq s \leq t \leq T$.

By re-arranging Eq. 4.2.5, we obtain the supermartingale property of the \mathfrak{F}^0 -adapted process $[0, T] \ni t \mapsto \widehat{L}(t, \bar{Z}) + \widehat{\ell}(t, \bar{Z})$:

$$\widehat{L}(s, \bar{Z}) + \widehat{\ell}(s, \bar{Z}) \geq \mathbb{E}^{\mathbb{Q}_{\bar{Z}}} \left(\widehat{L}(t, \bar{Z}) + \widehat{\ell}(t, \bar{Z}) \mid \mathcal{F}_s^0 \right) \quad (4.2.6)$$

The process is a $\mathbb{Q}_{\bar{Z}}$ -martingale iff the equality holds for all $0 \leq s \leq t \leq T$. If \bar{Z} is a worst-case density process, then according to Definition 4.2.1 \bar{Z} solves Eq. 4.2.1 for all $t \in [0, T]$. We may set $Z' = \bar{Z}$ in Eq. 4.2.5 so that the first line takes the equal sign for all $s \in [0, t]$. Conversely, if the equality holds for all $0 \leq s \leq t \leq T$, then it holds for all $0 \leq s \leq t = T$. By taking the equal sign in Eq. 4.2.6 and replacing t by T , we get

$$\widehat{L}(s, \bar{Z}) = \mathbb{E}^{\mathbb{Q}_{\bar{Z}}} \left(\widehat{\ell}(T, \bar{Z}) - \widehat{\ell}(s, \bar{Z}) \mid \mathcal{F}_s^0 \right)$$

for all $s \in [0, T]$, confirming that \bar{Z} is a worst-case density process by Definition 4.2.1. \square

Proposition. 4.2.2 can be regarded as generalization of the dynamic programming equation. In fact, given an optimal martingale density $Z^* \in \mathcal{M}_+(1)$, we take an arbitrary $\bar{Z} \in \mathcal{Z}(s, Z^*)$ and substitute it into Eq. 4.2.6. By observing that $\bar{Z} \in \mathcal{Z}(s, Z^*)$ matches Z^* up to time s , we transform Eq. 4.2.6 into

$$\widehat{L}(s, Z^*) + \widehat{\ell}(s, Z^*) \geq \mathbb{E}^{\mathbb{Q}_{\bar{Z}}} \left(\widehat{L}(t, \bar{Z}) + \widehat{\ell}(t, \bar{Z}) \mid \mathcal{F}_s^0 \right)$$

⁵The conditional expectation of a \mathcal{F}_t^0 -measurable function $X : \Omega \rightarrow \mathbb{R}$ w.r.t a sub- σ -algebra $\mathcal{F}_s^0 \subseteq \mathcal{F}_t^0$ is

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}_{Z'}} (X \mid \mathcal{F}_s^0) &= \mathbb{E} \left(\frac{Z'(T)}{Z'(s)} X \mid \mathcal{F}_s^0 \right) = \mathbb{E} \left(\frac{Z'(t)}{Z'(s)} \mathbb{E} \left(\frac{Z'(T)}{Z'(t)} X \mid \mathcal{F}_t^0 \right) \mid \mathcal{F}_s^0 \right) = \mathbb{E} \left(\frac{Z'(t)}{Z'(s)} \mathbb{E}^{\mathbb{Q}_{Z'}} (X \mid \mathcal{F}_t^0) \mid \mathcal{F}_s^0 \right) \\ &= \mathbb{E} \left(\frac{\bar{Z}(t)}{\bar{Z}(s)} X \mid \mathcal{F}_s^0 \right) \\ &= \mathbb{E}^{\mathbb{Q}_{\bar{Z}}} (X \mid \mathcal{F}_s^0) \end{aligned}$$

The inequality holds for all $\bar{Z} \in \mathcal{Z}(s, Z^*)$. It takes the equal sign when $\bar{Z} = Z^*$. This leads to the following dynamic programming equation with respect to the density process,

$$\widehat{L}(s, Z^*) + \widehat{\ell}(s, Z^*) = \max_{Z \in \mathcal{Z}(s, Z^*)} \mathbb{E}^{\mathbb{Q}_{Z^*}} \left(\widehat{L}(t, Z) + \widehat{\ell}(t, Z) \mid \mathcal{F}_s^0 \right)$$

for all s and t that satisfy $0 \leq s \leq t \leq T$.

4.3 General Result of Model Risk Measurement

We have shown in Proposition. 4.2.2 that the \mathfrak{F}^0 -adapted process $[0, T] \ni t \mapsto \widehat{L}(t, Z^*) + \widehat{\ell}(t, Z^*)$ is a \mathbb{Q}_{Z^*} -martingale iff Z^* is a worst-case density process. In this section, we will show that such Z^* indeed exists under certain conditions and is characterized by an equation. This leads to a complete solution to the problem formulated in Eq. 4.1.3. First we prove a lemma.

Lemma 4.3.1. *Fix a martingale density $\bar{Z} \in \mathcal{M}_+(1)$. A measurable process $C : [0, T] \times \Omega \rightarrow \mathbb{R}$, satisfying*

$$\mathbb{E}^{\mathbb{Q}_{\bar{Z}}} (C(t, \cdot) \mid \mathcal{F}_t^0) \leq \mathbb{E}^{\mathbb{Q}_{\bar{Z}}} (C(t, \cdot) \mid \mathcal{F}_t^0) \quad (4.3.1)$$

for all $t \in [0, T]$ and all $Z \in \mathcal{Z}(t, \bar{Z})$, admits a progressively measurable modification, i.e. there exists a progressively measurable process $\tilde{C} : [0, T] \times \Omega \rightarrow \mathbb{R}$, regarded as a non-anticipative functional, satisfying $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = \tilde{C}(t, \omega)\}) = 1$ for every $t \in [0, T]$.

Proof. The \mathcal{F}_t^0 -measurable function $u(t, \cdot) := \mathbb{E}^{\mathbb{Q}_{\bar{Z}}} (C(t, \cdot) \mid \mathcal{F}_t^0)$ forms a \mathfrak{F}^0 -adapted process $(u(t, \cdot))_{t \in [0, T]}$. It admits a progressively measurable modification $(\tilde{C}(t, \cdot))_{t \in [0, T]}$ (Karatzas and Shreve 1991). We would like to show that $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = \tilde{C}(t, \omega)\}) = 1$ for every $t \in [0, T]$.

We prove this lemma by contradiction. Suppose there exists a $t \in [0, T]$ such that $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = \tilde{C}(t, \omega)\}) < 1$, then $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = u(t, \omega)\}) < 1$.⁶ This implies that either $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) < u(t, \omega)\}) > 0$ or $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) > u(t, \omega)\}) > 0$. Without losing generality, we assume $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) < u(t, \omega)\}) > 0$.

⁶We only need to prove $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = u(t, \omega)\}) = 1$ leads to $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = \tilde{C}(t, \omega)\}) = 1$. In fact, assuming $\mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = u(t, \omega)\}) = 1$ we have

$$\begin{aligned} \mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) = \tilde{C}(t, \omega)\}) &= 1 - \mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) \neq \tilde{C}(t, \omega)\}) \\ &\geq 1 - \mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) \neq u(t, \omega)\} \cup \{\omega \in \Omega \mid u(t, \omega) \neq \tilde{C}(t, \omega)\}) \\ &\geq 1 - \mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid C(t, \omega) \neq u(t, \omega)\}) - \mathbb{Q}_{\bar{Z}}(\{\omega \in \Omega \mid u(t, \omega) \neq \tilde{C}(t, \omega)\}) \\ &= 1 \end{aligned}$$

For notational simplicity, in the rest of the proof we use C to denote the random variable $C(t, \cdot)$ and u to denote the \mathcal{F}_t^0 -measurable function $u(t, \cdot)$. We construct an alternative martingale density $Z' \in \mathcal{Z}(t, \bar{Z})$ by

$$Z'(s) = \begin{cases} \bar{Z}(s) & s \in [0, t] \\ \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} + e^u \mathbf{1}_{C \geq u} | \mathcal{F}_s^0)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} + e^u \mathbf{1}_{C \geq u} | \mathcal{F}_t^0)} \bar{Z}(s) & s \in (t, T] \end{cases} \quad (4.3.2)$$

To show that indeed $Z' \in \mathcal{Z}(t, \bar{Z})$, we need to prove that $Z'(0) = 1$, $Z' \geq 0$, $Z'(t) = \bar{Z}(t)$ and Z' is a P-martingale. The first three conditions are obvious from the definition. The martingale property of $(Z'(s))_{s \in [0, t]}$ is clear. The martingale property of $(Z'(s))_{s \in [t, T]}$ is confirmed by

$$\begin{aligned} \mathbb{E}(Z'(r) | \mathcal{F}_s^0) &= \mathbb{E}\left(\frac{\mathbb{E}(\bar{Z}(T)(e^C \mathbf{1}_{C < u} + e^u \mathbf{1}_{C \geq u}) | \mathcal{F}_r^0)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} + e^u \mathbf{1}_{C \geq u} | \mathcal{F}_t^0)} \middle| \mathcal{F}_s^0\right) \\ &= \frac{\mathbb{E}(\bar{Z}(T)(e^C \mathbf{1}_{C < u} + e^u \mathbf{1}_{C \geq u}) | \mathcal{F}_s^0)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} + e^u \mathbf{1}_{C \geq u} | \mathcal{F}_t^0)} = Z'(s) \end{aligned}$$

for all $s \in [t, T]$ and $r \in [t, T]$ satisfying $s \leq r$.

Because $\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0) = \mathbb{Q}_{\bar{Z}}(C < u) > 0$, there exists a $\omega \in \Omega$ such that $\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega) > 0$. We define

$$\begin{aligned} w_l &:= \mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega) \\ w_u &:= 1 - w_l = \mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C \geq u} | \mathcal{F}_t^0)(\omega) \\ L &:= \ln \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} \\ c_l &:= \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(Ce^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} \\ c_u &:= \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C \mathbf{1}_{C \geq u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C \geq u} | \mathcal{F}_t^0)(\omega)} \end{aligned}$$

then the LHS of Eq. 4.3.1 (with Z replaced by Z') satisfies

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}_{Z'}}(C | \mathcal{F}_t^0)(\omega) &= \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(Ce^C \mathbf{1}_{C < u} + Ce^u \mathbf{1}_{C \geq u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} + e^u \mathbf{1}_{C \geq u} | \mathcal{F}_t^0)(\omega)} = \frac{w_l c_l e^L + w_u c_u e^u}{w_l e^L + w_u e^u} \\ &> w_l c_l + w_u c_u \end{aligned} \quad (4.3.3)$$

Note that the inequality is given by the Chebyshev's sum inequality, which states that $w_1, w_2 > 0$ and $w_1 + w_2 = 1$, one have $(w_1 a_1 + w_2 a_2)(w_1 b_1 + w_2 b_2) < w_1 a_1 b_1 + w_2 a_2 b_2$ if $a_1 < a_2$ and $b_1 < b_2$. This inequality can be easily proved by expanding the left-hand side. In Eq. 4.3.3, we have $w_l > 0$, $w_u > 0$ ⁷ and

$$c_l = \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(Ce^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} < \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(ue^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} = u \leq c_u$$

⁷ $w_u = 0$ would lead to $\mathbb{E}^{\mathbb{Q}_{Z'}}(C | \mathcal{F}_t^0)(\omega) = \mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega) < u(\omega)$ in contradiction with the definition of u .

and

$$e^L = \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} < \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^u \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} = e^u$$

Therefore Chebyshev's sum inequality is applicable.

We further apply Jensen's inequality to the following expression twice ($x \ln x$ is a convex function while $\ln x$ is a concave function),

$$\mathbb{E}(x \ln x) \geq \mathbb{E}(x) \ln \mathbb{E}(x) \geq \mathbb{E}(x) \mathbb{E}(\ln x) \quad (4.3.4)$$

Following the inequality above, we take expectation w.r.t \mathcal{F}_t^0 and under the alternative measure generated by the Radon-Nikodym derivative

$$\frac{\bar{Z}(T)}{\bar{Z}(t)} \frac{\mathbf{1}_{C < u}}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0)}$$

By further assigning $x = e^C$, we get the following inequality

$$\begin{aligned} \frac{\mathbb{E}(\bar{Z}(T) C e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}(\bar{Z}(T) \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} &\geq \frac{\mathbb{E}(\bar{Z}(T) e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}(\bar{Z}(T) \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} \times \frac{\mathbb{E}(\bar{Z}(T) C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}(\bar{Z}(T) \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} \\ \Rightarrow \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(e^C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} &\geq \frac{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)}{\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega)} \end{aligned}$$

The LHS is simply c_t . Substituting the inequality into Eq. 4.3.3 one gets

$$\mathbb{E}^{\mathbb{Q}_{Z'}}(C | \mathcal{F}_t^0)(\omega) > \mathbb{E}^{\mathbb{Q}_{Z'}}(C \mathbf{1}_{C < u} | \mathcal{F}_t^0)(\omega) + \mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C \mathbf{1}_{C \geq u} | \mathcal{F}_t^0)(\omega) = \mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C | \mathcal{F}_t^0)(\omega)$$

This violates the condition stated in Eq. 4.3.1. We therefore conclude that $(C(t, \cdot))_{t \in [0, T]}$ admits a progressively measurable modification $(\tilde{C}(t, \cdot))_{t \in [0, T]}$. \square

A process C that satisfies the conditions in Lemma 4.3.1 admits a progressively measurable modification \tilde{C} w.r.t $\mathbb{Q}_{\bar{Z}}$, but not necessarily w.r.t the reference measure \mathbb{P} . However, if it also holds w.r.t \mathbb{P} , then we get the converse of Lemma 4.3.1. In fact, for $\bar{Z} \in \mathcal{M}_+(1)$ and any $Z \in \mathcal{Z}(t, \bar{Z})$, both $\mathbb{Q}_{\bar{Z}}$ and \mathbb{Q}_Z are absolutely continuous w.r.t \mathbb{P} , implying \tilde{C} is a modification of C w.r.t $\mathbb{Q}_{\bar{Z}}$ and \mathbb{Q}_Z . This results in

$$\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C(t, \cdot) | \mathcal{F}_t^0) = \mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(\tilde{C}(t, \cdot) | \mathcal{F}_t^0) \quad \text{and} \quad \mathbb{E}^{\mathbb{Q}_Z}(C(t, \cdot) | \mathcal{F}_t^0) = \mathbb{E}^{\mathbb{Q}_Z}(\tilde{C}(t, \cdot) | \mathcal{F}_t^0)$$

The progressively measurable process \tilde{C} is adapted to the filtration \mathfrak{F}^0 . Therefore

$$\mathbb{E}^{\mathbb{Q}_{\bar{Z}}}(C(t, \cdot) | \mathcal{F}_t^0) = \tilde{C}(t, \cdot) = \mathbb{E}^{\mathbb{Q}_Z}(C(t, \cdot) | \mathcal{F}_t^0)$$

for all $t \in [0, T]$ and $Z \in \mathcal{Z}(t, \bar{Z})$. We use Lemma 4.3.1 to prove the following proposition.

Proposition 4.3.2. $Z^* \in \mathcal{M}_+(1)$ is a worst-case martingale density iff the random variable

$$\Omega \ni \omega \mapsto \ell(T, \omega) - \frac{f'(Z^*(T)(\omega))}{\vartheta}$$

is constant, \mathbb{Q}_{Z^*} -a.s., and is dominated by that constant P -a.s.

Proof. Suppose $Z^* \in \mathcal{M}_+(1)$ is a worst-case martingale density. According to Definition 4.2.1,

$$Z^* = \arg \max_{Z \in \mathcal{Z}(t, Z^*)} \mathbb{E}^{\mathbb{Q}^Z}(\widehat{\ell}(T, Z) - \widehat{\ell}(t, Z) \mid \mathcal{F}_t^0) \quad (4.3.5)$$

for all $t \in [0, T]$. Given any $t \in [0, T]$ and any $Z \in \mathcal{Z}(t, Z^*)$, we construct a new martingale density that lies between Z^* and Z by

$$Z_\lambda = (1 - \lambda)Z^* + \lambda Z$$

where $\lambda \in [0, 1]$. $Z_\lambda \in \mathcal{Z}(t, Z^*)$ for all $\lambda \in [0, 1]$ due to the convexity of $\mathcal{Z}(t, Z^*)$. Since Z^* solves Eq. 4.3.5, the maximum value of

$$K(\lambda) := \mathbb{E}^{\mathbb{Q}^{Z_\lambda}}(\widehat{\ell}(T, Z_\lambda) - \widehat{\ell}(t, Z_\lambda) \mid \mathcal{F}_t^0) \quad (4.3.6)$$

is reached when $\lambda = 0$. Taking the first and second derivatives with respect to λ , we get

$$K'(\lambda) = \mathbb{E} \left(\frac{Z(T) - Z^*(T)}{Z^*(t)} \left(\ell(T, Z^*) - \ell(t, \cdot) - \frac{f'(Z_\lambda(T))}{\vartheta} \right) \mid \mathcal{F}_t^0 \right) \quad (4.3.7)$$

$$K''(\lambda) = - \mathbb{E} \left(\frac{(Z(T) - Z^*(T))^2}{Z^*(t)} \frac{f''(Z_\lambda(T))}{\vartheta} \mid \mathcal{F}_t^0 \right) \quad (4.3.8)$$

Notice that the twice-differentiable function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is convex as required by the non-negativity of the f -divergence (Ali and Silvey 1966,). This implies that $f''(z) > 0$ for all $z \in \mathbb{R}_+$. Combined with Eq. 4.3.8, this condition leads to $K''(\lambda) < 0$ for all $\lambda \in [0, 1]$. For $K(0) = \max_{\lambda \in [0, 1]} K(\lambda)$ to hold, the first derivative at $\lambda = 0$ must satisfy

$$K'(0) \leq 0 \Leftrightarrow \mathbb{E}^{\mathbb{Q}^Z}(C_{Z^*}(t, \cdot) \mid \mathcal{F}_t^0) \leq \mathbb{E}^{\mathbb{Q}^{Z^*}}(C_{Z^*}(t, \cdot) \mid \mathcal{F}_t^0) \quad (4.3.9)$$

where the process $C_{Z^*} : [0, T] \times \Omega \rightarrow \mathbb{R}$ is defined by

$$C_{Z^*}(t, \cdot) := \ell(T, \cdot) - \ell(t, \cdot) - \frac{f'(Z^*(T))}{\vartheta}$$

The inequality above holds for all $t \in [0, T]$ and all $Z \in \mathcal{Z}(t, Z^*)$. According to Lemma 4.3.1, C_{Z^*} admits a progressively measurable modification, say \tilde{C}_{Z^*} . In particular, at $t = 0$

$$C_{Z^*}(0, \cdot) = \ell(T, \cdot) - \frac{f'(Z^*(T))}{\vartheta}$$

takes a constant value $c := \tilde{C}_{Z^*}(0, 0)$, \mathbb{Q}_{Z^*} -a.s. In fact, \tilde{C}_{Z^*} is regarded as a non-anticipative functional so that $\tilde{C}_{Z^*}(0, \omega) = \tilde{C}_{Z^*}(0, 0) = c$ for all $\omega \in \Omega$ satisfying $(0, \omega) \sim (0, 0)$. As a result,

$$\begin{aligned} \mathbb{Q}_{Z^*}(C_{Z^*}(0, \cdot) = c) &\geq \mathbb{Q}_{Z^*}(\tilde{C}_{Z^*}(0, \cdot) = c) - \mathbb{Q}_{Z^*}(C_{Z^*}(0, \cdot) \neq \tilde{C}_{Z^*}(0, \cdot)) \\ &= \mathbb{Q}_{Z^*}(\tilde{C}_{Z^*}(0, \cdot) = c) - 0 \geq \mathbb{Q}_{Z^*}((0, \cdot) \sim (0, 0)) \\ &= \mathbb{Q}_{Z^*}(\{\omega \in \Omega \mid \omega(0) = 0\}) = 1 \end{aligned} \quad (4.3.10)$$

Next we prove $\mathbb{P}(C_{Z^*}(0, \cdot) \leq c) = 1$ by contradiction. Suppose on the contrary that $\mathbb{P}(C_{Z^*}(0, \cdot) > c) > 0$. We construct a martingale density $Z' \in \mathcal{Z}(0, Z^*) = \mathcal{M}_+(1)$ by setting

$$Z'(t) = \frac{\mathbb{E}(\mathbf{1}_{C_{Z^*}(0, \cdot) > c} \mid \mathcal{F}_t^0)}{\mathbb{P}(C_{Z^*}(0, \cdot) > c)}$$

for all $t \in [0, T]$. This leads to

$$\mathbb{E}^{\mathbb{Q}_{Z'}}(C_{Z^*}(0, \cdot) \mid \mathcal{F}_0^0) = \mathbb{E}(Z'(T)C_{Z^*}(0, \cdot)) = \frac{\mathbb{E}(\mathbf{1}_{C_{Z^*}(0, \cdot) > c} C_{Z^*}(0, \cdot))}{\mathbb{P}(C_{Z^*}(0, \cdot) > c)} > \frac{c\mathbb{E}(\mathbf{1}_{C_{Z^*}(0, \cdot) > c})}{\mathbb{P}(C_{Z^*}(0, \cdot) > c)} = c$$

Because we have already shown that $C_{Z^*}(0, \cdot) = c$, \mathbb{Q}_{Z^*} -a.s. (Eq. 4.3.10),

$$\mathbb{E}^{\mathbb{Q}_{Z^*}}(C_{Z^*}(0, \cdot) \mid \mathcal{F}_0^0) = c < \mathbb{E}^{\mathbb{Q}_{Z'}}(C_{Z^*}(0, \cdot) \mid \mathcal{F}_0^0)$$

According to Eq. 4.3.9, $K'(0) > 0$ (where the generic density process Z is replaced by the constructed process $Z' \in \mathcal{Z}(0, Z^*)$). This contradicts the assumption that Z^* is a worst-case martingale density.

Conversely, given a process $Z^* \in \mathcal{M}_+(1)$, suppose $C_{Z^*}(0, \cdot) : \Omega \rightarrow \mathbb{R}$ takes a constant value, say c , \mathbb{Q}_{Z^*} -a.s., and $C_{Z^*}(0, \cdot) \leq c$ P-a.s. Given any $t \in [0, T]$ and any $Z \in \mathcal{Z}(t, Z^*)$, $C_{Z^*}(0, \cdot) \leq c$ \mathbb{Q}_Z -a.s. due to the absolute continuity of \mathbb{Q}_Z w.r.t. \mathbb{P} . These properties lead to conditional expectations

$$\mathbb{E}^{\mathbb{Q}_{Z^*}}(C_{Z^*}(0, \cdot) \mid \mathcal{F}_t^0) = c \quad \text{and} \quad \mathbb{E}^{\mathbb{Q}_Z}(C_{Z^*}(0, \cdot) \mid \mathcal{F}_t^0) \leq c$$

Noticing that $C_{Z^*}(t, \cdot) = C_{Z^*}(0, \cdot) - \ell(t, \cdot)$ where $\ell(t, \cdot)$ is \mathcal{F}_t^0 -measurable, We have

$$\mathbb{E}^{\mathbb{Q}_Z}(C_{Z^*}(t, \cdot) \mid \mathcal{F}_t^0) \leq c - \ell(t, \cdot) = \mathbb{E}^{\mathbb{Q}_{Z^*}}(C_{Z^*}(t, \cdot) \mid \mathcal{F}_t^0)$$

According to Eq. 4.3.9, $K'(0) \leq 0$. Because $K''(\lambda) < 0$ (Eq. 4.3.8) for all $\lambda \in [0, 1]$, $K(0) \geq K(1)$. According to the definition of $K(\lambda)$ (Eq. 4.3.6), we have

$$\mathbb{E}^{\mathbb{Q}_{Z^*}}(\hat{\ell}(T, Z^*) - \hat{\ell}(t, Z^*) \mid \mathcal{F}_t^0) = K(0) \geq K(1) = \mathbb{E}^{\mathbb{Q}_Z}(\hat{\ell}(T, Z) - \hat{\ell}(t, Z) \mid \mathcal{F}_t^0)$$

This inequality applies to every $t \in [0, T]$ and every $Z \in \mathcal{Z}(t, Z^*)$. As a result, Z^* solves Eq. 4.3.5 for all $t \in [0, T]$ and is indeed a worst-case martingale density. \square

It is noted that Proposition 4.2.2 is a general result that works for any \mathfrak{F}^0 -adapted process $(\widehat{\ell}(t, Z))_{t \in [0, T]}$, irrespective of its actual formulation (Eq. 4.1.4). On the other hand, Proposition 4.3.2 makes use of the formulation, thus specifying the condition of a worst-case martingale density w.r.t the function $f(x)$. Note that any worst-case density process $Z^* \in \mathcal{M}_+(1)$ solves the original problem formulated in Eq. 4.1.6. Assuming the existence of such Z^* , we regard Eq. 4.1.6 as the initial value (at $t = 0$) of a particular process, termed as the value process. In general, we define three \mathfrak{F}^0 -adapted processes as below.

Definition 4.3.3. Given $\vartheta \in (0, \infty)$ and a worst-case martingale density $Z^* \in \mathcal{M}_+(1)$, the value process, $U : [0, T] \times \Omega \rightarrow \mathbb{R}$, the worst-case risk, $V : [0, T] \times \Omega \rightarrow \mathbb{R}$, and the budget process $\eta : [0, T] \times \Omega \rightarrow \mathbb{R}$,⁸ regarded as non-anticipative functionals, are defined by

$$\begin{aligned} U(t, \cdot) &:= \widehat{L}(t, Z^*) + \ell(t, \cdot) \\ V(t, \cdot) &:= \widehat{L}(t, Z^*) + \widehat{\ell}(t, Z^*) + F(t, Z^*) \\ \eta(t, \cdot) &:= \vartheta (V(t, \cdot) - U(t, \cdot)) \end{aligned}$$

where $(F(t, Z^*))_{t \in [0, T]}$ is the \mathcal{Q}_{Z^*} -martingale that satisfies $F(T, Z^*) = f(Z^*(T))/Z^*(T)$.

Intuitively, $U(t, \cdot)$ gives the worst-case expected loss, subtracting the on-going cost of perturbing the nominal model from time t to T . According to the definition of the worst-case martingale density (Eq. 4.2.3),

$$\begin{aligned} U(t, \cdot) &= \mathbb{E}^{\mathcal{Q}_{Z^*}} (\widehat{\ell}(T, Z^*) - \widehat{\ell}(t, Z^*) \mid \mathcal{F}_t^0) + \ell(t, \cdot) \\ &= \mathbb{E}^{\mathcal{Q}_{Z^*}} (\ell(T, \cdot) \mid \mathcal{F}_t^0) - \vartheta^{-1} Z^*(t)^{-1} \mathbb{E}^{\mathcal{Q}} (f(Z^*(T)) - f(Z^*(t)) \mid \mathcal{F}_t^0) \mathbf{1}_{Z^*(t) > 0} \end{aligned} \quad (4.3.11)$$

The second term is the penalization term for perturbing the nominal model from time t onwards. For continuity it is defined to be zero in the limiting case of $Z^*(t) = 0$. According to Definition 4.3.3, $V(t, \cdot)$ is the worst-case expected loss,

$$V(t, \cdot) = \mathbb{E}^{\mathcal{Q}_{Z^*}} (\widehat{\ell}(T, Z^*) \mid \mathcal{F}_t^0) + \vartheta^{-1} Z^*(t)^{-1} \mathbb{E}^{\mathcal{Q}} (f(Z^*(T)) \mid \mathcal{F}_t^0) \mathbf{1}_{Z^*(t) > 0} = \mathbb{E}^{\mathcal{Q}_{Z^*}} (\ell(T, \cdot) \mid \mathcal{F}_t^0)$$

The difference between $V(t, \cdot)$ and $U(t, \cdot)$ gives the cost for perturbing the nominal model (measured by the f -divergence), characterized by the process η :

$$\eta(t, \cdot) = Z^*(t)^{-1} \mathbb{E} (f(Z^*(T)) - f(Z^*(t)) \mid \mathcal{F}_t^0) \mathbf{1}_{Z^*(t) > 0}$$

We may further consider the terminal and initial values of the three processes. The value process, $U(t, \cdot)$, measures the target formulated in Eq. 4.1.6 from backwards, in the sense that

$$U(T) = \ell(T, \cdot) \quad \text{and} \quad U(0) = \mathbb{E}^{\mathcal{Q}_{Z^*}} (\widehat{\ell}(T, Z^*)) = \max_{Z \in \mathcal{M}_+(1)} \mathbb{E}^{\mathcal{Q}_Z} (\widehat{\ell}(T, Z)) \quad (4.3.12)$$

⁸We name it the budget process as it measures the remaining budget of the fictitious adversary (Glasserman and Xu 2014,). $\eta(0, \cdot)$ is referred as the relative entropy budget in (Glasserman and Xu 2014).

The worst-case risk process measures the model risk, Eq. 4.1.3, from backwards. According to Lemma. 4.1.1(4), the worst-case density Z^* solves the primal problem with $\eta := \eta(0, \cdot) = \mathbb{E}(f(Z^*(T)))$. Therefore

$$V(T) = \ell(T, \cdot) \quad \text{and} \quad V(0) = \mathbb{E}^{\mathbb{Q}_{Z^*}}(\ell(T, \cdot)) = \sup_{Z \in \mathcal{Z}_\eta} \mathbb{E}^{\mathbb{Q}_{Z^*}}(\ell(T, \cdot))$$

The cumulative budget η (i.e. relative entropy budget in (Glasserman and Xu 2014)) is measured by the budget process from backwards,

$$\eta(T) = 0 \quad \text{and} \quad \eta(0) = \mathbb{E}(f(Z^*(T))) = \eta$$

To solve the problem formulated in Eq. 4.1.6, Eq. 4.3.12 suggests solving the process U by backward induction. In a similar way, the model risk, Eq. 4.1.3, and its corresponding cumulative budget, η , may be quantified by solving the processes V and η by backward induction. The full procedure is given by the following theorem.

Theorem 4.3.4. *Given $\vartheta \in (0, \infty)$, suppose there exists a function $z : \mathbb{R} \rightarrow \mathbb{R}_+$ that satisfies*

$$x - \frac{f'(z(x))}{\vartheta} = c \quad \text{if } x \in I_c := \{\vartheta^{-1}y + c \mid y \in \text{range}(f')\} \quad (4.3.13)$$

$$z(x) = 0 \quad \text{if } x \notin I_c \quad (4.3.14)$$

where $c \in \mathbb{R}$ is a constant such that $\mathbb{E}(z \circ \ell(T, \cdot)) = 1$ and $\mathbb{P}(\ell(T, \cdot) < \sup I_c) = 1$. Then the value process, U , the worst-case risk, V , and the budget process, η , satisfy the following equations

$$\begin{aligned} U(t, \cdot) &= \frac{M(t) + f(Z(t))}{\vartheta Z(t)} + c \\ V(t, \cdot) &= \frac{W(t)}{Z(t)} \\ \eta(t, \cdot) &= \frac{\vartheta W(t) - M(t) - f(Z(t))}{Z(t)} - \vartheta c \end{aligned} \quad (4.3.15)$$

for all $t \in [0, T]$ and a.a. $\omega \in \{Z(t) > 0\}$, where (Z, M, W) is a \mathfrak{F}^0 -adapted \mathbb{P} -martingale that satisfies the following terminal condition:

$$\begin{pmatrix} Z \\ M \\ W \end{pmatrix} (T) = \begin{pmatrix} z \circ \ell(T, \cdot) \\ f' \circ z \circ \ell(T, \cdot) \times z \circ \ell(T, \cdot) - f \circ z \circ \ell(T, \cdot) \\ z \circ \ell(T, \cdot) \times \ell(T, \cdot) \end{pmatrix} \quad (4.3.16)$$

Proof. The function z defined by Eq. 4.3.13 provides a martingale density $Z \in \mathcal{M}_+(1)$ by composition:

$$Z(t) = \mathbb{E}(z \circ \ell(T, \cdot) \mid \mathcal{F}_t^0) \quad (4.3.17)$$

for all $t \in [0, T]$. Z is exactly the first element of the vectorized process defined in Eq. 4.3.16. It is indeed an element of $\mathcal{M}_+(1)$, for $Z(T) = z \circ \ell(T, \cdot) \geq 0$ and $Z(0) = \mathbb{E}(z \circ \ell(T, \cdot)) = 1$. The random variable

$$C_Z(0, \cdot) := \ell(T, \cdot) - \frac{f'(Z(T, \cdot))}{\vartheta} = \ell(T, \cdot) - \frac{f' \circ z \circ \ell(T, \cdot)}{\vartheta} \quad (4.3.18)$$

is equal to the constant c \mathbb{Q}_Z -a.s. In fact, $c \in \mathbb{R}$ is selected such that

$$1 = \mathbb{E}(z \circ \ell(T, \cdot)) = \mathbb{E}(z \circ \ell(T, \cdot) \mathbf{1}_{\ell(T, \cdot) \in I_c}) + \mathbb{E}(z \circ \ell(T, \cdot) \mathbf{1}_{\ell(T, \cdot) \notin I_c}) = \mathbb{E}(z \circ \ell(T, \cdot) \mathbf{1}_{\ell(T, \cdot) \in I_c})$$

by virtue of $z(x) = 0$ for all $x \notin I_c$. Since $C_Z(0, \omega) = c$ for all $\omega \in \Omega$ satisfying $\ell(T, \omega) \in I_c$, we have

$$\begin{aligned} \mathbb{Q}_Z(C_Z(0, \cdot) = c) &= \mathbb{E}(Z(T) \mathbf{1}_{C_Z(0, \cdot) = c}) \geq \mathbb{E}(z \circ \ell(T, \cdot) \mathbf{1}_{\ell(T, \cdot) \in I_c} \mathbf{1}_{C_Z(0, \cdot) = c}) \\ &= \mathbb{E}(z \circ \ell(T, \cdot) \mathbf{1}_{\ell(T, \cdot) \in I_c}) = 1 \end{aligned}$$

Next we need to show that $C_Z(0, \cdot) \leq c$ P-a.s. Notice that the function $f' : (0, \infty) \rightarrow \mathbb{R}$ is continuous and strictly increasing due to the convexity of f , implying that $\text{range}(f') = (f'(0_+), f'(\infty_-))$. We conclude that $\text{range}(f')$ is an open interval and denote it by (a, b) , where a and b can be either real numbers or $\pm\infty$. According to the assumption, we have

$$1 = \mathbb{P}(\ell(T, \cdot) < \sup I_c) = \mathbb{P}(\ell(T, \cdot) \in I_c \cup \ell(T, \cdot) \leq \{\vartheta^{-1}a + c\})$$

We extend the function f' continuously to zero by assigning $f'(0) = a$.

$$\begin{aligned} \mathbb{P}(C_Z(0, \cdot) \leq c) &= \mathbb{E}(\mathbf{1}_{\ell(T, \cdot) \in I_c} \mathbf{1}_{C_Z(0, \cdot) \leq c}) + \mathbb{E}(\mathbf{1}_{\ell(T, \cdot) \notin I_c} \mathbf{1}_{C_Z(0, \cdot) \leq c}) \\ &= \mathbb{E}(\mathbf{1}_{\ell(T, \cdot) \in I_c}) + \mathbb{E}(\mathbf{1}_{\ell(T, \cdot) \notin I_c} \mathbf{1}_{\ell(T, \cdot) - \vartheta^{-1}f'(0) \leq c}) \\ &= \mathbb{E}(\mathbf{1}_{\ell(T, \cdot) \in I_c}) + \mathbb{E}(\mathbf{1}_{\ell(T, \cdot) \notin (\vartheta^{-1}a + c, \vartheta^{-1}b + c)} \mathbf{1}_{\ell(T, \cdot) \leq \vartheta^{-1}a + c}) \\ &= \mathbb{P}(\ell(T, \cdot) \in I_c \cup \ell(T, \cdot) \leq \vartheta^{-1}a + c) = 1 \end{aligned}$$

We conclude that $C_Z(0, \cdot) = c$ \mathbb{Q}_Z -a.s. and $C_Z(0, \cdot) \leq c$ P-a.s. According to Proposition 4.3.2, Z defined in Eq. 4.3.17 is a worst-case density process.

The second component of Eq. 4.3.16 is a P-martingale given by

$$\begin{aligned} M(t) &= \mathbb{E}(f' \circ z \circ \ell(T, \cdot) \times z \circ \ell(T, \cdot) - f \circ z \circ \ell(T, \cdot) \mid \mathcal{F}_t^0) \\ &= \mathbb{E}(Z(T)f'(Z(T)) - f(Z(T)) \mid \mathcal{F}_t^0) \end{aligned}$$

for all $t \in [0, T]$. Substituting Eq. 4.3.18 into Eq. 4.3.11, we have

$$\begin{aligned} U(t, \cdot) &= \mathbb{E}^{\mathbb{Q}_Z} \left(C_Z(0, \cdot) + \frac{f'(Z(T, \cdot))}{\vartheta} \mid \mathcal{F}_t^0 \right) - \mathbb{E} \left(\frac{f(Z(T)) - f(Z(t))}{\vartheta Z(t)} \mid \mathcal{F}_t^0 \right) \\ &= c + \mathbb{E} \left(\frac{Z(T)f'(Z(T)) - f(Z(T)) + f(Z(t))}{\vartheta Z(t)} \mid \mathcal{F}_t^0 \right) \\ &= \frac{M(t) + f(Z(t))}{\vartheta Z(t)} + c \end{aligned}$$

By virtue of $C_Z(0, \cdot) = c$ \mathbb{Q}_Z -a.s., the equation above holds \mathbb{Q}_Z -a.s. More precisely, it

holds for a.a. $\omega \in \{Z(t) > 0\}$.⁹ The third element of Eq. 4.3.16, $W(t) = \mathbb{E}(z \circ \ell(T, \cdot) \times \ell(T, \cdot) | \mathcal{F}_t^0) = \mathbb{E}(Z(T)\ell(T, \cdot) | \mathcal{F}_t^0)$, characterizes the worst-case risk by

$$V(t, \cdot) = \mathbb{E}^{\mathbb{Q}^Z}(\ell(T, \cdot) | \mathcal{F}_t^0) = \mathbb{E}\left(\frac{Z(T)}{Z(t)}\ell(T, \cdot) \middle| \mathcal{F}_t^0\right) = \frac{W(t)}{Z(t)}$$

for all $\omega \in \{\omega \in \Omega | Z(t)(\omega) > 0\}$. Thus the equation above holds \mathbb{Q}_Z -a.s. Following the expressions for $U(t, \cdot)$ and $V(t, \cdot)$, we get the formula for the budget process

$$\eta(t, \cdot) = \vartheta(V(t, \cdot) - U(t, \cdot)) = \frac{\vartheta W(t) - M(t) - f(Z(t))}{Z(t)} - \vartheta c$$

□

In the proof above, we propose the inverse of the function f' , denoted by $g : \text{range}(f') \rightarrow (0, \infty)$. Using this inverse function, we have the following proposition which states that certain integrability conditions guarantee the existence of the solution, given by Theorem 4.3.4, to the problem of model risk quantification.

Proposition 4.3.5. *Denote $g : (a, b) \rightarrow (0, \infty)$ as the inverse function of f' . If $f'(\infty_-) = \infty$ and $g(\vartheta(\ell(T, \cdot) - c))\mathbf{1}_{\ell(T, \cdot) \in I_c}$ is integrable under the reference measure \mathbb{P} for every $c \in \mathbb{R}$, then the assumptions in Theorem 4.3.4 hold.*

Proof. We need to prove the existence of $c \in \mathbb{R}$ and $z : \mathbb{R} \rightarrow \mathbb{R}_+$, such that Eq. 4.3.13 for all $x \in I_c$ and $z(x) = 0$ for all $x \notin I_c$, $\mathbb{E}(z \circ \ell(T, \cdot)) = 1$ and $\mathbb{P}(\ell(T, \cdot) < \sup I_c) = 1$.

We have shown in the proof of Theorem 4.3.4 that $\text{range}(f') = (a, b)$. Here b takes ∞ as the strictly increasing function f' diverges at infinity. For a given $c \in \mathbb{R}$, the implicit equation Eq. 4.3.13 gives

$$z(x) = g(\vartheta(x - c))$$

for all $x \in I_c = (\vartheta^{-1}a + c, \infty)$. For all $x \notin I_c$, $z(x) = 0$ which gives

$$\begin{aligned} \mathbb{E}(z \circ \ell(T, \cdot)) &= \mathbb{E}(z \circ \ell(T, \cdot)\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a + c}) + \mathbb{E}(z \circ \ell(T, \cdot)\mathbf{1}_{\ell(T, \cdot) \leq \vartheta^{-1}a + c}) \\ &= \mathbb{E}(g(\vartheta(\ell(T, \cdot) - c))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a + c}) \end{aligned}$$

We would like to show that the function $K : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$K(c) := \mathbb{E}(g(\vartheta(\ell(T, \cdot) - c))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a + c}) \quad (4.3.19)$$

⁹ According to the definition of $C_Z(0, \cdot)$ (Eq. 4.3.18), $C_Z(0, \omega) = c$ for all $\omega \in \Omega$ satisfying $\ell(T, \omega) \in I_c$. It follows from $Z(T)(\omega) = z \circ \ell(T, \omega) = 0$ for a.a. $\omega \in \{\omega \in \Omega | \ell(T, \omega) \notin I_c\}$ that

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}^Z}(C_Z(0, \cdot) | \mathcal{F}_t^0) &= \mathbb{E}^{\mathbb{Q}^Z}(C_Z(0, \cdot)\mathbf{1}_{\ell(T, \cdot) \in I_c} | \mathcal{F}_t^0) + \mathbb{E}^{\mathbb{Q}^Z}(C_Z(0, \cdot)\mathbf{1}_{\ell(T, \cdot) \notin I_c} | \mathcal{F}_t^0) \\ &= Z(t)^{-1}\mathbb{E}(Z(T, \cdot)C_Z(0, \cdot)\mathbf{1}_{\ell(T, \cdot) \in I_c} | \mathcal{F}_t^0) + Z(t)^{-1}\mathbb{E}(Z(T, \cdot)C_Z(0, \cdot)\mathbf{1}_{\ell(T, \cdot) \notin I_c} | \mathcal{F}_t^0) \\ &= cZ(t)^{-1}\mathbb{E}(Z(T, \cdot)\mathbf{1}_{\ell(T, \cdot) \in I_c} | \mathcal{F}_t^0) \\ &= cZ(t)^{-1}\mathbb{E}(Z(T, \cdot) | \mathcal{F}_t^0) = c \end{aligned}$$

for a.a. $\omega \in \{Z(t) > 0\}$.

takes value of one for some $c \in \mathbb{R}$.

First we will show that K is continuous. Fix an arbitrary $c_0 \in \mathbb{R}$ and $\varepsilon \in (0, \infty)$. Resulted from the continuity of g , the function $y(\cdot, \omega) : (-\infty, c_0] \rightarrow \mathbb{R}$ defined by

$$y(c, \omega) := (g(\vartheta(\ell(T, \omega) - c)) - g(\vartheta(\ell(T, \omega) - c_0))) \mathbf{1}_{\ell(T, \omega) > \vartheta^{-1}a + c_0}$$

is continuous for every $\omega \in \Omega$. Therefore, the function $Y : (-\infty, c_0] \rightarrow \mathbb{R}$, defined by $Y(c) := \mathbb{E}(y(c, \cdot))$, is continuous at c_0 .¹⁰ Its continuity implies the existence of $\delta > 0$ such that $|Y(c)| = |Y(c) - Y(c_0)| < \varepsilon/2$ for all $c_0 \in \mathbb{R}$ satisfying $c_0 - \delta < c \leq c_0$. Let

$$\delta_- := \min\left(\delta, \frac{f'(\varepsilon/2) - a}{\vartheta}\right)$$

Then for all $c_0 - \delta_- < c \leq c_0$ we have

$$\begin{aligned} 0 \leq K(c) - K(c_0) &= \mathbb{E}(g(\vartheta(\ell(T, \cdot) - c)) \mathbf{1}_{\ell(T, \cdot) - \vartheta^{-1}a \in (c, c_0]}) + Y(c) \\ &\leq \mathbb{E}(g(\vartheta(\vartheta^{-1}a + c_0 - c)) \mathbf{1}_{\ell(T, \cdot) - \vartheta^{-1}a \in (c, c_0]}) + Y(c) \\ &< g(a + \vartheta\delta_-) + \varepsilon/2 \\ &\leq g(f'(\varepsilon/2)) + \varepsilon/2 \\ &= \varepsilon \end{aligned}$$

We may prove in a similar way that there exists $\delta_+ > 0$ such that $K(c) - K(c_0) \in (-\varepsilon, 0]$ for all $c_0 < c < c_0 + \delta_+$. Combining the two arguments, $|K(c) - K(c_0)|$ is less than ε for all $c \in \mathbb{R}$ satisfying $|c - c_0| < \min(\delta_+, \delta_-)$. This proves that the function K , defined in Eq. 4.3.19, is continuous.

Next we need to prove that there exist $c_+, c_- \in \mathbb{R}$ such that $K(c_+) \leq 1$ and $K(c_-) \geq 1$. In fact, the limit $\lim_{c \rightarrow -\infty} \mathbb{P}(\ell(T, \cdot) > \vartheta^{-1}a + c) = 1$ implies the existence of $c \in \mathbb{R}$ such that $\mathbb{P}(\ell(T, \cdot) > \vartheta^{-1}a + c) \geq 1/\xi$ for some $\xi > 1$. Defining

$$c_- := c - \frac{\max(0, f'(\xi) - a)}{\vartheta} \leq c$$

¹⁰It follows from the dominated convergence theorem that Y is continuous at c_0 . In fact, the sequence, $\{y(c_0 - 1/n, \cdot)\}_{n=1}^\infty$, of real-valued measurable functions converges pointwise to $y(c_0, \cdot)$ by virtue of its continuity. The sequence is dominated by $y(c_0 - 1, \cdot)$ due to the fact that g increases monotonically. $y(c_0 - 1, \cdot)$ is integrable as

$$\mathbb{E}(|y(c_0 - 1, \cdot)|) \leq \mathbb{E}(g(\vartheta(\ell(T, \cdot) - c_0 + 1)) \mathbf{1}_{\ell(T, \omega) > \vartheta^{-1}a + c_0}) \leq \mathbb{E}(g(\vartheta(\ell(T, \cdot) - c_0 + 1)) \mathbf{1}_{\ell(T, \omega) > I_{c_0 - 1}}) < \infty$$

The dominated convergence theorem guarantees the convergence of the expectation

$$\lim_{n \rightarrow \infty} \mathbb{E}(y(c_0 - 1/n, \cdot)) = \mathbb{E}(y(c_0, \cdot)) = 0$$

This means that given an arbitrary $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $|\mathbb{E}(y(c_0 - 1/n, \cdot))| < \varepsilon$ for all $n \geq n_0$. Due to the fact that g increases monotonically, for every $c \in [c_0 - 1/n_0, c_0]$ we have

$$0 \leq \mathbb{E}(y(c, \cdot)) - \mathbb{E}(y(c_0, \cdot)) = \mathbb{E}(y(c, \cdot)) \leq \mathbb{E}(y(c_0 - 1/n, \cdot)) < \varepsilon$$

This proves that Y is continuous at c_0 .

we have

$$\begin{aligned}
K(c_-) &\geq \mathbb{E}(g(\vartheta(\ell(T, \cdot) - c_-))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c}) \geq g(\vartheta(\vartheta^{-1}a + c - c_-))\mathbb{E}(\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c}) \\
&\geq g(\vartheta(\vartheta^{-1}a + \vartheta^{-1}(f'(\xi) - a)))\mathbb{E}(\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c}) \\
&= \xi \mathbb{P}(\ell(T, \cdot) > \vartheta^{-1}a + c) \\
&\geq 1
\end{aligned}$$

On the other hand, the following limit¹¹

$$\lim_{c \rightarrow \infty} \mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) - \vartheta^{-1}a \in (0, c)}) = \mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a}) < \infty$$

implies the existence of $c \in \mathbb{R}$ such that

$$\mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) - \vartheta^{-1}a \in (0, c)}) \geq \mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a}) - 1$$

Letting $c_+ = \max(0, c)$, we have

$$\begin{aligned}
K(c_+) &\leq \mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c_+}) \\
&\leq \mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c}) \\
&= \mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a}) - \mathbb{E}(g(\vartheta(\ell(T, \cdot))\mathbf{1}_{\ell(T, \cdot) - \vartheta^{-1}a \in (0, c)}) \\
&\leq 1
\end{aligned}$$

According to the intermediate value theorem, there exists $c \in \mathbb{R}$ such that the continuous function K , defined in Eq. 4.3.19, takes the value of one.¹²

The condition $\mathbb{P}(\ell(T, \cdot) < \sup I_c) = 1$ holds irrespective of the actual measure \mathbb{P} , for

$$\bigcup_{x \in I_c} \ell(T, \cdot) \leq x = \bigcup_{x > \vartheta^{-1}a+c} \{\omega \in \Omega \mid \ell(T, \omega) \leq x\} = \{\omega \in \Omega \mid \ell(T, \omega) \in \mathbb{R}\}$$

has probability one. As a result, the assumptions stated in Theorem 4.3.4 are valid, which guarantees the existence of the worst-case solution provided by the theorem. \square

We consider a special class of f -divergence, including the renowned Kullback-Leibler divergence, of which the function $\mathbb{R} \ni x \mapsto xf'(x) - f(x)$ is linear (or equivalently $x \mapsto xf''(x)$ is constant). This type of f -divergence has a particular advantage on applying Theorem 4.3.4, because the process

$$\begin{aligned}
M(t) &= \mathbb{E}(Z(T)f'(Z(T)) - f(Z(T)) \mid \mathcal{F}_t^0) \\
&= \mathbb{E}(Z(T) \mid \mathcal{F}_t^0) \times f'(\mathbb{E}(Z(T) \mid \mathcal{F}_t^0)) - f(\mathbb{E}(Z(T) \mid \mathcal{F}_t^0)) \\
&= Z(t)f'(Z(t)) - f(Z(t))
\end{aligned} \tag{4.3.20}$$

can be calculated directly from $Z(t)$. Therefore in practice we only need to apply backward induction to the two-dimensional P-martingale $(Z(t), W(t))_{t \in [0, T]}$. By substituting Eq. 4.3.20 into Eq. 4.3.15, we have the following proposition.

¹¹ The convergence is guaranteed by the dominated convergence theorem. See the previous footnote.

¹² Such $c \in \mathbb{R}$ is also unique by noticing that the function K is strictly decreasing.

Corollary 4.3.6. *Suppose in Theorem 4.3.4 there exists $d \in (0, \infty)$ such that $x f''(x) = d$ for all $x \in \mathbb{R}_+$. Then the value process, U , the worst-case risk, V , and the budget process, η , satisfy the following equations*

$$\begin{aligned} U(t, \cdot) &= \frac{f'(Z(t))}{\vartheta} + c \\ V(t, \cdot) &= \frac{W(t)}{Z(t)} \\ \eta(t, \cdot) &= \frac{\vartheta W(t)}{Z(t)} - f'(Z(t)) - \vartheta c \end{aligned} \tag{4.3.21}$$

for all $t \in [0, T]$ and all $\omega \in \Omega$ such that $Z(t)(\omega) > 0$, where (Z, W) is a \mathfrak{F}^0 -adapted P-martingale that satisfies the following terminal condition:

$$\begin{pmatrix} Z \\ W \end{pmatrix} (T) = \begin{pmatrix} z \circ \ell(T, \cdot) \\ z \circ \ell(T, \cdot) \times \ell(T, \cdot) \end{pmatrix}$$

Corollary 4.3.6 applies to the Kullback-Leibler divergence. In particular, the calculation of the constant c is pretty straightforward. We illustrate this in the following corollary.

Corollary 4.3.7. *Under the Kullback-Leibler divergence, suppose $\mathbb{E}(e^{\vartheta \ell(T, \cdot)}) < \infty$. Then there exists an unique solution to the problem of model risk quantification, given by*

$$\begin{aligned} U(t, \cdot) &= \frac{\ln \tilde{Z}(t)}{\vartheta} \\ V(t, \cdot) &= \frac{\tilde{W}(t)}{\tilde{Z}(t)} \\ \eta(t, \cdot) &= \frac{\vartheta \tilde{W}(t)}{\tilde{Z}(t)} - \ln \tilde{Z}(t) \end{aligned}$$

where (\tilde{Z}, \tilde{W}) is a \mathfrak{F}^0 -adapted P-martingale that satisfies the terminal condition:

$$\begin{pmatrix} \tilde{Z} \\ \tilde{W} \end{pmatrix} (T) = \begin{pmatrix} \exp(\vartheta \ell(T, \cdot)) \\ \ell(T, \cdot) \exp(\vartheta \ell(T, \cdot)) \end{pmatrix}$$

Proof. The Kullback-Leibler divergence adopts $f'(x) = (x \ln x)' = \ln x + 1$ for all $x \in (0, \infty)$. f' diverges at ∞ . The inverse function $g : \mathbb{R} \rightarrow (0, \infty)$ is given by $g(x) = e^{x-1}$. Since $\mathbb{E}(e^{\vartheta \ell(T, \cdot)}) < \infty$, we have

$$\mathbb{E}(|g(\vartheta(\ell(T, \cdot) - c)) \mathbf{1}_{\ell(T, \cdot) \in I_c}|) = e^{-\vartheta c - 1} \mathbb{E}(e^{\vartheta \ell(T, \cdot)}) < \infty$$

for all $c \in \mathbb{R}$. Proposition 4.3.5 guarantees the existence of a unique $c \in \mathbb{R}$ and $z : \mathbb{R} \rightarrow \mathbb{R}_+$ satisfying $\mathbb{E}(z \circ \ell(T, \cdot)) = 1$, therefore a unique solution to the problem of model risk quantification.

More specifically, we calculate the function $z : \mathbb{R} \rightarrow \mathbb{R}_+$ from Eq. 4.3.13:

$$z(x) = e^{\vartheta(x-c)-1}$$

for all $x \in \mathbb{R}$. The constant $c \in \mathbb{R}$ is given by

$$1 = \mathbb{E}(z \circ \ell(T, \cdot)) = \mathbb{E}\left(e^{\vartheta(\ell(T, \cdot)-c)-1}\right) \Leftrightarrow c = \frac{1}{\vartheta} \ln \mathbb{E}\left(e^{\vartheta \ell(T, \cdot)-1}\right) = \frac{\ln \tilde{Z}(0) - 1}{\vartheta}$$

The corollary defines two \mathbb{P} -martingales by

$$\begin{aligned} \tilde{Z}(t) &= \mathbb{E}\left(e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0\right) \\ \tilde{W}(t) &= \mathbb{E}\left(\ell(T, \cdot) e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0\right) \end{aligned}$$

The process Z and W in Corollary 4.3.6 are simply normalized versions of \tilde{Z} and \tilde{W} ,

$$\begin{aligned} Z(t) &= \mathbb{E}\left(z \circ \ell(T, \cdot) \mid \mathcal{F}_t^0\right) = \mathbb{E}\left(e^{\vartheta(\ell(T, \cdot)-c)-1} \mid \mathcal{F}_t^0\right) = \frac{\tilde{Z}(t)}{\tilde{Z}(0)} \\ W(t) &= \mathbb{E}\left(z \circ \ell(T, \cdot) \times \ell(T, \cdot) \mid \mathcal{F}_t^0\right) = \mathbb{E}\left(\ell(T, \cdot) e^{\vartheta(\ell(T, \cdot)-c)-1} \mid \mathcal{F}_t^0\right) = \frac{\tilde{W}(t)}{\tilde{Z}(0)} \end{aligned}$$

Substituting the equations above into Eq. 4.3.21, we have

$$\begin{aligned} U(t, \cdot) &= \frac{\ln(Z(t)) + 1}{\vartheta} + c = \frac{\ln \tilde{Z}(t)}{\vartheta} \\ V(t, \cdot) &= \frac{W(t)}{Z(t)} = \frac{\tilde{W}(t)}{\tilde{Z}(t)} \\ \eta(t, \cdot) &= \frac{\vartheta W(t)}{Z(t)} - (\ln(Z(t)) - 1) - \vartheta c = \frac{\vartheta \tilde{W}(t)}{\tilde{Z}(t)} - \ln \tilde{Z}(t) \end{aligned}$$

Note that $Z(T)(\omega) = e^{\vartheta(\ell(T, \omega)-c)-1} > 0$ for all $\omega \in \Omega$. $Z(t) = \mathbb{E}(Z(T) \mid \mathcal{F}_t^0) > 0$, implying that the equations above hold for all $t \in [0, T]$ and all $\omega \in \Omega$. \square

4.4 Model Risk Measurement with Continuous Semimartingales

The last section of this chapter provides the general theory on quantifying the model risk. In this section, we focus on the class of continuous semimartingales. It has an important property formulated by the functional Itô formula. To introduce the formula we need to briefly review the functional Itô calculus (Bally et al. 2016). First we define the horizontal derivative and the vertical derivative of a non-anticipative functional $F : \Lambda_T^d \rightarrow \mathbb{R}$. Its horizontal derivative at $(t, \omega) \in \Lambda_T^d$ is defined by the limit

$$\mathcal{D}F(t, \omega) := \lim_{h \rightarrow 0^+} \frac{F(t+h, \omega) - F(t, \omega)}{h}$$

if it exists. Intuitively, it describes the rate of change w.r.t time, assuming no change of the state variable from t onwards, and conditional on its history up to t given by the stopped path ω_t . On the other hand, the vertical derivative describes the rate of change w.r.t the state variable from t onwards. Formally, the vertical derivative at $(t, \omega) \in \Lambda_T^d$, denoted by $\nabla_\omega F(t, \omega)$, is defined as the gradient of the function $\mathbb{R}^d \ni x \mapsto F(t, \omega_t + x\mathbf{1}_{[t, T]})$ at 0, assuming its existence. The horizontal and vertical derivatives of a non-anticipative functional are also non-anticipative functionals.

We define the left-continuous non-anticipative functionals by noticing that the space of stopped paths, Λ_T^d , is endowed with a metric d_∞ . Suppose $F : \Lambda_T^d \rightarrow \mathbb{R}$ is a non-anticipative functional. F is left-continuous if for every $(t, \omega) \in \Lambda_T^d$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $|F(t, \omega) - F(t', \omega')| < \varepsilon$ for all $(t', \omega') \in \Lambda_T^d$ satisfying $t' < t$ and $d_\infty((t, \omega), (t', \omega')) < \delta$. We may further impose a boundedness condition on a non-anticipative functional F . It states that for any compact $K \subset \mathbb{R}^d$ and $t_0 < T$, there exists a $C > 0$ such that $|F(t, \omega)| \leq C$ for all $t \leq t_0$ and $\omega \in \Omega$. Suppose a non-anticipative functional F is horizontally differentiable and vertically twice-differentiable for all $(t, \omega) \in \Lambda_T^d$, and $\mathcal{D}F$, $\nabla_\omega F$ and $\nabla_\omega^2 F$ satisfy the boundedness condition above. In addition, F , $\nabla_\omega F$ and $\nabla_\omega^2 F$ are left-continuous, and $\mathcal{D}F$ is continuous for all $(t, \omega) \in \Lambda_T^d$. Then we call F a regular functional.

Suppose the canonical process X on Ω is a continuous semimartingale and $F : \Lambda_T^d \rightarrow \mathbb{R}$ is a regular functional. The \mathbb{R} -valued process $(Y(t))_{t \in [0, T]}$, defined by $Y(t) = F(t, \cdot)$ for all $t \in [0, T]$, follows the functional Itô formula P-a.s. (Bally et al. 2016, pp. 190-191)

$$Y(t) - Y(0) = \int_0^t \mathcal{D}F(u, \cdot) du + \int_0^t \nabla_\omega F(u, \cdot) dX(u) + \frac{1}{2} \int_0^t \text{Tr}(\nabla_\omega^2 F(u, \cdot) d[X](u))$$

If we further impose the constraint that $\int_0^T \xi(t) dX(t) = 0$ for all bounded predictable processes ξ satisfying $\int_0^T \xi(t) dt = 0$, then the canonical process X is a strong solution to the SDE (Revuz and Yor 2013)

$$dX(t) = \mu(t)dt + \sigma(t)dW(t) \tag{4.4.1}$$

where $(W(t))_{t \in [0, T]}$ is a \mathbb{R}^d -valued standard Wiener process on the underlying filtered probability space (assuming its existence). $(\mu(t))_{t \in [0, T]}$ is a \mathbb{R}^d -valued predictable process, and $(\sigma(t))_{t \in [0, T]}$ is a \mathbb{R}^{d^2} -valued predictable process. We may identify their elements, say $(\mu_i(t))_{t \in [0, T]}$ and $(\sigma_{ij}(t))_{t \in [0, T]}$, with non-anticipative functionals. The SDE Eq. 4.4.1 may be regarded as a path-dependent generalisation of the renowned Itô diffusion process. The existence and uniqueness of its solutions have been given in the literature by imposing various conditions (e.g. boundedness and Lipschitz properties, see Bally et al. (2016)). Now if X satisfies Eq. 4.4.1 P-a.s., then it follows from the functional Itô formula that the process Y is a strong solution to the SDE

$$dY(t) = \left(\mathcal{D}F(t, \cdot) + \mu(t)\nabla_\omega F(t, \cdot) + \frac{\text{Tr}(\sigma(t)\sigma(t)^T \nabla_\omega^2 F(t, \cdot))}{2} \right) dt + \sigma(t)\nabla_\omega F(t, \cdot)dW(t)$$

For simplicity we may define a nonlinear differential operator \mathcal{A} that sends a regular functional to a non-anticipative functional by

$$\mathcal{A}F := \mathcal{D}F + \mu(t)\nabla_\omega F + \frac{1}{2} \text{Tr}(\sigma(t)\sigma(t)^T \nabla_\omega^2 F) \quad (4.4.2)$$

Then the process Y , defined by $Y(t) = F(t, \cdot)$, is a strong solution to

$$dY(t) = \mathcal{A}F(t, \cdot)dt + \sigma(t)\nabla_\omega F(t, \cdot)dW(t) \quad (4.4.3)$$

Suppose Y is a P-martingale, then the regular functional F satisfies $\mathcal{A}F = 0$ P-a.s. Applying this property, we may convert the martingale statement in Theorem 4.3.4 to an analytical statement. This is formulated in the following corollary.

Corollary 4.4.1. *Given $\vartheta \in (0, \infty)$, suppose there exist $c \in \mathbb{R}$ and $z : \mathbb{R} \rightarrow \mathbb{R}_+$ defined in Theorem 4.3.4. If the canonical process X satisfies Eq. 5.3.1 for some \mathbb{R}^d -valued predictable process $(\mu(t))_{t \in [0, T]}$ and \mathbb{R}^{d^2} -valued predictable process $(\sigma(t))_{t \in [0, T]}$, then the value process, U , the worst-case risk, V , and the budget process, η , satisfy the following equations*

$$\begin{aligned} U(t, \cdot) &= \frac{M(t) + f(Z(t))}{\vartheta Z(t)} + c \\ V(t, \cdot) &= \frac{W(t)}{Z(t)} \\ \eta(t, \cdot) &= \frac{\vartheta W(t) - M(t) - f(Z(t))}{Z(t)} - \vartheta c \end{aligned}$$

for all $t \in [0, T]$ and all $\omega \in \Omega$ such that $Z(t)(\omega) > 0$, where Z , M and W are identified by the solutions to the equation $\mathcal{A}F = 0$ (P-a.s.), subject to their respective terminal conditions:

$$\begin{pmatrix} Z \\ M \\ W \end{pmatrix} (T) = \begin{pmatrix} z \circ \ell(T, \cdot) \\ f' \circ z \circ \ell(T, \cdot) \times z \circ \ell(T, \cdot) - f \circ z \circ \ell(T, \cdot) \\ z \circ \ell(T, \cdot) \times \ell(T, \cdot) \end{pmatrix}$$

In practice, we are more interested in the type of f -divergence that gives the constant function $x \mapsto x f''(x)$. Such an f -divergence allows us to solve U and V directly using path-dependent partial differential equations.

Proposition 4.4.2. *Suppose there exists $d \in (0, \infty)$ such that $x f''(x) = d$ for all $x \in \mathbb{R}_+$, and $f'(\infty-) = \infty$. In addition, the inverse function, $g : \text{Im} f' \rightarrow (0, \infty)$, provides a twice-differentiable function $\mathbb{R} \ni x \mapsto g(x) \mathbf{1}_{x \in \text{Im} f'}$. The value process and the worst-case risk, identified with the regular functionals $U_t := U(t, \cdot)$ and $V_t := V(t, \cdot)$, solve the following path-dependent partial differential equations \mathbb{Q}_Z -a.s.*

$$\begin{aligned} \mathcal{A}U_t + \frac{\theta g''(\vartheta(U_t - c))}{2 g'(\vartheta(U_t - c))} (\sigma_t \nabla_\omega U_t)^2 &= 0 \\ \mathcal{A}V_t + \frac{\vartheta g'(\vartheta(U_t - c)) \nabla_\omega U_t \sigma_t \sigma_t^T}{g(\vartheta(U_t - c))} \nabla_\omega V_t &= 0 \end{aligned} \quad (4.4.4)$$

subject to the terminal condition $U_T = V_T = \ell(T, \cdot)$. The budget process $\eta_t = \vartheta(V_t - U_t)$ for all $t \in [0, T]$. Defining $I_c := \{\vartheta^{-1}y + c \mid y \in \text{Im}f'\}$, the solution exists if $g(\vartheta(\ell(T, \cdot) - c))\mathbf{1}_{\ell(T, \cdot) \in I_c}$ is integrable for every $c \in \mathbb{R}$.

Proof. It follows from Corollary 4.3.6 that¹³

$$Z(t) = g(\vartheta(U_t - c))\mathbf{1}_{Z(t) > 0} = g(\vartheta(U_t - c))\mathbf{1}_{U_t \in I_c} := \mathbf{g}(\vartheta(U_t - c))$$

for all $t \in [0, T]$, where \mathbf{g} denotes the function $\mathbb{R} \ni x \mapsto g(x)\mathbf{1}_{x \in \text{Im}f'}$ which is twice-differentiable by assumption. Since $(Z(t))_{t \in [0, T]}$ is a P-martingale that can be identified with a solution to the equation $\mathcal{A}F = 0$ (P-a.s.), we have

$$0 = \mathcal{A}\mathbf{g}(\vartheta(U_t - c)) = \mathbf{g}'(\vartheta(U_t - c))\mathcal{A}U_t + \frac{\theta}{2}\mathbf{g}''(\vartheta(U_t - c))(\sigma_t \nabla_\omega U_t)^2$$

For all $\omega \in \Omega$ such that $U(t, \omega) \in I_c$, the equation is equivalent to¹⁴

$$\mathcal{A}U_t + \frac{\theta}{2} \frac{\mathbf{g}''(\vartheta(U_t - c))}{\mathbf{g}'(\vartheta(U_t - c))} (\sigma_t \nabla_\omega U_t)^2 = 0 \quad (4.4.5)$$

Noticing that $\{\omega \in \Omega \mid U(t, \omega) \in I_c\}$ has measure one under \mathbb{Q}_Z ¹⁵, the equation above holds \mathbb{Q}_Z -a.s.

It follows from Eq. 4.4.3 that the P-martingale $(Z(t))_{t \in [0, T]}$ solves the SDE

$$dZ(t) = \mathcal{A}\mathbf{g}(\vartheta(U_t - c))dt + \sigma_t \nabla_\omega \mathbf{g}(\vartheta(U_t - c))dW(t) = \vartheta \mathbf{g}'(\vartheta(U_t - c))\sigma_t \nabla_\omega U_t dW(t)$$

¹³We have shown in the proof of Proposition 4.3.5 that f' diverges at infinity implies that $\text{Im}f'$ is an open interval in the form of (a, ∞) . Then

$$U(t, \omega) = \vartheta^{-1}f'(Z(t)(\omega)) + c > \vartheta^{-1}a + c \in I_c$$

for all $\omega \in \{\omega \in \Omega \mid Z(t)(\omega) > 0\}$. On the other hand, for all $\omega \in \{\omega \in \Omega \mid Z(t)(\omega) = 0\}$,

$$\begin{aligned} 0 = Z(t)(\omega) &= \mathbb{E}^{\mathbb{Q}_Z}(Z(T)\mathbf{1}_{\ell(T, \cdot) \leq \vartheta^{-1}a+c} \mid \mathcal{F}_t^0)(\omega) + \mathbb{E}^{\mathbb{Q}_Z}(Z(T)\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c} \mid \mathcal{F}_t^0)(\omega) \\ &\geq \mathbb{E}^{\mathbb{Q}_Z}(g(\vartheta(\ell(T, \cdot) - c))\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c} \mid \mathcal{F}_t^0)(\omega) \end{aligned}$$

This implies that $\mathbb{E}^{\mathbb{Q}_Z}(\mathbf{1}_{\ell(T, \cdot) > \vartheta^{-1}a+c} \mid \mathcal{F}_t^0)(\omega) = 0$ by virtue of $\text{Im}g = (0, \infty)$, which gives

$$U(t, \omega) = \mathbb{E}^{\mathbb{Q}_Z}(\ell(T, \cdot) \mid \mathcal{F}_t^0)(\omega) = \mathbb{E}^{\mathbb{Q}_Z}(\ell(T, \cdot)\mathbf{1}_{\ell(T, \cdot) \leq \vartheta^{-1}a+c} \mid \mathcal{F}_t^0)(\omega) \leq \vartheta^{-1}a + c \notin I_c$$

¹⁴For all $x \in (a, \infty)$, $\mathbf{g}'(x) = g'(x) > 0$ (due to the convexity of f), and for all $x \in (-\infty, a]$,

$$\mathbf{g}'(x) = \lim_{h \rightarrow 0^-} \frac{\mathbf{g}(x) - \mathbf{g}(x-h)}{h} = 0$$

Therefore, $\mathbf{g}(x) = g(x)\mathbf{1}_{x \in (a, \infty)}$ implies that $\mathbf{g}'(x) = g'(x)\mathbf{1}_{x \in (a, \infty)}$, which in turns implies $\mathbf{g}''(x) = g''(x)\mathbf{1}_{x \in (a, \infty)}$. For all $\omega \in \{\omega \in \Omega \mid U(t, \omega) \in I_c\}$, $\vartheta U(t, \omega) - c \in (a, \infty)$ and thus

$$\mathbf{g}'(\vartheta(U(t, \omega) - c)) = g'(\vartheta(U(t, \omega) - c)) > 0 \quad \text{and} \quad \mathbf{g}''(\vartheta(U(t, \omega) - c)) = g''(\vartheta(U(t, \omega) - c))$$

¹⁵ $\mathbb{Q}_Z(U(t, \cdot) \in I_c) = \mathbb{Q}_Z(Z(t) > 0) = \mathbb{E}(Z(T)\mathbf{1}_{Z(t) > 0}) = \mathbb{E}(Z(t)\mathbf{1}_{Z(t) > 0}) = \mathbb{E}(Z(t)) = 1$

We may define a process $(Y(t))_{t \in [0, T]}$ by the stochastic integral

$$Y(t) := \int_0^t \left(\frac{\vartheta g'(\vartheta(U_s - c))}{g(\vartheta(U_s - c))} \mathbf{1}_{U_s \in I_c} \sigma_t \nabla_\omega U_s \right) dW(s)$$

for all $t \in [0, T]$. This transforms the SDE above into

$$dZ(t) = \vartheta g'(\vartheta(U_t - c)) \mathbf{1}_{U_t \in I_c} \sigma_t \nabla_\omega U_t dW(t) = g(\vartheta(U_t - c)) \mathbf{1}_{U_t \in I_c} dY(t) = Z(t) dY(t)$$

suggesting that the process $(Z(t))_{t \in [0, T]}$ is a Doleans-Dade exponential, i.e. $Z = \mathcal{E}(Y)$. Note that the SDE above ensures that $(Z(t))_{t \in [0, T]}$ is a local martingale. We assume that $(Z(t))_{t \in [0, T]}$ is a martingale.¹⁶

According to the Girsanov theorem, the Brownian motion under \mathbb{Q}_Z is given by adding an extra drift term. Noticing that $U_t \in I_c$ \mathbb{Q}_Z -a.s., the Girsanov theorem transforms the SDE of the canonical process under \mathbb{P} (Eq. 5.3.1) to the following SDE (in the sense that $(X(t))_{t \in [0, T]}$ is a strong solution of the following under \mathbb{Q}_Z),

$$dX(t) = \left(\mu_t + \frac{\vartheta g'(\vartheta(U_t - c))}{g(\vartheta(U_t - c))} \sigma_t \sigma_t^T \nabla_\omega U_t \right) dt + \sigma_t dW^{\mathbb{Q}_Z}(t) \quad (4.4.6)$$

The functional Itô formula, Eq. 4.4.2-4.4.3, applies to the alternative measure \mathbb{Q}_Z as well. Following the definition of the operator \mathcal{A} , we have

$$\begin{aligned} \mathcal{A}^{\mathbb{Q}_Z} F(t, \cdot) &= DF(t, \cdot) + \left(\mu_t + \frac{\vartheta g'(\vartheta(U_t - c))}{g(\vartheta(U_t - c))} \nabla_\omega U_t \sigma_t \sigma_t^T \right) \nabla_\omega F(t, \cdot) + \frac{1}{2} \text{Tr}(\sigma(t) \sigma(t)^T \nabla_\omega^2 F(t, \cdot)) \\ &= \mathcal{A}F(t, \cdot) + \frac{g'(\vartheta \sigma_t \sigma_t^T (U_t - c)) \nabla_\omega U_t \sigma_t \sigma_t^T}{g(\vartheta(U_t - c))} \nabla_\omega F(t, \cdot) \end{aligned}$$

for some regular functional $F : \Lambda_T^d \rightarrow \mathbb{R}$ and all $t \in [0, T]$. The worst-case model risk, $\mathbb{E}^{\mathbb{Q}_Z}(\ell(T, \cdot) | \mathcal{F}_t)$, is a \mathbb{Q}_Z -martingale. Identified with the regular functional V , it satisfies the following equation \mathbb{Q}_Z -a.s.

$$0 = \mathcal{A}^{\mathbb{Q}_Z} V_t = \mathcal{A}V_t + \frac{\vartheta g'(\vartheta(U_t - c)) \nabla_\omega U_t \sigma_t \sigma_t^T}{g(\vartheta \sigma_t (U_t - c))} \nabla_\omega V_t \quad (4.4.7)$$

Combined with the terminal condition $U_T = V_T = \ell(T, \cdot)$, Eq. 4.4.5 and Eq. 4.4.7 provide the path-dependent partial differential equations that govern the value process and the worst-case risk, respectively. It follows from Proposition 4.3.5 that the solution indeed exists if $g(\vartheta(\ell(T, \cdot) - c)) \mathbf{1}_{\ell(T, \cdot) \in I_c}$ is integrable for every $c \in \mathbb{R}$. \square

Kullback-Leibler divergence provides a convenient setting for applying Proposition 4.4.2. The function $f'(x) = \ln x + 1$ diverges at ∞ , and its inverse $g : \mathbb{R} \rightarrow (0, \infty)$ given by $g(x) = e^{x-1}$ is twice-differentiable. In addition, the worst-case martingale density $Z(T) = e^{\vartheta(\ell(T, \cdot) - c) - 1} > 0$ supplies a measure \mathbb{Q}_Z that is equivalent to the reference measure \mathbb{P} . Combining Corollary 4.3.7 with Proposition 4.4.2, and substituting $g(x) = e^{x-1}$ into Eq. 4.4.4, we get the following corollary that applies to the Kullback-Leibler divergence.

¹⁶One may verify this case by case using some sufficient condition that ensures a local martingale to be a martingale, e.g. the Novikov's condition.

Corollary 4.4.3. *Under the Kullback-Leibler divergence, suppose $\mathbb{E} \left(e^{\vartheta \ell(T, \cdot)} \right) < \infty$. Then there exists a unique solution to the problem of model risk quantification. The value process and the worst-case risk, identified with regular functionals $U_t := U(t, \cdot)$ and $V_t := V(t, \cdot)$, solve the following path-dependent partial differential equations P-a.s.*

$$\begin{aligned} \mathcal{A}U_t + \frac{\vartheta}{2} (\sigma_t \nabla_\omega U_t)^2 &= 0 \\ \mathcal{A}V_t + \vartheta \nabla_\omega U_t \sigma_t \sigma_t^T \nabla_\omega V_t &= 0 \end{aligned} \quad (4.4.8)$$

subject to the terminal condition $U_T = V_T = \ell(T, \cdot)$. The budget process $\eta_t = \vartheta(V_t - U_t)$ for all $t \in [0, T]$.

In practice, the path-dependent partial differential equations, Eq. 4.4.8, are generally difficult to solve. However, we may convert Eq. 4.4.8 into normal non-linear partial differential equations for a special type of path dependency, formulated by

$$\ell(T, \cdot) = h_0(T, X(T)) + \int_0^T h_1(t, X(t)) dt + \int_0^T h(t, X(t)) dX(t) \quad (4.4.9)$$

for some functions $h : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $h_i : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ ($i = 1, 2$). We further restrict the canonical process X to the class of Itô diffusions. This means that the process is Markovian, and there exist functions $\mu : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d^2}$ such that $\mu_t = \mu(t, X(t))$ and $\sigma_t = \sigma(t, X(t))$. The path-dependent partial differential equations, Eq. 4.4.8, degenerates to normal partial differential equations.

Corollary 4.4.4. *Under the Kullback-Leibler divergence, suppose $\mathbb{E} \left(e^{\vartheta \ell(T, \cdot)} \right) < \infty$, the canonical process $(X(t))_{t \in [0, T]}$ solves the SDE, $dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t)$, and the cumulative loss $\ell(T, \cdot)$ takes the form of Eq. 4.4.9. If there exists a function $\tilde{u} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that solves the partial differential equation*

$$\begin{aligned} \frac{\partial \tilde{u}(t, x)}{\partial t} + \mu(t, x) \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) + \frac{\vartheta}{2} \left(\sigma(t, x) \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) \right)^2 \\ + \frac{1}{2} \text{Tr} \left(\sigma(t, x)^2 \frac{\partial^2 \tilde{u}(t, x)}{\partial x^2} \right) + h_1(t, x) = 0 \end{aligned} \quad (4.4.10)$$

and a function $\tilde{v} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that solves the partial differential equation

$$\begin{aligned} \frac{\partial \tilde{v}(t, x)}{\partial t} + \vartheta \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) \sigma(t, x)^2 \left(\frac{\partial \tilde{v}(t, x)}{\partial x} + h(t, x) \right) \\ + \mu(t, x) \left(\frac{\partial \tilde{v}(t, x)}{\partial x} + h(t, x) \right) + \frac{1}{2} \text{Tr} \left(\sigma(t, x)^2 \frac{\partial^2 \tilde{v}(t, x)}{\partial x^2} \right) + h_1(t, x) = 0 \end{aligned} \quad (4.4.11)$$

subject to the terminal condition $\tilde{u}(T, \cdot) = \tilde{v}(T, \cdot) = h_0(T, \cdot)$, then the value process, the worst-case risk and the budget process, identified with regular functionals, follow

$$\begin{aligned} U_t &= \tilde{u}(t, X(t)) + \int_0^t h_1(s, X(s))ds + \int_0^t h(s, X(s))dX(s) \\ V_t &= \tilde{v}(t, X(t)) + \int_0^t h_1(s, X(s))ds + \int_0^t h(s, X(s))dX(s) \end{aligned}$$

and $\eta_t = \vartheta(\tilde{v}(t, X(t)) - \tilde{u}(t, X(t)))$ for all $t \in [0, T]$.

Proof. We first define regular functionals $\tilde{U}, \tilde{V} : \Lambda_T^d \rightarrow \mathbb{R}$ by

$$\begin{aligned} \tilde{U}_t &:= U_t - \int_0^t h_1(s, X(s))ds - \int_0^t h_2(s, X(s))dX(s) \\ \tilde{V}_t &:= V_t - \int_0^t h_1(s, X(s))ds - \int_0^t h_2(s, X(s))dX(s) \end{aligned} \tag{4.4.12}$$

The horizontal and vertical derivatives can be derived from Eq. 4.4.12,

$$\begin{aligned} \mathcal{D}\tilde{U}_t &= \mathcal{D}U_t - h_1(t, X(t)) & \text{and} & & \mathcal{D}\tilde{V}_t &= \mathcal{D}V_t - h_1(t, X(t)) \\ \nabla_\omega \tilde{U}_t &= \nabla_\omega U_t - h_2(t, X(t)) & \text{and} & & \nabla_\omega \tilde{V}_t &= \nabla_\omega V_t - h_2(t, X(t)) \\ \nabla_\omega^2 \tilde{U}_t &= \nabla_\omega^2 U_t & \text{and} & & \nabla_\omega^2 \tilde{V}_t &= \nabla_\omega^2 V_t \end{aligned}$$

Substituting the equations above into Eq. 4.4.8, we transform Eq. 4.4.8 to

$$\begin{aligned} \mathcal{D}\tilde{U}_t + \mu(t, X(t))(\nabla_\omega \tilde{U}_t + h(t, X(s))) + \frac{\vartheta}{2} \left(\sigma(t, X(t))(\nabla_\omega \tilde{U}_t + h(t, X(s)))^2 \right) \\ + \frac{1}{2} \text{Tr} \left(\sigma(t, X(t))^2 \nabla_\omega^2 \tilde{U}_t \right) + h_1(t, X(t)) = 0 \end{aligned} \tag{4.4.13}$$

and

$$\begin{aligned} \mathcal{D}\tilde{V}_t + \vartheta(\nabla_\omega \tilde{U}_t + h(t, X(s)))^2 \sigma(t, X(t))^2 (\nabla_\omega \tilde{V}_t + h(t, X(s)))^2 \\ + \mu(t, X(t))(\nabla_\omega \tilde{V}_t + h(t, X(s))) + \frac{1}{2} \text{Tr} \left(\sigma(t, X(t))^2 \nabla_\omega^2 \tilde{V}_t \right) + h_1(t, X(t)) = 0 \end{aligned} \tag{4.4.14}$$

If there exists a function $\tilde{u} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that solves the partial differential equation

$$\begin{aligned} \frac{\partial \tilde{u}(t, x)}{\partial t} + \mu(t, x) \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) + \frac{\vartheta}{2} \left(\sigma(t, x) \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) \right)^2 \\ + \frac{1}{2} \text{Tr} \left(\sigma(t, x)^2 \frac{\partial^2 \tilde{u}(t, x)}{\partial x^2} \right) + h_1(t, x) = 0 \end{aligned}$$

and a function $\tilde{v} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that solves

$$\begin{aligned} \frac{\partial \tilde{v}(t, x)}{\partial t} + \vartheta \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) \sigma(t, x)^2 \left(\frac{\partial \tilde{v}(t, x)}{\partial x} + h(t, x) \right) \\ + \mu(t, x) \left(\frac{\partial \tilde{v}(t, x)}{\partial x} + h(t, x) \right) + \frac{1}{2} \text{Tr} \left(\sigma(t, x)^2 \frac{\partial^2 \tilde{v}(t, x)}{\partial x^2} \right) + h_1(t, x) = 0 \end{aligned}$$

then the regular functionals defined by $\tilde{U}_t := \tilde{u}(t, X(t))$ and $\tilde{V}_t := \tilde{v}(t, X(t))$, for all $t \in [0, T]$, satisfy Eqs. 4.4.13 and 4.4.14. The terminal condition $\tilde{U}_T = \tilde{V}_T = h_0(T, X(T))$ is satisfied if $\tilde{u}(T, x) = \tilde{v}(T, x) = h_0(T, x)$ holds for all $x \in \mathbb{R}$. \square

Note that Eq. 4.4.10-4.4.11 are non-linear parabolic partial differential equations and in general have to be solved numerically.

4.5 Concluding Remarks

This chapter provides a theoretical framework for model risk quantification. In particular, it goes beyond existing methods by incorporating path-dependent losses. We need several ingredients to formulate the problem, including terminal time T , a (path-dependent) loss function ℓ , a nominal model (i.e. a canonical process $(X_t)_{t \in [0, T]}$ under a nominal measure \mathbb{P}) and some f -divergence. The non-parametric nature of this approach relies on the f -divergence to restrict the set of proper alternative models. This is, however, only applicable to measures that are absolutely continuous w.r.t the nominal measure. More generic distance measure, such as the Wasserstein metric, may be applied instead (see Chapter 6). Despite of this shortcoming, f -divergence, especially the Kullback-Leibler divergence, is most tractable and yields simple results for path-dependent problems.

Chapter 5

Application on Dynamic Model Risk Measurement

Understanding and measuring model risk is important to financial practitioners. This type of risk is inherent in the application of any models, requiring a general approach of quantification. Here we provide approaches that incorporate model uncertainty into financial applications. They are similar to the well-known approaches to derivative pricing, namely the martingale approach, the tree approach and the differential equation approach. The non-parametric nature of our methodology allows practitioners to combine robustness into tasks such as derivative pricing and portfolio management with sufficient computational efficiency.

5.1 Introduction

The existing work on model uncertainty either focuses on parameter uncertainty or relies on comparisons between models. To go beyond that, Glasserman and Xu (2014) recently proposed a non-parametric approach. Under this framework, a worst-case model is found among all alternative models in a neighborhood of the reference model. Glasserman and Xu (2014) adopted the Kullback-Leibler divergence (i.e. relative entropy) to measure the distance between an alternative model and the reference model. They also proposed the use of the α -divergence to avoid the heavy tails that cause integrability issues under the Kullback-Leibler divergence.

Glasserman and Xu's approach, however, is based on a static setting under which the state variable does not change over time. Financial practitioners are mostly concerned about security prices which evolve continuously in time. An appropriate approach needs to be capable of quantifying model risk associated with a stochastic state process.

This chapter introduces three approaches to quantifying model risk, namely the martingale approach, the tree approach and the differential equation approach. Their counterparts in derivative pricing have been applied extensively. Essentially, the martingale approach utilizes the fact that the discounted future price of a derivative con-

tract acts as a martingale under the risk-neutral measure. The tree approach models the underlying process with a recombining tree and the differential equation approach aims to solve the Feynmann-Kac formula for derivative pricing. All three approaches may be generalised to a robust setting, producing upper and lower bounds on prices to account for model risk.

A particularly interesting topic in derivative pricing and risk management is path dependency. The general results presented in this chapter use the functional Ito calculus, which is very well suited for dealing with path-dependent problems. We will present several examples ranging from portfolio optimization to robust derivative pricing to illustrate the quantification of model uncertainty.

5.2 Martingale and Tree Approaches to Model Risk Quantification

It is well known among practitioners that derivative instruments may be priced using three equivalent approaches, namely the martingale approach, the tree (or lattice) approach and the analytical approach. In particular, we consider the path space $\Omega := D([0, T], \mathbb{R}^d)$ of the underlying asset prices and the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$, where \mathbb{P} is the risk-neutral measure. A derivative product that expires at time T pays a cumulative amount $\ell(t, \omega)$ at time $t \in [0, T]$, given the path $\omega \in \Omega$. Ignoring issues related to the time value of money, the fair price at $t \in [0, T]$ is the expectation, $\mathbb{E}(\ell(T, \cdot) | \mathcal{F}_t)$, under the measure \mathbb{P} . The fact that the fair price is a martingale allows financial practitioners to readily apply the martingale approach and the tree approach. If $\ell(T, \omega)$ takes some special form, one may also apply the Feynman-Kac formula to price the product. In this section, we will show that the three practical approaches may be generalized to account for the robustness, in the sense that the model risk is quantified in a non-parametric way.

Sec. 4 paves the theoretical ground for solving the model risk problem under some f -divergence,

$$\max_{D_f(\mathbb{Q}||\mathbb{P}) \leq \eta} \mathbb{E}^{\mathbb{Q}}(\ell(T, \cdot) | \mathcal{F}_0) \quad (5.2.1)$$

In most practical application, we choose the Kullback-Leibler divergence or the relative entropy, since it provides the greatest tractability. Under the framework of model risk management, Eq. 5.2.1 is termed as the worst-case risk and $\ell : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is termed as the loss function.

The primary purpose of dynamic model risk measurement is to determine the value process U^ϑ , the risk process V^ϑ , and the budget process $\eta^\vartheta := \vartheta(V^\vartheta - U^\vartheta)$ for some $\vartheta \in \mathbb{R}_+$. Given $\eta = \eta_0^\vartheta$, then the original problem has the solution given by V_0^ϑ . Now need to solve both U^ϑ and V^ϑ by backward induction, starting from $U_T^\vartheta = V_T^\vartheta = \ell(T, \cdot)$. Corollary 4.3.7 states that there exists a unique solution if $\mathbb{E}(e^{\vartheta \ell(T, \cdot)}) < \infty$. The three

processes may be calculated as follows:

$$\begin{aligned} U_t^\vartheta &= \frac{\ln \tilde{Z}(t)}{\vartheta} \\ V_t^\vartheta &= \frac{\tilde{W}(t)}{\tilde{Z}(t)} \\ \eta_t^\vartheta &= \frac{\vartheta \tilde{W}(t)}{\tilde{Z}(t)} - \ln \tilde{Z}(t) \end{aligned}$$

where (\tilde{Z}, \tilde{W}) is a \mathfrak{F}^0 -adapted P-martingale that satisfies the terminal condition:

$$\begin{pmatrix} \tilde{Z} \\ \tilde{W} \end{pmatrix} (T) = \begin{pmatrix} \exp(\vartheta \ell(T, \cdot)) \\ \ell(T, \cdot) \exp(\vartheta \ell(T, \cdot)) \end{pmatrix}$$

According to this corollary, the process $\tilde{Z}(t) := \exp(\vartheta U_t^\vartheta)$ is a P-martingale, subject to the terminal condition $\tilde{Z}(T) = \exp(\vartheta \ell(T, \cdot))$. This provides a martingale approach to the problem of model risk quantification. In particular, we may write down the expression for U_t^ϑ explicitly:

$$U_t^\vartheta = \vartheta^{-1} \ln \tilde{Z}(t) = \vartheta^{-1} \ln \mathbb{E} \left(e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0 \right) \quad (5.2.2)$$

The corollary also leads to the explicit expression for the worst-case risk V_t ,

$$V_t^\vartheta = \frac{\mathbb{E} \left(\ell(T, \cdot) e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0 \right)}{\mathbb{E} \left(e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0 \right)} \quad (5.2.3)$$

It is noted that Eq. 5.2.3 is a path-dependent generalisation of the worst-case risk derived in (Glasserman and Xu 2014). Eqs. 5.2.2-5.2.3 allow us to evaluate the budget process η_t by

$$\eta_t^\vartheta = \frac{\vartheta \mathbb{E} \left(\ell(T, \cdot) e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0 \right)}{\mathbb{E} \left(e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0 \right)} - \ln \mathbb{E} \left(e^{\vartheta \ell(T, \cdot)} \mid \mathcal{F}_t^0 \right) \quad (5.2.4)$$

In analogy to the martingale approach to derivative pricing, Eqs. 5.2.3-5.2.4 allow us to evaluate the worst-case risk V_0^ϑ along with its corresponding relative entropy budget η_0^ϑ . A practical issue is that the evaluation of expectations in general requires the probability density function. In cases where the probability density function is not readily available, Monte Carlo simulation may be adopted, which is regarded as a variation of the martingale approach. In particular, V_t may be estimated using Monte Carlo simulation as a weighted average of the losses of all sampled paths. The weighting factor, $e^{\vartheta \ell(T, \cdot)}$, takes the exponential form, amplifying the possibility of a heavier loss.

Another numerical approach to the problem of model risk quantification requires the construction of a tree. The tree resembles the stochastic evolution of the state variable $X(t, \omega) = \omega(t)$. It is noted that in general the tree is not recombining to account

for the path-dependency of the loss functional ℓ . The tree approach allows us to apply the dynamic model risk measurement to a large variety of canonical processes. A famous example is the CRR tree that models the celebrated Black-Scholes dynamics (Cox, Ross, Rubinstein et al. 1979). After constructing the tree, we iterate backwards from the terminal values evaluated at the last layer of the tree, i.e. $U_T^\vartheta = V_T^\vartheta = \ell(T, \cdot)$. For generality, we assume that there are n nodes at the t -th layer that connects to one node at the $(t-1)$ -th layer. The values assigned to these nodes are denoted by $U_t^{(k)}, \dots, U_t^{(k+n-1)}$ and $V_t^{(k)}, \dots, V_t^{(k+n-1)}$. We further denote the probability of the tree branch that links the k -th node at the $(t-1)$ -th layer to the j -th node at the t -th layer by $p(t-1, k; t, j)$. We can then assign the following values to the node at the $(t-1)$ -th layer by

$$U_{t-1}^{(k)} = \frac{1}{\vartheta} \ln \left(\sum_{j=k}^{k+n-1} p(t-1, k; t, j) e^{\vartheta U_t^{(j)}} \right) \quad (5.2.5)$$

and

$$V_{t-1}^{(k)} = \frac{\sum_{j=k}^{k+n-1} p(t-1, k; t, j) V_t^{(j)} e^{\vartheta U_t^{(j)}}}{\sum_{j=k}^{k+n-1} p(t-1, k; t, j) e^{\vartheta U_t^{(j)}}} \quad (5.2.6)$$

Applying Eqs. 5.2.5-6.3.15 iteratively, we are able to calculate the values of U_t^ϑ and V_t^ϑ at each node of the tree. The relative entropy budget is calculated by $\eta_t^\vartheta = \vartheta(V_t^\vartheta - U_t^\vartheta)$. Note that Jensen's inequality implies that

$$U_{t-1}^{(k)} \geq \sum_{j=k}^{k+n-1} p(t-1, k; t, j) U_t^{(j)}$$

illustrating that $(U_t^\vartheta)_{t=1,2,\dots,n}$ is a P-supermartingale.

Compared to Monte Carlo simulation, the tree approach suffers from the computational complexity that grows exponentially with the number of time steps. However, it provides deterministic results which could be a more important factor in practice. In particular, Monte Carlo simulation assigns uneven weights that are exponential to the path-wise risks. If a randomly sampled path has a significantly higher risk, the result can be dominated by this specific path. This problem is not encountered by the tree approach.

We have mentioned that in general we need to account for path dependency by requiring evaluation of a non-recombining tree. This is in practice computationally intensive. We will show that the path dependency may be removed if the loss functional $\ell(T, \cdot)$ takes the special integral form

$$\ell(T, \cdot) = h_0(T, X(T)) + \int_0^T h_1(t, X(t)) dt + \int_0^T h(t, X(t)) dX(t) \quad (5.2.7)$$

where h , h_0 and h_1 are real-valued functions $[0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$. In this case may define new processes, \tilde{U}^ϑ and \tilde{V}^ϑ , by

$$\begin{aligned}\tilde{U}_t^\vartheta &:= U_t^\vartheta - \int_0^t h_1(s, X(s))ds - \int_0^t h_2(s, X(s))dX(s) \\ \tilde{V}_t^\vartheta &:= V_t^\vartheta - \int_0^t h_1(s, X(s))ds - \int_0^t h_2(s, X(s))dX(s)\end{aligned}\tag{5.2.8}$$

Starting from the terminal condition $\tilde{U}_T^\vartheta = \tilde{V}_T^\vartheta = h_0(T, \cdot)$, we iterate backwards to solve \tilde{U}_t and \tilde{V}_t by

$$\tilde{U}_{t-1}^{(k)} = \frac{1}{\vartheta} \ln \left(\sum_{j=k}^{k+n-1} q(t-1, k; t, j) \right)\tag{5.2.9}$$

and

$$\tilde{V}_{t-1}^{(k)} = \frac{\sum_{j=k}^{k+n-1} q(t-1, k; t, j) \left(\tilde{V}_t^{(j)} + g(X_{t-1}^{(k)})\Delta t + h(X_{t-1}^{(k)}) (X_t^{(j)} - X_{t-1}^{(k)}) \right)}{\sum_{j=k}^{k+n-1} q(t-1, k; t, j)}\tag{5.2.10}$$

where

$$q(t-1, k; t, j) := p(t-1, k; t, j) e^{\vartheta \left(\tilde{U}_t^{(j)} + g(X_{t-1}^{(k)})\Delta t + h(X_{t-1}^{(k)}) (X_t^{(j)} - X_{t-1}^{(k)}) \right)}$$

and Δt is the time interval between two neighboring layers of the tree. Applying Eqs. 5.2.9-5.2.10 instead of Eqs. 5.2.5-6.3.15, the path dependency of the terminal conditions is removed, allowing the construction of a recombining tree (or lattice). It is noted that $\tilde{U}_0 = U_0$ and $\tilde{V}_0 = V_0$, demonstrating that evaluation of the modified processes, \tilde{U} and \tilde{V} , serves the purpose of quantifying the model risk at $t = 0$.

5.3 Differential Equations for Model Risk Quantification

Suppose the canonical process X , assumed to be a continuous process, is a strong solution to the SDE

$$dX_t = \mu_t dt + \sigma_t dW_t\tag{5.3.1}$$

where $(W_t)_{t \in [0, T]}$ is a \mathbb{R}^d -valued standard Wiener process on the underlying filtered probability space. $(\mu_t)_{t \in [0, T]}$ is a \mathbb{R}^d -valued progressively measurable process, and $(\sigma_t)_{t \in [0, T]}$ is a $\mathbb{R}^{d \times d}$ -valued progressively measurable process. We may alternatively approach the problem of model risk quantification by solving differential equations. In general the equations are path-dependent, as illustrated in Corollary 4.4.3. For convenience we repeat this corollary here:

Corollary 5.3.1. *Under the Kullback-Leibler divergence, suppose $\mathbb{E}(e^{\vartheta \ell(T, \cdot)}) < \infty$. Then there exists unique solution to the problem of model risk quantification. The value process $(U_t^\vartheta)_{t \in [0, T]}$ and the risk process $(V_t^\vartheta)_{t \in [0, T]}$ solve the following path-dependent partial differential equations P-a. s.*

$$\begin{aligned} \mathcal{D}U_t^\vartheta + \mu(t)\nabla_\omega U_t^\vartheta + \frac{1}{2} \text{Tr}(\sigma(t)^2 \nabla_\omega^2 U_t^\vartheta) + \frac{\vartheta}{2} (\sigma_t \nabla_\omega U_t^\vartheta)^2 &= 0 \\ \mathcal{D}V_t^\vartheta + \mu(t)\nabla_\omega V_t^\vartheta + \frac{1}{2} \text{Tr}(\sigma(t)^2 \nabla_\omega^2 V_t^\vartheta) + \vartheta \nabla_\omega U_t^\vartheta \sigma_t^2 \nabla_\omega V_t^\vartheta &= 0 \end{aligned} \quad (5.3.2)$$

subject to the terminal condition $U_T^\vartheta = V_T^\vartheta = \ell(T, \cdot)$. The budget process $\eta_t^\vartheta = \vartheta(V_t^\vartheta - U_t^\vartheta)$ for all $t \in [0, T]$.

In the corollary displayed above, the path-dependency is handled by the functional Ito calculus, in which \mathcal{D} and ∇_ω are the horizontal derivative operator and the vertical derivative operator (Bally et al. 2016). These operators are path-dependent generalisation of the partial derivatives w.r.t time and state variable, respectively. It is noted that on deriving Eq. 5.3.2 we applied the Girsanov theorem to characterize the worst-case change of measure. In fact, the worst-case measure \mathbb{Q}^* simply introduces an additional drift term $\vartheta \sigma_t^2 \nabla_\omega U_t^\vartheta$. Therefore, the state process under the worst-case measure is a strong solution to the SDE

$$dX_t = (\mu_t + \vartheta \sigma_t^2 \nabla_\omega U_t^\vartheta) dt + \sigma_t dW_t^{\mathbb{Q}^*}$$

Eq. 5.3.2 takes a functional form and is hard to solve. We would like to specify the loss function in order to remove the path dependency. This is indeed the case when the loss function takes the special form in Eq. 5.2.7, as given by Corollary 4.4.4. For convenience we repeat the corollary here:

Corollary 5.3.2. *Under the Kullback-Leibler divergence, suppose $\mathbb{E}(e^{\vartheta \ell(T, \cdot)}) < \infty$ and the canonical process $(X_t)_{t \in [0, T]}$ solves the SDE, $dX_t = \mu_t dt + \sigma_t dW_t$. In addition, the loss function $\ell(T, \cdot)$ takes the form of Eq. 5.2.7. If there exists a function $\tilde{u} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that solves the partial differential equation*

$$\begin{aligned} \frac{\partial \tilde{u}(t, x)}{\partial t} + \mu(t, x) \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) + \frac{\vartheta}{2} \left(\sigma(t, x) \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) \right)^2 \\ + \frac{1}{2} \text{Tr} \left(\sigma(t, x)^2 \frac{\partial^2 \tilde{u}(t, x)}{\partial x^2} \right) + h_1(t, x) = 0 \end{aligned} \quad (5.3.3)$$

and a function $\tilde{v} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ that solves the partial differential equation

$$\begin{aligned} \frac{\partial \tilde{v}(t, x)}{\partial t} + \vartheta \left(\frac{\partial \tilde{u}(t, x)}{\partial x} + h(t, x) \right) \sigma(t, x)^2 \left(\frac{\partial \tilde{v}(t, x)}{\partial x} + h(t, x) \right) \\ + \mu(t, x) \left(\frac{\partial \tilde{v}(t, x)}{\partial x} + h(t, x) \right) + \frac{1}{2} \text{Tr} \left(\sigma(t, x)^2 \frac{\partial^2 \tilde{v}(t, x)}{\partial x^2} \right) + h_1(t, x) = 0 \end{aligned} \quad (5.3.4)$$

subject to the terminal condition $\tilde{u}(T, \cdot) = \tilde{v}(T, \cdot) = h_0(T, \cdot)$, then the value process, the worst-case risk and the budget process, identified with regular functionals, follow

$$\begin{aligned} U_t^\vartheta &= \tilde{u}(t, X_t) + \int_0^t h_1(s, X_s) ds + \int_0^t h(s, X_s) dX(s) \\ V_t^\vartheta &= \tilde{v}(t, X_t) + \int_0^t h_1(s, X_s) ds + \int_0^t h(s, X_s) dX(s) \end{aligned}$$

and $\eta_t^\vartheta = \vartheta(\tilde{v}(t, X_t) - \tilde{u}(t, X_t))$ for all $t \in [0, T]$.

To use the corollary above, we solve the $(d + 1)$ -dimensional partial differential equations Eqs. 5.3.3-5.3.4. Numerical approaches such as finite difference schemes may be applicable. The functions u and v immediately result in the worst-case risk $V_0^\vartheta = \tilde{v}(0, X_0)$ along with the relative entropy budget $\eta_0^\vartheta = \vartheta(\tilde{v}(0, X_0) - \tilde{u}(0, X_0))$. The worst-case measure is characterized by the additional drift term

$$\vartheta \sigma_t^2 \nabla_\omega U_t^\vartheta = \vartheta \sigma_t^2 \nabla_\omega \frac{\partial \tilde{u}(t, X_t)}{\partial t} + h_1(t, X_t) \quad (5.3.5)$$

We want to comment on the link between the problems of derivative pricing and model risk quantification. If $\vartheta = 0$, it is evident from the relation $\eta_0 = \vartheta(V_0 - U_0)$ that the relative entropy budget η_0 is also zero. As a result, the reference model is the only legitimate model. The process V in fact provides the fair price of a derivative product whose payoff function is given by ℓ . The martingale approach, Eq. 5.2.3, is reduced to $V_t = \mathbb{E}(\ell(T, \cdot) | \mathcal{F}_t^0)$, showing that V is a P-martingale. The tree approach, Eq. 6.3.15, is also reduced to the tree approach to derivative pricing, $V_{t-1}^{(k)} = \sum_{j=k}^{k+n-1} p(t-1, k; t, j) V_t^{(j)}$. The partial differential equation, Eq. 5.3.4, is reduced to

$$\frac{\partial \tilde{v}(t, x)}{\partial t} + \mu(t, x) \left(\frac{\partial \tilde{v}(t, x)}{\partial x} + h(t, x) \right) + \frac{1}{2} \text{Tr} \left(\sigma(t, x)^2 \frac{\partial^2 \tilde{v}(t, x)}{\partial x^2} \right) + h_1(t, x) = 0 \quad (5.3.6)$$

subject to the terminal condition $\tilde{v}(T, x) = h_0(T, x)$. This equation is a generalized form of the Feynman-Kac formula for solving the following expectation:

$$\tilde{v}(t, x) = \mathbb{E} \left(h_0(T, X(T)) + \int_t^T h_1(s, X(s)) ds + \int_t^T h(s, X(s)) dX(s) \middle| X(t) = x \right)$$

From the other point of view, Eqs. 5.3.3 and 5.3.4 may be regarded as a robust generalisation of the pricing formula Eq. 5.3.6. Consider robust derivative pricing as an example, we may set the loss function as the discounted pay off of the derivative product. After choosing a value for $\vartheta > 0$, we solve Eqs. 5.3.3 and 5.3.4 to find \tilde{u}_0^ϑ and \tilde{v}_0^ϑ . Then \tilde{v}_0^ϑ may be viewed as the upper bound of the derivative price accounting for model ambiguity, at a parametric significance level of ϑ . In a similar way, we calculate $\tilde{v}_0^{-\vartheta}$ as the lower bound, and thus $[\tilde{v}_0^{-\vartheta}, \tilde{v}_0^\vartheta]$ provides robustness to the standard derivative pricing at a parametric significance level of ϑ . One may actually calculate the entropic cost corresponding to ϑ , by $\tilde{\eta}_0^\vartheta = \vartheta(\tilde{v}_0^\vartheta - \tilde{u}_0^\vartheta)$. Unfortunately, in general $\tilde{\eta}_0^\vartheta$ is not equal to $\tilde{\eta}_0^{-\vartheta}$. If we would like to fix an entropic significance level of η for both the upper bound and the lower bound, we may have to sacrifice computational tractability by applying numerical routines and solving Eqs. 5.3.3 and 5.3.4 multiple times.

5.4 Solutions to Problems with Linear and Quadratic Loss Functions

Now let us consider the closed-form solutions to the partial differential equations Eqs. 5.3.3-5.3.4, when the technique of variable separation is applicable. For example, we guess the non-linear partial differential equation Eq. 5.3.3 has a simple solution with the following form,

$$\tilde{u}(t, x) = h_0(T, p(t)x) + \int_t^T P(s)ds \quad (5.4.1)$$

for some integrable function $P : [0, T] \rightarrow \mathbb{R}$ and some function $p : [0, T] \rightarrow \mathbb{R}^{d \times d}$ subject to $p(T) = I$ ($d \times d$ identity matrix). Substituting Eq. 5.4.1 into Eq. 5.3.3, we obtain

$$\begin{aligned} P(t) = g_1(t, x)p'(t)x + \mu_t(p(t)g_1(t, x) + h(t, x)) + \frac{\vartheta}{2}(\sigma_t(p(t)g_1(t, x) + h(t, x)))^2 \\ + \frac{1}{2}\text{Tr}(\sigma_t^2 p(t)^2 g_2(t, x)) + h_1(t, x) \end{aligned} \quad (5.4.2)$$

where

$$g_1(t, x) := \left. \frac{\partial h_0(T, y)}{\partial y} \right|_{y=p(t)x} \quad \text{and} \quad g_2(t, x) := \left. \frac{\partial^2 h_0(T, y)}{\partial y^2} \right|_{y=p(t)x}$$

Eq. 5.3.3 is separable if there exists a function $p : [0, T] \rightarrow \mathbb{R}^{d \times d}$ satisfying $p(T) = I$, such that the right-hand side of Eq. 5.4.2 is independent of x . Applying this technique of variable separation, we may solve the non-linear partial differential equation, Eq. 5.3.3, in a simple analytical way.

After solving Eq. 5.3.3 for $\tilde{u}(t, x)$, we will switch to the partial differential equation Eq. 5.3.4 to solve for $\tilde{v}(x, t)$. In a similar way, if $\tilde{v}(t, x)$ can be decomposed into

$$\tilde{v}(t, x) = h_0(T, q(t)x) + \int_t^T Q(s)ds \quad (5.4.3)$$

for some integrable function $Q : [0, T] \rightarrow \mathbb{R}$ and some function $q : [0, T] \rightarrow \mathbb{R}^{d \times d}$ subject to $q(T) = I$, then Eq. 5.3.4 is transformed to

$$\begin{aligned} Q(t) = g_1(t, x)q'(t)x + \mu_t(q(t)g_1(t, x) + h(t, x)) + \frac{1}{2}\text{Tr}(\sigma_t^2 q(t)^2 g_2(t, x)) \\ + \vartheta(p(t)g_1(t, x) + h(t, x))\sigma_t^2(q(t)g_1(t, x) + h(t, x)) + h_1(t, x) \end{aligned} \quad (5.4.4)$$

If there exists a function $q : [0, T] \rightarrow \mathbb{R}^{d \times d}$ satisfying $q(T) = I$, such that the right-hand side of Eq. 5.4.4 is independent of x , then we immediately get the expression for $\tilde{v}(x, t)$ by substituting Eq. 5.4.4 into Eq. 5.4.3.

We will apply the technique of variable separation to some simple examples, assuming a diffusive canonical process $(X_t)_{t \in [0, T]}$ whose drift and diffusion terms are independent of x , i.e. $\mu_t = \mu(t)$ and $\sigma_t = \sigma(t)$. First we consider the case when the terminal part of the total loss takes a linear form, i.e. $h_0(T, x) = ax + b$ (see Eq. 5.2.7). In addition, the path-dependent part in Eq. 5.2.7 is assumed to have rates of losses independent of x , so that we may write $h_1(t, x)$ and $h(t, x)$ as $h_1(t)$ and $h(t)$ respectively. According to these assumptions, we transform Eq. 5.4.2 into

$$P(t) = ap'(t)x + \mu(t)(ap(t) + h(t)) + \frac{\vartheta}{2}(\sigma(t)(ap(t) + h(t)))^2 + h_1(t)$$

To make the right-hand side independent of x , we have $p'(t) = 0$ which leads to $p(t) = I$ for all $t \in [0, T]$. Substituting the equation above into Eq. 5.4.1, we get the expression for $\tilde{u}(x, t)$,

$$\tilde{u}(x, t) = ax + \int_t^T \left(\mu(s)(a + h(s)) + \frac{\vartheta}{2}(\sigma(s)(a + h(s)))^2 + h_1(s) \right) ds + b$$

In a similar way, we solve Eqs. 5.4.3-5.4.4 to get the expression for $\tilde{v}(x, t)$,

$$\tilde{v}(x, t) = ax + \int_t^T \left(\mu(s)(a + h(s)) + \frac{\vartheta}{2}(\sigma(s)(a + h(s)))^2 + h_1(s) \right) ds + b$$

The budget process is given by

$$\eta_t^\vartheta = \vartheta(\tilde{v}(x, t) - \tilde{u}(x, t)) = \frac{\vartheta}{2} \int_t^T (\sigma(s)(a + h(s)))^2 ds$$

From the equations above, we may express the worst-case risk, $V_0 = \tilde{v}(0, 0)$, as a function of the relative entropy budget $\eta = \eta_0^\vartheta$:

$$V_0^\pm = b + \int_0^T (h_1(t) + \mu(t)(a + h(t))) dt \pm \sqrt{2\eta \int_0^T \sigma(t)^2 (a + h(t))^2 dt} \quad (5.4.5)$$

where the negative sign corresponds to the case where we minimize the expectation, accounting for model risk, instead of maximize it. $[V_0^-, V_0^+]$ may be regarded as a confidence interval (of the expected loss) resulted from the ambiguity of models.

This simple example resembles the risk of holding a portfolio of financial instruments (e.g. risky stocks and riskless bonds) along with their hedging instruments. The terminal risk function, $f(x) = ax + b$, reflects the risk of holding a fixed position a of the risky instruments and b of the riskless instruments. Loss incurred by the hedging positions, as part of the loss functional Eq. 5.2.7, is given by the stochastic integral, $\int_0^T h_1(t) dX(t)$, with a time-varying hedging ratio $h_1(t)$. Eq. 5.4.5 gives the maximum (or minimum) loss accounting for model ambiguity. We note that the component associated with the model risk, i.e. the last term in Eq. 5.4.5, increases with the square root of the relative entropy budget η . This is similar to a fixed percentile (say 5% and

95% percentiles) of a Brownian motion, which increases proportional to the square-root of time to maturity. In practice, one may regard the relative entropy budget as a cumulative quantity over time, which may be justified by certain information-theoretic consideration. It affects the worst-case risk just as the total variance does to the percentiles.

We now consider a path-independent loss function that is quadratic to the state variable x , i.e. $h_0(T, x) = x^T A x$ and $h_1 = h = 0$. For simplicity we assume A is a $d \times d$ symmetric matrix. According to Eq. 5.4.1, we guess the solution to have the form of

$$\tilde{u}(x, t) = x^T p(t)^T A p(t) x + \int_t^T P(s) ds$$

Let $H(t) := p(t)^T A p(t)$. H is a $d \times d$ symmetric matrix due to the symmetry of A . We will solve for $H(t)$ instead of $p(t)$. According to Eq. 5.4.2, we have

$$P(t) = x^T H'(t) x + 2\vartheta x^T H(t)^T \sigma(t)^2 H(t) x + \text{Tr}(\sigma(t)^2 A) \quad (5.4.6)$$

In the equation above, we only consider the driftless dynamics, i.e. $\mu(t) = 0$ for all $t \in [0, T]$. The right-hand side of Eq. 5.4.6 contains only quadratic terms of x . These terms vanish only if H satisfies the ordinary differential equation,

$$H'(t) + 2\vartheta H(t)^T \sigma(t)^2 H(t) = 0$$

subject to $H(T) = A$ by virtue of $p(T) = I$. Solving the equation gives $H(t) = A(I - 2\vartheta A \mathcal{V}(t))^{-1}$ in which

$$\mathcal{V}(t) := \int_t^T \sigma(s)^2 ds$$

Substituting the solution into Eq. 5.4.6, we get $P(t) = \text{Tr}(\sigma(t)^2 (I - 2\vartheta A \mathcal{V}(t))^{-1})$. Eq. 5.4.1 then gives

$$\begin{aligned} \tilde{u}(x, t) &= x^T A (I - 2\vartheta A \mathcal{V}(t))^{-1} x - \int_t^T \text{Tr}((I - 2\vartheta A \mathcal{V}(s))^{-1} A d\mathcal{V}(s)) \\ &= x^T A (I - 2\vartheta A \mathcal{V}(t))^{-1} x + \frac{1}{2\vartheta} \int_t^T d \ln |I - 2\vartheta A \mathcal{V}(s)| \\ &= x^T A (I - 2\vartheta A \mathcal{V}(t))^{-1} x - \frac{1}{2\vartheta} \ln |I - 2\vartheta A \mathcal{V}(t)| \end{aligned}$$

The worst-case risk $\tilde{v}(x, t)$ is calculated by solving Eq. 5.4.4. In a similar way, we solve the following ordinary differential equation w.r.t $G(t) := q(t)^T A q(t)$,

$$G'(t) + 4\vartheta (I - 2\vartheta A \mathcal{V}(t))^{-1} A^T \sigma(t)^2 G(t) = 0$$

subject to $G(T) = A$. The equation has a solution $G(t) = A(I - 2\vartheta A \mathcal{V}(t))^{-2}$. Substituting this into Eq. 5.4.4, we get

$$\tilde{v}(x, t) = x^T A (I - 2\vartheta A \mathcal{V}(t))^{-2} x - \int_t^T \text{Tr}((I - 2\vartheta A \mathcal{V}(s))^{-2} A d\mathcal{V}(s))$$

$$= x^T A(I - 2\vartheta AV(t))^{-2}x + \frac{1}{2\vartheta} \text{Tr}((I - 2\vartheta AV(t))^{-1} - I)$$

The process of relative entropy is given by

$$\eta_t = 2\vartheta^2 x^T A(I - 2\vartheta AV(t))^{-2} AV(t)x + \frac{1}{2} (\text{Tr}((I - 2\vartheta AV(t))^{-1} - I) + \ln |I - 2\vartheta AV(t)|)$$

This result can be applied to the problem of evaluating the model risk of portfolio variance. To evaluate the portfolio variance, the symmetric matrix $A = aa^T$, where a is the vector of portfolio weights. The equations above provide the worst-case (maximum) portfolio variance and the corresponding relative entropy budget as functions of ϑ . They serve as a dynamical generalization of the results derived in literature (Glasserman and Xu 2014) using the joint distribution of x . It is noted that here we use a completely different approach. Instead of dealing with the distribution explicitly, we solve the partial differential equations that govern the worst-case risk. This approach applies to a wider range of problems, as illustrated in the next section.

5.5 Robust Pricing of Equity Swap

We illustrate the approach of partial differential equations using a simple example. Consider an equity swap that exchanges fixed-rate payments with equity returns regularly. The equity price is assumed to follow a geometric Brownian motion:

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

under the reference model. The swap is assumed to be fairly priced under the same model, i.e. with an initial value of zero. To utilize the convenience of continuous-time approaches, we assume continuous cash flow payments. This allows us to express the present value of the equity leg by

$$P_e = \mathbb{E} \left(\int_0^T \frac{e^{-rt}}{X_t} dX_t \right)$$

where the interest rate r is assumed to be a constant. To manage the risk associated with the equity model, we need to evaluate the maximum (or minimum) value under alternative models ¹. We may therefore define the loss functional by

$$\ell(t, \cdot) = \int_0^t \frac{e^{-rs}}{X_s} dX_s$$

¹In this and following applications, we assume that the holder of the risky asset does not hedge the risk. In fact, if it is allowed to hedge with the underlying asset continuously, then there is zero market risk under the reference model. As f -divergence only works for equivalent measures under which the market risk is zero, our approach would result in zero model risk. In addition, to allow for alternative models we relax the martingale condition of the discounted price of the underlying asset.

which is a special form of Eq. 5.2.7. According to Eqs. 5.3.3-5.3.4, we may remove the path dependency and write down the partial differential equations for $\tilde{u}_t = \tilde{u}(t, x)$ and $\tilde{v}_t = \tilde{v}(t, x)$ that satisfy the terminal conditions $\tilde{u}_T = \tilde{v}_T = 0$,

$$\begin{aligned} \frac{\partial \tilde{u}_t}{\partial t} + \mu x \left(\frac{\partial \tilde{u}_t}{\partial x} + \frac{e^{-rt}}{x} \right) + \frac{\vartheta \sigma^2 x^2}{2} \left(\frac{\partial \tilde{u}_t}{\partial x} + \frac{e^{-rt}}{x} \right)^2 + \frac{\sigma^2 x^2}{2} \frac{\partial^2 \tilde{u}_t}{\partial x^2} &= 0 \\ \frac{\partial \tilde{v}_t}{\partial t} + \mu x \left(\frac{\partial \tilde{v}_t}{\partial x} + \frac{e^{-rt}}{x} \right) + \vartheta \sigma^2 x^2 \left(\frac{\partial \tilde{u}_t}{\partial x} + \frac{e^{-rt}}{x} \right) \left(\frac{\partial \tilde{v}_t}{\partial x} + \frac{e^{-rt}}{x} \right) + \frac{\sigma^2 x^2}{2} \frac{\partial^2 \tilde{v}_t}{\partial x^2} &= 0 \end{aligned}$$

Despite of the non-linearity, the equations above indeed have fairly simple solutions:

$$\begin{aligned} \tilde{u}_t &= \mu \frac{e^{-rt} - e^{-rT}}{r} + \vartheta \sigma^2 \frac{e^{-2rt} - e^{-2rT}}{4r} \\ \tilde{v}_t &= \mu \frac{e^{-rt} - e^{-rT}}{r} + \vartheta \sigma^2 \frac{e^{-2rt} - e^{-2rT}}{2r} \end{aligned}$$

The relative entropy budget

$$\eta_t = \vartheta(\tilde{v}_t - \tilde{u}_t) = \vartheta^2 \sigma^2 \frac{e^{-2rt} - e^{-2rT}}{4r}$$

At $t = 0$, the worst-case risk $V_0 = \tilde{v}_0$ reflects the maximum (or minimum if $\vartheta < 0$) value of the equity leg, controlled by the parameter ϑ . If we set ϑ to zero to prohibit any alternative model, V_0 should be the fair price of the equity leg under the reference model. This equals the price of the fixed-rate leg assuming the swap has an initial value of zero. As a result, the maximum (or minimum) value of the swap accounting for model risk is

$$\vartheta \sigma^2 (2r)^{-1} (1 - e^{-2rT})$$

accompanied by the relative entropy budget

$$\eta = \eta_0 = \vartheta^2 \sigma^2 (4r)^{-1} (1 - e^{-2rT})$$

We may substitute ϑ by its expression w.r.t η to obtain the range of the swap price, accounting for model risk, as a function of the relative entropy budget:

$$\left[-\sigma \sqrt{\frac{1 - e^{-2rT}}{r}} \eta, \sigma \sqrt{\frac{1 - e^{-2rT}}{r}} \eta \right]$$

The dependence of this interval on the model parameters is illustrated by Fig. 5.5.1.

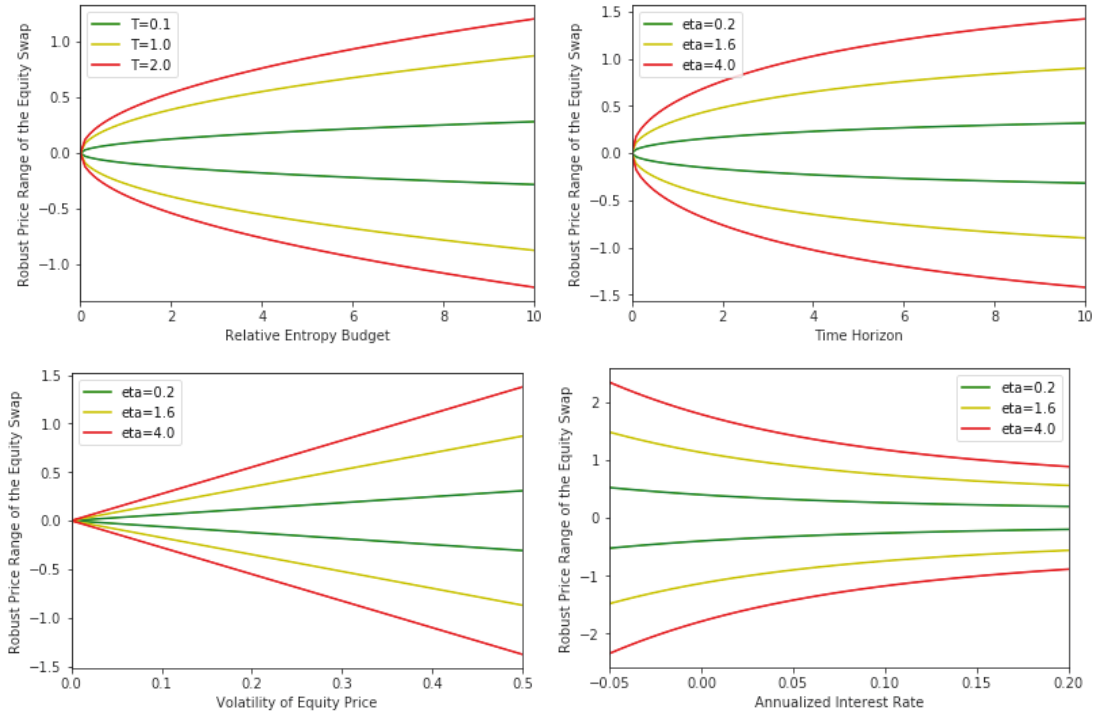


Fig 5.5.1: (Upper Left) Robust price range of the hypothetical equity swap as function of the relative entropy η at different time horizons T . (Upper Right) Price range as function of time horizons at three η levels. (Lower Left) Price range as function of the volatility σ . (Lower Right) Price range as function of the interest rate r .

An interesting observation is that the worst-case risk does not depend on the initial value of the state variable (i.e. the stock price). This is because the volatility is independent of x , decoupling the equity return from the price level. This is no longer the case when we consider diffusion processes in general. In fact, we may numerically evaluate the model risk assuming a more sophisticated local volatility model. For instance, we may consider about the CEV model:

$$dX_t = \mu X_t dt + \sigma X_t^\gamma dW_t \tag{5.5.1}$$

When $\gamma < 1$, the volatility decreases with X_t , implying a leverage effect that is common in equity markets. On the other hand, $\gamma > 1$ implies high volatility with high price, a common feature of commodity markets. Here we adopt $\gamma = 0.5$ and $\gamma = 1.1$ for numerical demonstration. Other parameters are the interest rate $r = 0.05$, drift $\mu = 0.05$, volatility $\sigma = 0.2$ and time horizon $T = 1$. The dependence of this interval on the model parameters is illustrated by Fig. 5.5.2.

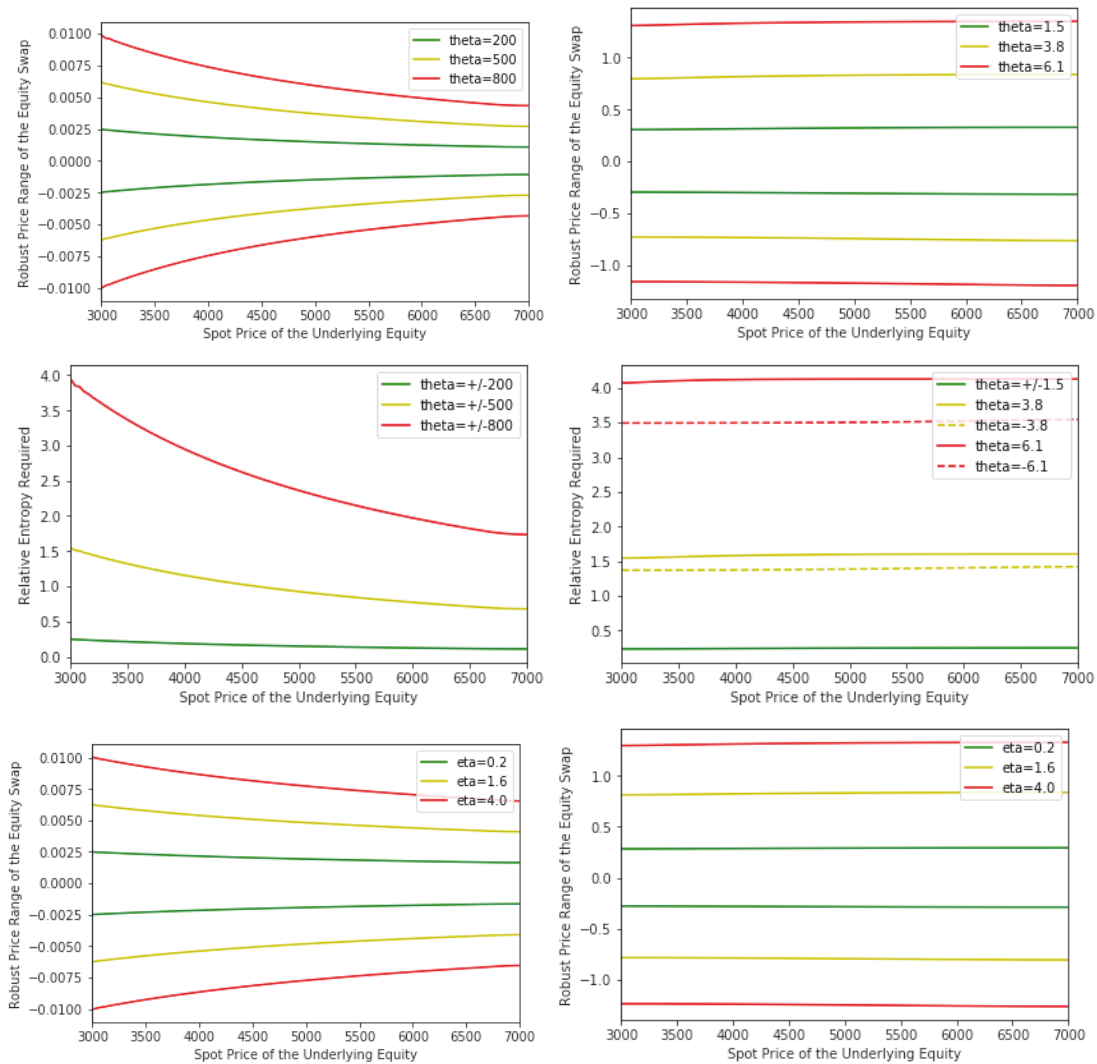


Fig 5.5.2: Robust pricing for the hypothetical equity swap under the CEV model with (Left) $\gamma = 0.5$ and (Right) $\gamma = 1.1$. Upper figures are price ranges changing with the equity price at three different parametric significance levels. Middle figures show the corresponding cost measured by the relative entropy. Bottom figures show price ranges at three different entropic significance levels.

5.6 Robust European Option Pricing

In many scenarios the loss function is contingent on the state variable, usually expressed by some piece-wise function. In general, the model risk problem under these scenarios cannot be solved analytically. Here we will approach the problem using numerical routines, including the finite difference scheme, the tree approach and the Monte Carlo simulation.

In this section, we attempt to measure the model risk of a short position in a European call option. The loss function is path-independent, since it includes only the terminal value $h_0(T, x) = (x - K)^+$. We further assume the canonical process $(X_t)_{t \in [0, T]}$ follows the geometric Brownian motion

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

The diffusive nature allows us to solve the problem of model risk quantification using a finite difference scheme.

In the following example, we parametrize the reference model by setting $\mu = 0$ and $\sigma = 0.2$. The strike price is $K = 5000$ and the loss is realised after $T = 0.1$. We choose to use $\vartheta = 0.005$ as it demonstrates significant model risk. The worst-case risk², $\tilde{v}(0, X_0)$, at different spot prices, X_0 , are plotted in Fig. 5.6.1(a), in comparison to the nominal risk which follows the Black-Scholes model.

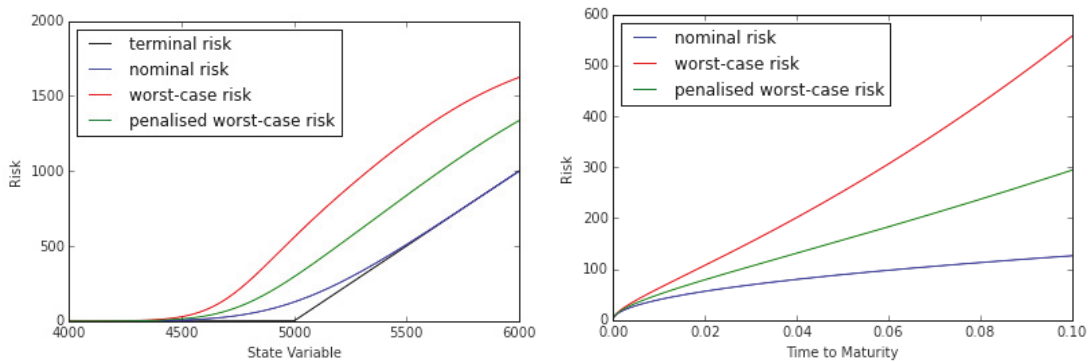


Fig 5.6.1: (a) worst-case risk, $\tilde{v}(0, X_0)$, at different spot prices X_0 in comparison to the nominal risk which follows the Black-Scholes PDE, (b) risks for a fixed spot price, $\tilde{v}(t, X_0)$ where $X_0 = 5000$, at different times to maturity $T - t$.

An advantage of the finite difference approach is that it solves the full grid of model risks along two dimensions, the time and the state variable (Fig. 5.6.2). Fig. 5.6.2(a) and (b) provide the bivariate functions of the worst-case risk, $\tilde{v}(t, x)$, and $\tilde{u}(t, x)$. Note that we term $\tilde{u}(t, x)$ as the penalized worst-case risk as it is indeed $\tilde{v}(t, x)$ subtracting the penalization term associated with the relative entropy. Fig. 5.6.2(c) illustrates the extra drift term in Eq. 6.3.18 that, according to the Girsanov's theorem, transforms the reference measure to the worst-case measure. The relative entropy, calculated by $\vartheta(\tilde{v}(t, x) - \tilde{u}(t, x))$, is given in Fig. 5.6.2(d).

²“Worst-case risk” is consist with the terminology used in Chapter 4 where $h_0(T, x)$ is the loss function. Here it is more suitable to refer $\tilde{v}(0, X_0)$ as the “worst-case price” as $h_0(T, x)$ is instead the payoff function of a derivative contract.

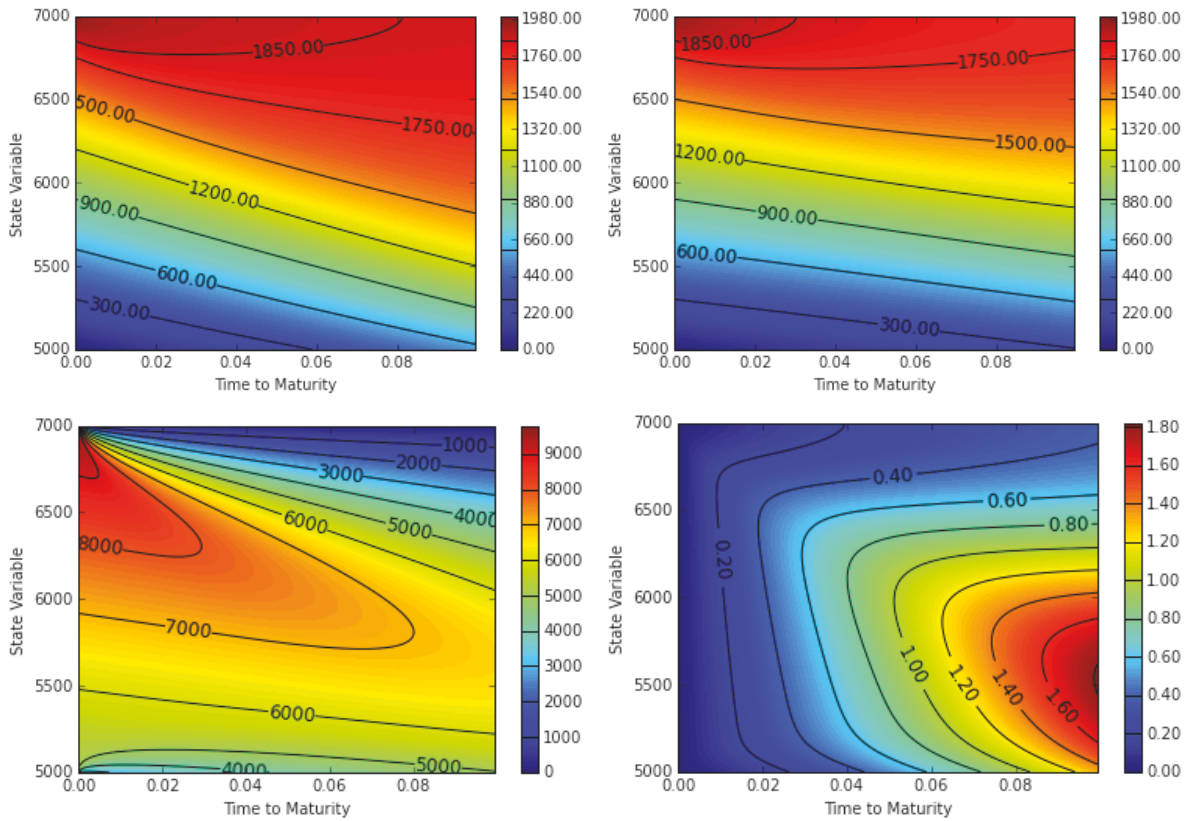


Fig 5.6.2: Two-dimensional grids of (a) the worst-case risk $\tilde{v}(t, x)$, (b) the penalized worst-case risk $\tilde{u}(t, x)$, (c) the drift term for the change of measure, (b) the relative entropy $\vartheta(\tilde{v}(x, t) - \tilde{u}(x, t))$.

Being a more general approach, the tree model applies to dynamics not restricted by the diffusion processes. Here though we consider the simple case of geometric Brownian motion, which is modelled by a CRR tree (Cox et al. 1979). We apply the backward induction formulated by Eqs. 5.2.9-5.2.10 to calculate the two processes $(\tilde{U}_t^\vartheta)_{t \in [0, T]}$ and $(\tilde{V}_t^\vartheta)_{t \in [0, T]}$ defined in Eq. 5.2.8. Fig. 5.6.3 illustrates an 8-layer example of the tree approach.

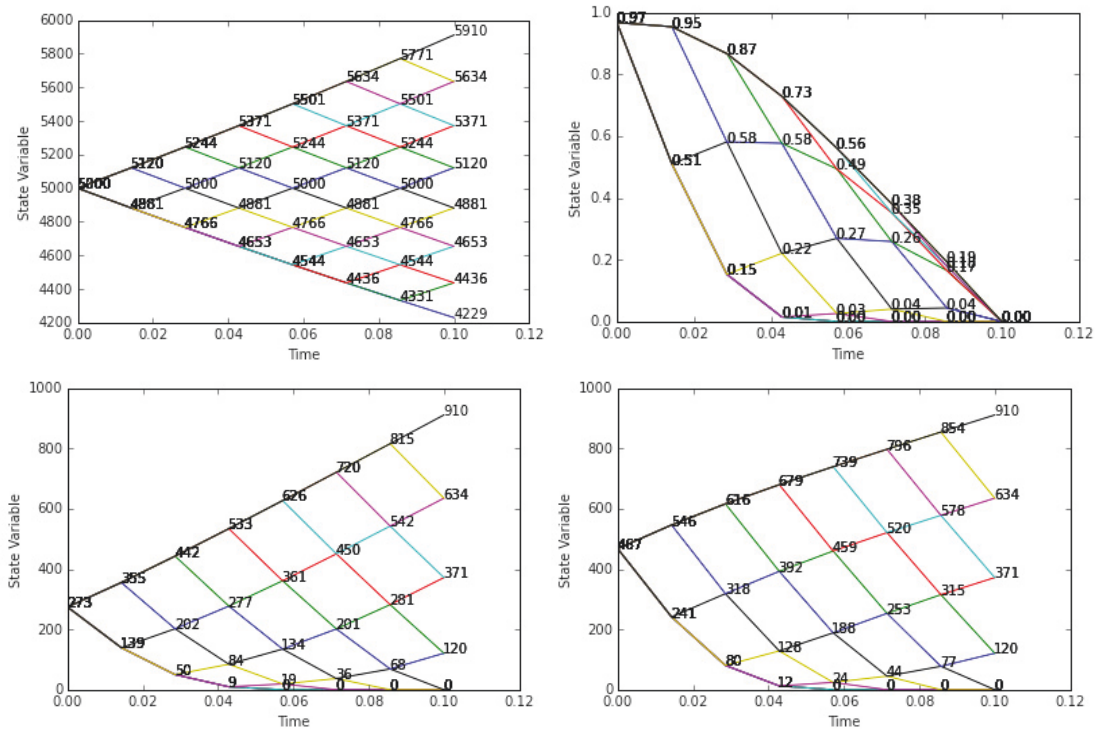


Fig 5.6.3: 8-layer trees of (a) the price process of the underlying asset, (b) the budget process calculated by $\eta_t^\vartheta = \vartheta(\tilde{V}_t^\vartheta - \tilde{U}_t^\vartheta)$, (c) the value process \tilde{U}_t^ϑ , and (d) the risk process \tilde{V}_t^ϑ .

We conduct the convergence test by running the tree approach for an increasing sequence of discretization levels. The worst-case risk, \tilde{V}_0^ϑ , and the relative entropy, η_0^ϑ , are plotted in Fig. 5.6.4 versus the number of layers in the CRR tree. Both quantities converge monotonically, providing sufficient accuracy with 500 layers.

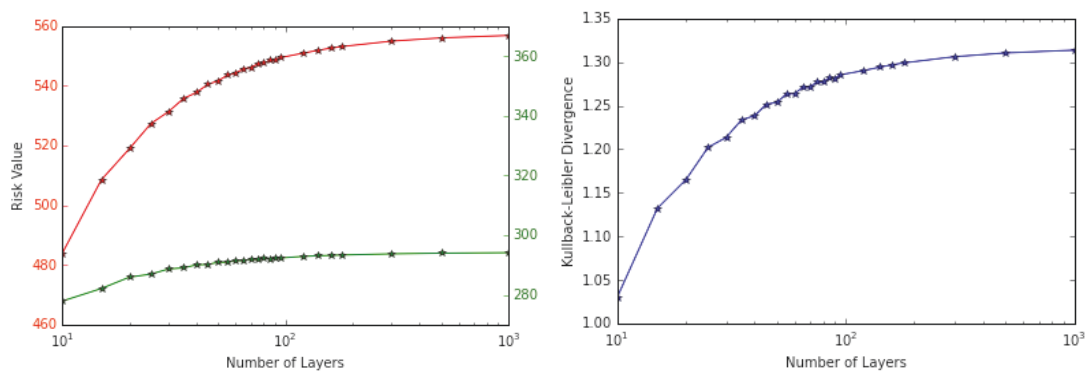


Fig 5.6.4: Convergence of (a) \tilde{V}_0^ϑ and \tilde{U}_0^ϑ , and (b) η_0^ϑ , with the number of layers in a CRR binomial tree, on the robust pricing of an European option.

In comparison with the finite difference and the tree methods, we also approach the same problem with the Monte Carlo simulation. This generic approach has been adopted in literature (Glasserman and Xu 2014,) to evaluate the model risk in hedging. Compared to the finite difference and the tree approach, it is computationally intensive and may produce inconsistent results. In Fig. 5.6.5, the final state variable is sampled according to the log-normal distribution. It appears that the convergence of the MC simulation is relatively slow. Even if the confidence interval for the nominal risk is tight, the confidence interval for the worst-case risk is still not satisfactory. If the dynamics is complicated enough for an analytical marginal distribution to be unavailable, a significant discretisation error could appear unless we sample frequently in time. This makes the MC simulation intractable.

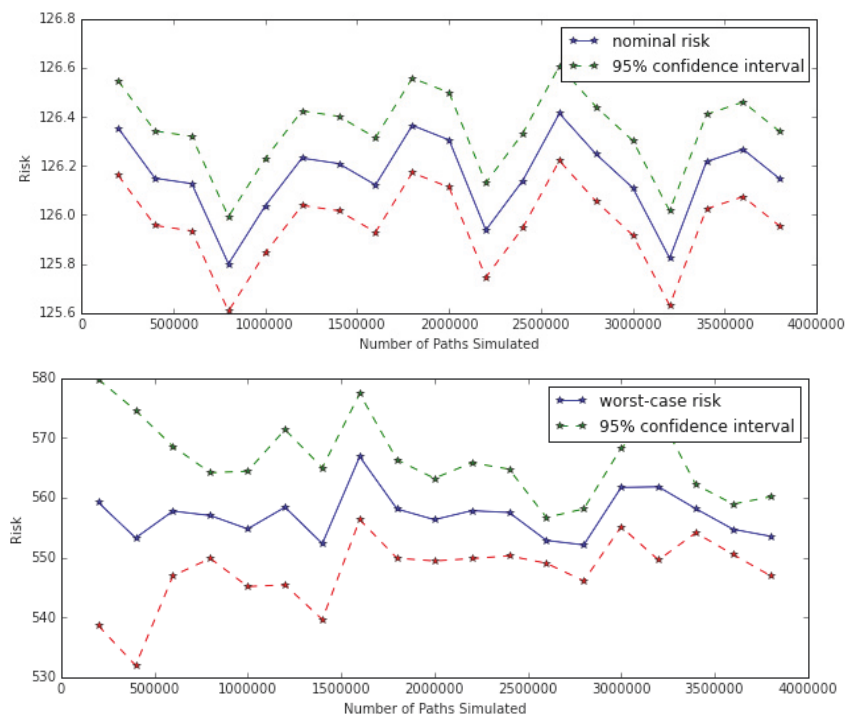


Fig 5.6.5: Convergence of (a) the nominal risk and (b) the worst-case risk simulated with different numbers of paths.

Compared with the MC simulation, the (recombining) tree approach and the finite difference scheme have a few advantages:

1. Consistency between time discretisation and path sampling
2. The number of sampled paths increases exponentially with the number of time steps
3. Deterministic method with proper probability assigned to each node / path
4. Smooth convergence with monotonically increasing worst-case risk and entropy
5. Iterative algorithm allows for computational efficiency
6. Quantities over the entire grid are obtained simultaneously

5.7 Robust American Option Pricing

Now let us consider measuring the model risk of an American option. This is a typical problem that has a free boundary (the loss is realised at a stopping time $\tau \leq T$). In the example of American options, alternative measures may modify the expectation of future payoff, thus changing the decision of early exercise. We may formulate the problem using the “adversary” interpretation (Glasserman and Xu 2014), which assumes that a fictional adversary (e.g. the counterparty) manipulates the measure aiming to maximise the risk. A free boundary problem is therefore classified into two categories based on the owner of the right of early termination. Now we first consider the case where we short an American option, and the adversary would switch the probability measure and choose the time for early exercise. The model risk problem is formulated as

$$\max_{Z, \tau} \mathbb{E}^{\mathbb{Q}^Z}((X_\tau - K)^+)$$

subject to the relative entropy constraint

$$\mathbb{E}(Z_T \ln Z_T) \leq \eta$$

On the other hand, if we possess the right of early termination (we are long the option), we may formulate the problem as a zero-sum game

$$\max_Z \min_\tau \mathbb{E}^{\mathbb{Q}^Z}((X_\tau - K)^+)$$

The equilibrium solution does sometimes involve mixed strategies though, which complicates the problem. In this example we consider the former case only.

In the example of American options, the decision of early exercise depends on the comparison between the immediate payoff and the expected future payoff. It is not straightforward to solve this type of problem using Monte Carlo simulation, as individual paths do not provide sufficient information on the future expectation. However, the tree approach is readily applicable as we evaluate the expected future payoff of the option at every node. To apply the tree approach, we construct the binomial tree for the canonical process, which is a geometric Brownian motion with $\mu = -0.55$ and $\sigma = 0.2$. Note that the negative drift term models a continuous flow of dividends. Without it the adversary (who long the option) would never exercise the option earlier than its maturity. The strike price is $K = 5000$ and the time to expiration is $T = 0.1$, and $\vartheta = 0.005$.

Starting from the initial price $X_0 = 5000$, we construct the CRR tree and iterate backwards to calculate the value process U^ϑ using the following rule,

$$U_{t-1}^{(k)} = \max \left((X_{t-1}^{(k)} - K)^+, \frac{1}{\vartheta} \ln \left(\sum_{j=k}^{k+n-1} p(t-1, k; t, j) e^{\vartheta U_t^{(j)}} \right) \right) \quad (5.7.1)$$

where $X_t^{(k)}$ and $U_t^{(k)}$ denote the values assigned to the k -th node at the t -th layer for the price process of the underlying asset and the value process, respectively. $p(t-1, k; t, j)$ is the probability (under the reference model) that connects the k -th node at the $(t-1)$ -th layer to the j -th node at the t -th layer. Eq. 5.7.1 allows us to determine whether or not a given node, say $(t-1, k)$, is an optimal exercise point. If $U_{t-1}^{(k)}$ takes the former value in Eq. 5.7.1, then the adversary would exercise the option, leaving $V_{t-1}^{(k)} = (X_{t-1}^{(k)} - K)^+$ and $\eta_{t-1}^{(k)} = 0$. Otherwise, $V_{t-1}^{(k)}$ and $\eta_{t-1}^{(k)}$ follow as if there is no possibility of early exercise (Eq. 6.3.15).

Fig. 5.7.1 illustrates a simple tree model with 8 layers of nodes. The zig-zag of the tree that models the budget process η^ϑ is due to early termination. Any node that triggers exercise (usually at the upper branches of the tree) automatically reverts the future dynamics back to the nominal one, reducing the overall relative entropy.

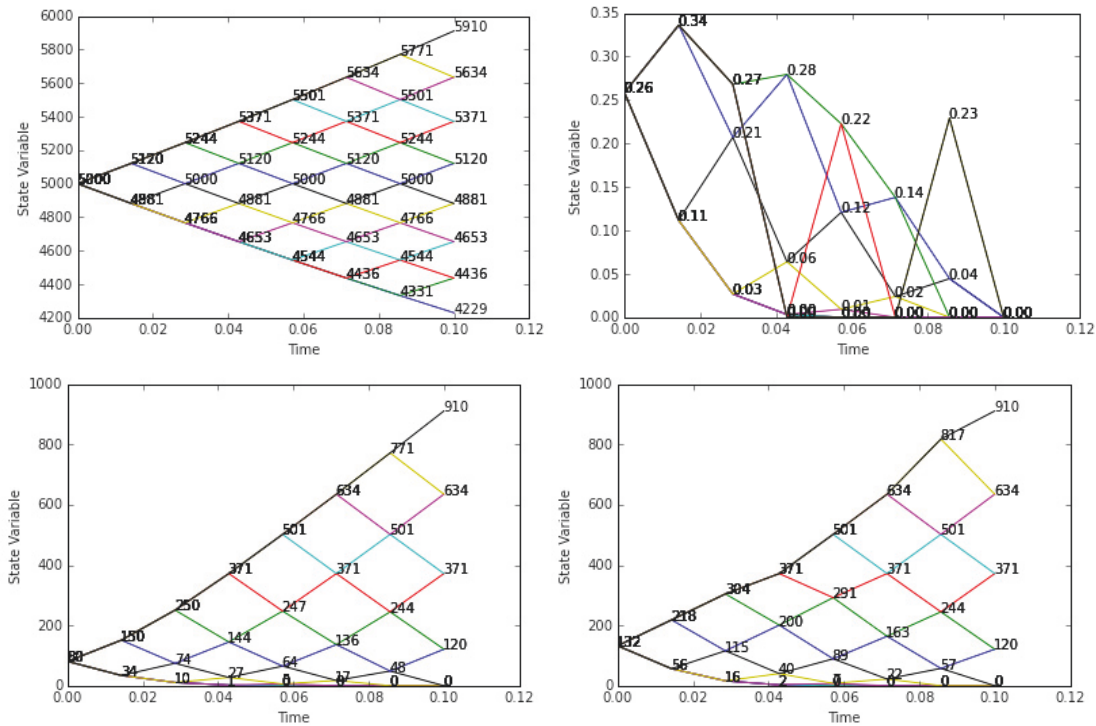


Fig 5.7.1: 8-layer trees of (a) the price process of the underlying asset, (b) the budget process η_t^ϑ , (c) the value process \tilde{U}_t^ϑ , and (d) the risk process \tilde{V}_t^ϑ .

The tree model adopted here demonstrates oscillation corresponding to an odd-even switch of the tree structure. Compared to Fig. 5.6.4, the oscillation is much more significant, as the decision of early exercise relies heavily on a finer evaluation grid. Nevertheless, the worst-case risk and the relative entropy both converge when the number of layers approaches 500.

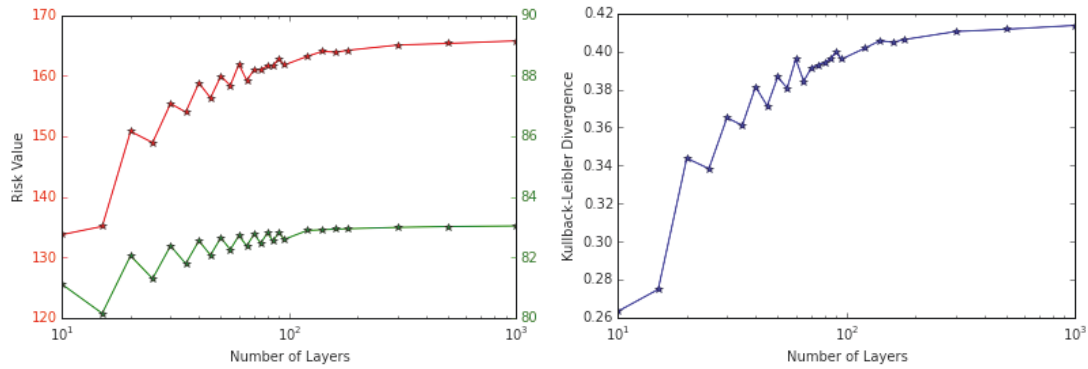


Fig 5.7.2: Convergence of (a) \tilde{V}_0^ϑ and \tilde{U}_0^ϑ , and (b) η_0^ϑ , with the number of layers in a CRR binomial tree, on the robust pricing of an American option.

The robust pricing of an American option may also be tackled using the finite difference scheme, as illustrated by Fig. 5.7.3 and Fig. 5.7.4. Fig. 5.7.3 shows that the penalised worst-case risk, $\tilde{u}(0, X_0)$, is only slightly higher than the nominal risk (i.e. the price of the American option under the reference model), resulted from the probable early termination.

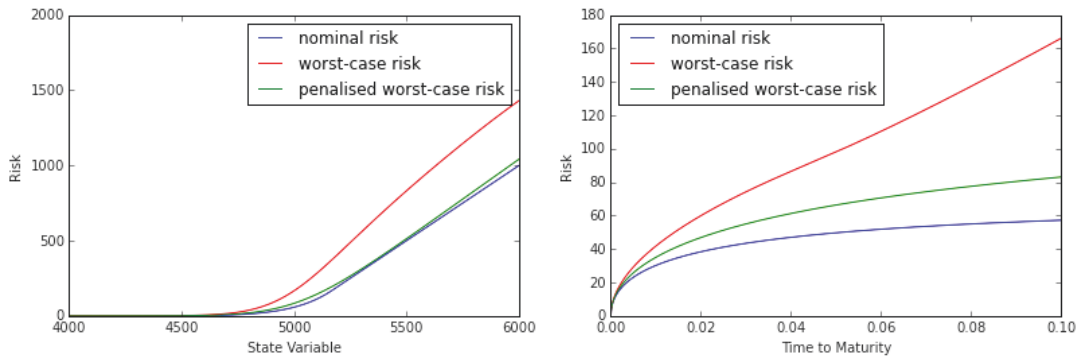


Fig 5.7.3: (a) worst-case risk, $V_0^\vartheta = \tilde{v}(0, X_0)$, at different spot prices X_0 in comparison to the nominal risk which follows the Black-Scholes PDE, (b) risks for a fixed spot price, $\tilde{v}(t, X_0)$ where $X_0 = 5000$, at different times to maturity $T - t$.

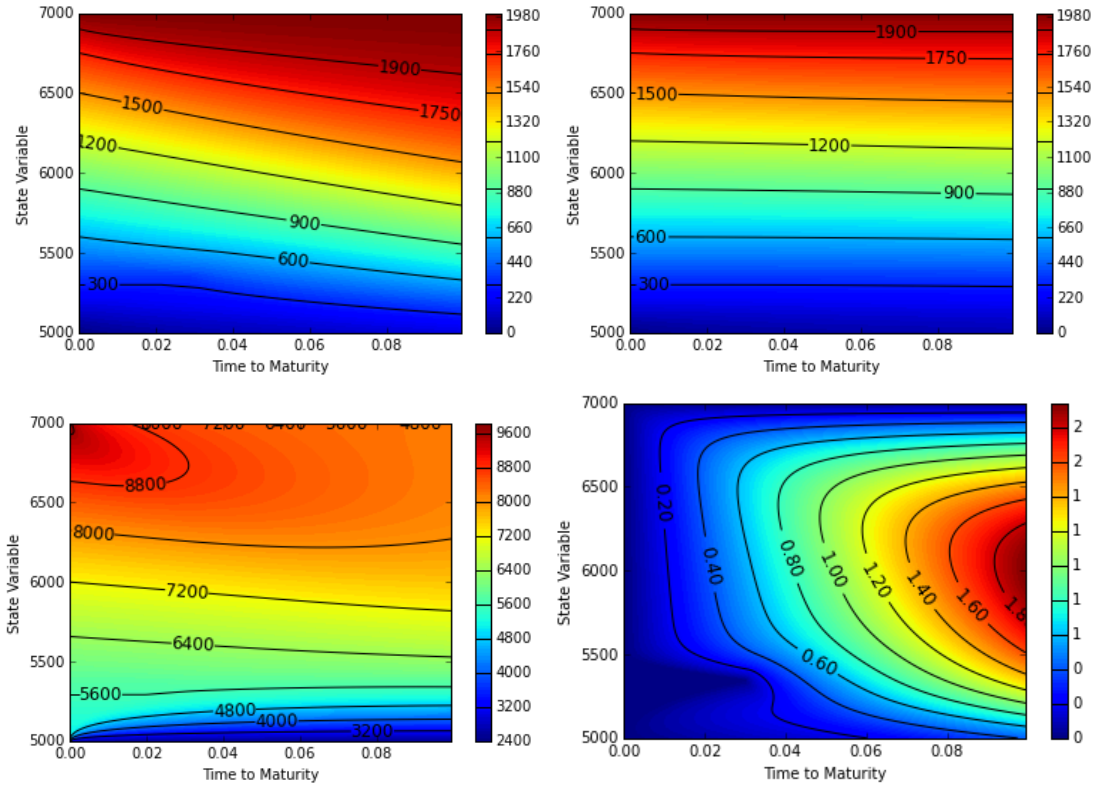


Fig 5.7.4: Two-dimensional grids of (a) the worst-case risk $\tilde{v}(t, x)$, (b) the penalized worst-case risk $\tilde{u}(t, x)$, (c) the drift term for the change of measure, (d) the relative entropy $\vartheta(\tilde{v}(x, t) - \tilde{u}(x, t))$.

5.8 Model Risk for Dynamic Portfolio Management

Another example is on portfolio rebalancing strategies. Suppose our investment universe included n risky assets whose prices are denoted by the vector X'_t following a Ito process

$$dX'_t = \mu'(t, X'_t)X'_t dt + \sigma'(t, X'_t)X'_t dW_t$$

Let $X_t := \ln(X'_t/X'_0)$ be the logarithmic return. It follows from Ito's Lemma that

$$\begin{aligned} dX_t &= \frac{dX'_t}{X'_t} - \frac{d[X']_t}{2X_t'^2} \\ &= \left(\mu'(t, X'_0 e^{X_t}) - \frac{\sigma'(t, X'_0 e^{X_t})^2}{2} \right) dt + \sigma'(t, X'_0 e^{X_t}) dW_t \end{aligned}$$

For simplicity, we denote $\mu'(t, X'_0 e^{X_t})$ by $\mu_t := \mu(t, X_t)$ and $\sigma'(t, X'_0 e^{X_t})$ by $\sigma_t := \sigma(t, X_t)$. We consider the wealth process given by the SDE

$$\begin{aligned} dY_t &= Y_t \Delta(t, X_t) \frac{dX'_t}{X'_t} + r_t Y_t (1 - 1' \Delta(t, X_t)) dt \\ &= Y_t \Delta(t, X_t) dX_t + Y_t \left(r_t (1 - 1' \Delta(t, X_t)) + \text{Tr} \left(\frac{\Delta(t, X_t)^2 \sigma_t^2}{2} \right) \right) dt \end{aligned}$$

where Y_t denotes the total wealth available for investment at time t . $\Delta(t, X_t)$ is a vector that contains the portfolio weights invested in risky assets, assuming it only depends on t and the vector of logarithmic return X_t . The rest of the total wealth, $Y_t(1 - 1' \Delta(t, X_t))$, generates interest at the risk-free rate r_t . The target here is to manage the worst possible loss terminated at a fixed time T , measured by the negative logarithmic return. Applying Ito's Lemma one gets

$$d \ln Y_t = \Delta(t, X_t) dX_t + r (1 - 1' \Delta(t, X_t)) dt$$

The loss function is therefore given by

$$\ell(T, \cdot) := -\ln \frac{Y_t}{Y_0} = \int_0^T r_t (1' \Delta(t, X_t) - 1) dt - \int_0^T \Delta(t, X_t) dX_t$$

the partial differential equation for $\tilde{u}(x, t)$ takes the form

$$\begin{aligned} \frac{\partial \tilde{u}}{\partial t} + \mu_t \left(\frac{\partial \tilde{u}}{\partial x} - \Delta \right) + \frac{\sigma_t^2}{2} \frac{\partial^2 \tilde{u}}{\partial x^2} + \frac{\vartheta \sigma_t^2}{2} \left(\frac{\partial \tilde{u}}{\partial x} - \Delta \right)^2 - r_t (1 - 1' \Delta) &= 0 \\ \frac{\partial \tilde{v}}{\partial t} + \mu_t \left(\frac{\partial \tilde{v}}{\partial x} - \Delta \right) + \frac{\sigma_t^2}{2} \frac{\partial^2 \tilde{v}}{\partial x^2} + \vartheta \sigma_t^2 \left(\frac{\partial \tilde{u}}{\partial x} - \Delta \right) \left(\frac{\partial \tilde{v}}{\partial x} - \Delta \right) - r_t (1 - 1' \Delta) &= 0 \end{aligned}$$

In the case that μ_t , σ_t and r_t depend only on time t and the asset allocation strategy takes a linear form $\Delta(t, x) = a(t) + b(t)x$, we may convert the two PDEs to six ODEs. In particular, $a(t) \in \mathbb{R}^n$ and $b(t) \in \mathbb{R}^{n \times n}$, and the solution of the PDE take the form of

$$\begin{aligned} \tilde{u}(t, x) &= \text{Tr}(d(t)x^2) + e(t)x + f(t) \\ \tilde{v}(t, x) &= \text{Tr}(i(t)x^2) + j(t)x + k(t) \\ \eta(t, x) &= \vartheta (\text{Tr}((i(t) - d(t))x^2) + (j(t) - e(t))x + k(t) - f(t)) \end{aligned}$$

where $d(t), i(t) \in \mathbb{R}^{n \times n}$, $e(t), j(t) \in \mathbb{R}^n$ and $f(t), k(t) \in \mathbb{R}$, subject to $d(T) = e(T) = f(T) = i(T) = j(T) = k(T) = 0$. By collecting terms with the same order of x , one converts the two PDEs to six ODEs (with respect to t):

$$\begin{aligned}
d'(t) + \frac{\vartheta}{2} \sigma_t^2 (2d(t) - b(t))^2 &= 0 \\
e'(t) + \mu_t (2d(t) - b(t)) + \vartheta \sigma_t^2 (2d(t) - b(t))(e(t) - a(t)) - 1b(t)r_t &= 0 \\
f'(t) + \mu_t (e(t) - a(t)) + \text{Tr} \left(\sigma_t^2 d(t) + \frac{\vartheta \sigma_t^2}{2} (e(t) - a(t))^2 \right) - r_t (1 - 1a(t)) &= 0 \\
i'(t) + \vartheta \sigma_t^2 (2d(t) - b(t))(2i(t) - b(t)) &= 0 \\
j'(t) + \vartheta \sigma_t^2 ((2d(t) - b(t))(j(t) - a(t)) + (2e(t) - a(t))(2i(t) - b(t))) & \\
+ \mu_t (2i(t) - b(t)) - 1b(t)r_t &= 0 \\
k'(t) + \mu_t (j(t) - a(t)) + \sigma_t^2 i(t) + \text{Tr} (\vartheta \sigma_t^2 (e(t) - a(t))(j(t) - a(t))) - r_t (1 - a(t)) &= 0
\end{aligned}$$

In particular, the solution at $b(t) \equiv 0$ is given by

$$\begin{aligned}
\tilde{u}(t, \cdot) &= \int_t^T (r_s (a(s) - 1) - \mu_s a(s)) ds + \frac{\vartheta}{2} \int_t^T \sigma_s^2 a(s)^2 ds \\
\tilde{u}(t, \cdot) &= \int_t^T (r_s (a(s) - 1) - \mu_s a(s)) ds + \vartheta \int_t^T \sigma_s^2 a(s)^2 ds \\
\eta(t, \cdot) &= \frac{\vartheta^2}{2} \int_t^T \sigma_s^2 a(s)^2 ds
\end{aligned}$$

The results are presented in Fig. 5.8.2, showing an optimal range of (a, b) if accounting for the robustness in optimisation. In the reference model, we prefer to hold more risky asset (larger a) and tends to hold even more when price goes up (larger b). But this strategy is very risky because the reference model is a poor toy model. However, even if we implement a sophistic model we still need to keep in mind model uncertainty, which tends to shift the optimal range to more conservative area.

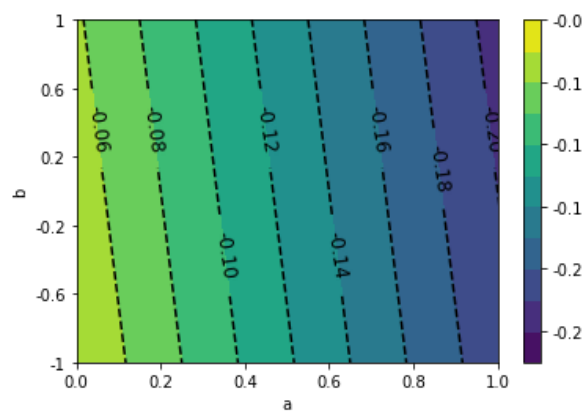


Fig 5.8.1: Expected loss (in terms of negative logarithmic return) over the investment horizon (time 0 to T) under the reference model, as a function of the rebalancing strategy $\Delta(x) = a + bx$.

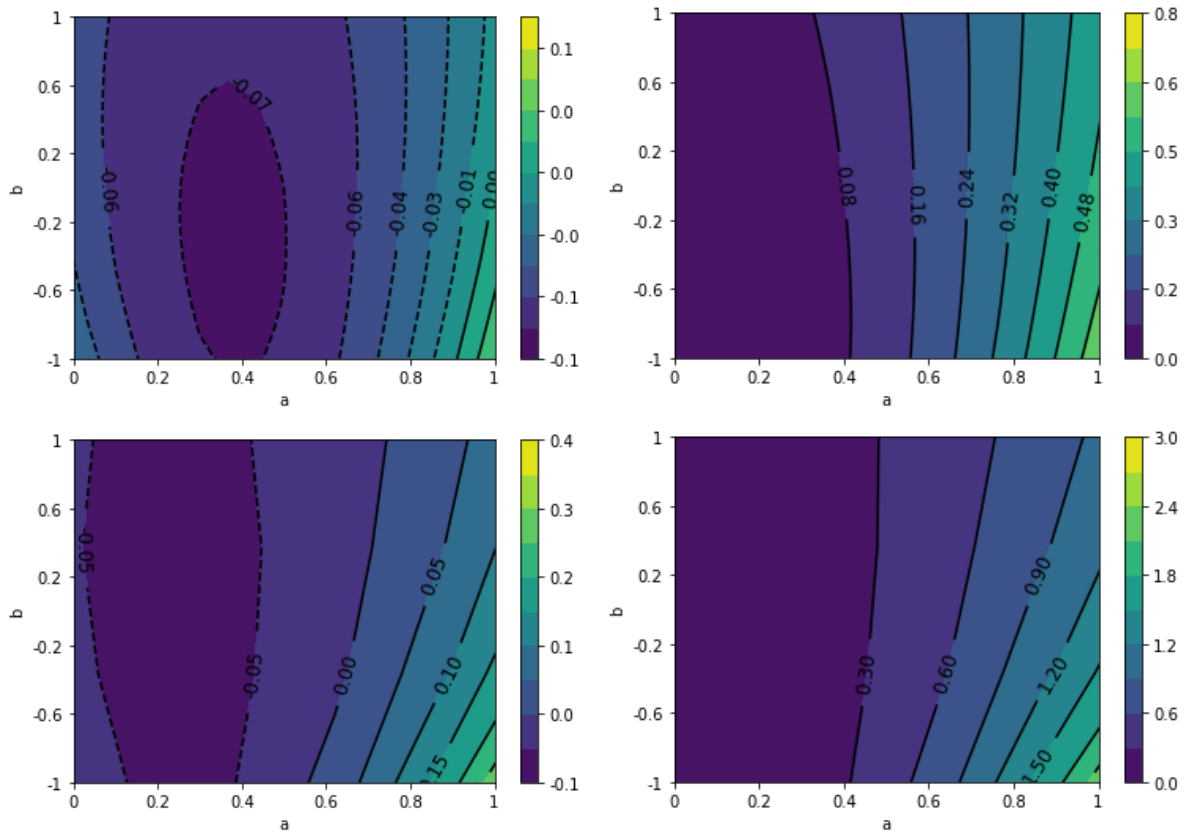


Fig 5.8.2: Expected loss (Left) and the associated relative entropy (Right) over the investment horizon (time 0 to T) under the worst-case model, as a function of the rebalancing strategy $\Delta(x) = a + bx$. Upper figures are at the parametric significance level of $\vartheta = 5$. Lower figures are at $\vartheta = 8$. Parameter choices are interest rate $r = 0.05$, drift $\mu = 0.1$, volatility $\sigma = 0.2$ and time horizon $T = 1$.

5.9 Model Risk of Portfolio Performance under loss constraint

Funding constraint is of particular interest in the investment industry. The borrower is usually required to meet some minimum margin. Once the overall portfolio value falls below a certain level during the investment horizon, the borrower would have to liquidate all the position which realizes the loss. Such loss constraint introduces some path-dependent ingredient to the investment performance, adding its sensitivity to the uncertainty of models applied. Here we studied a simple case in which the reference model states that the portfolio performance (without the loss constraint) follows an Ito

process

$$dX'_t = \mu_t X'_t dt + \sigma_t X'_t dW_t$$

The logarithmic return follows the SDE shown in the last section

$$dX_t := d \ln \left(\frac{X'_t}{X'_0} \right) = \left(\mu_t - \frac{\sigma_t^2}{2} \right) dt + \sigma_t dW_t$$

We impose the loss constraint such that the loss gets realized once X_t drops below the loss limit $-l$. Ignoring time value issues, we may write the final return at time T by

$$(X_t + l) \mathbf{1}_{\min\{X_t\}_{t \in [0, T]} > -l} - l$$

where l is the percentage loss limit. The loss function is defined by the negative of the logarithmic return,

$$\ell(T, \cdot) := l - (X_t + l) \mathbf{1}_{\min\{X_t\}_{t \in [0, T]} > -l}$$

We may solve the model risk problem numerically using the finite difference approach.

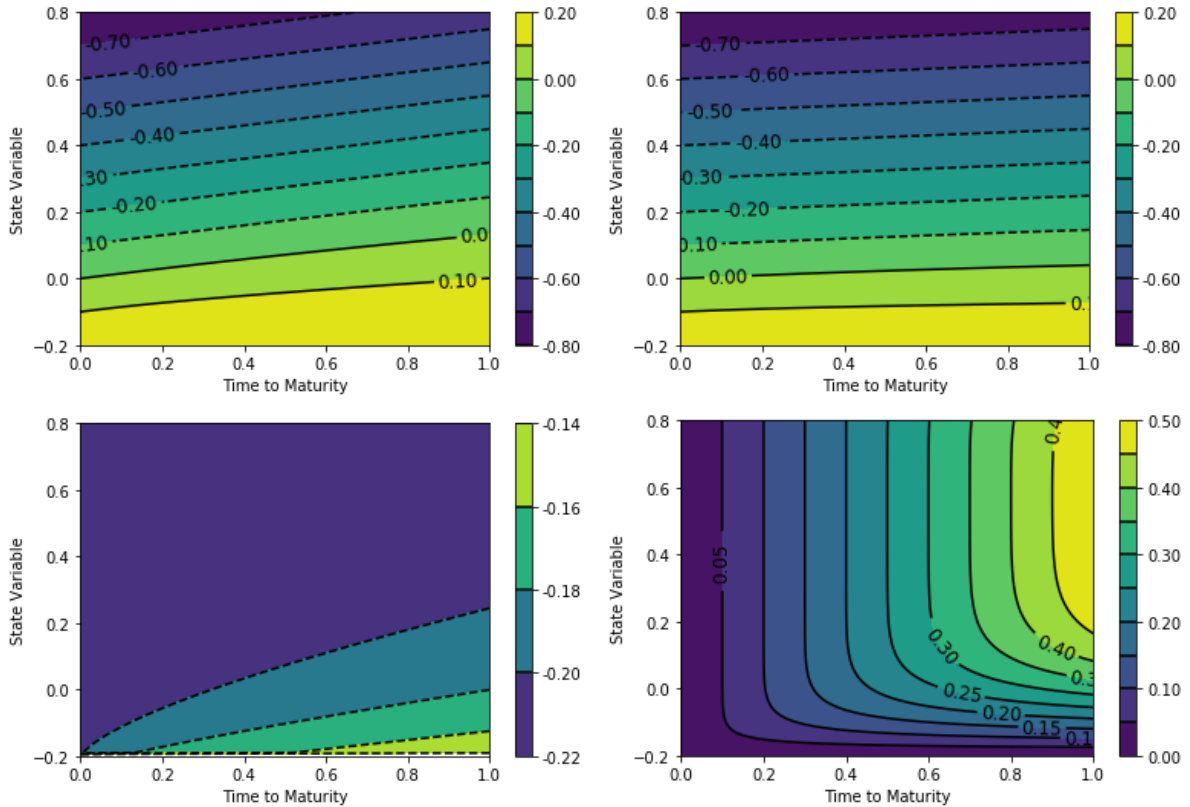


Fig 5.9.1: Two-dimensional grids of (Upper Left) the worst-case risk $\tilde{v}(t, x)$, (Upper Right) the penalized worst-case risk $\tilde{u}(t, x)$, (Bottom Left) the drift term for the change of measure, (Bottom Right) the relative entropy. $\vartheta = 5$, $\mu_t = 0.05$, $\sigma_t = 0.2$ and $l = 0.2$.

Fig. 5.9.2 shows that at a given η level the worst-case loss increases with the loss limit at a decreasing rate. A tighter loss constraint generally limits the amount of loss under the worst-case model. With a tight constraint, model uncertainty has less effect as evidenced by the increase of the relative entropy. When the constraint gets relaxed, the worst-case loss increases quickly with the relative entropy budget.

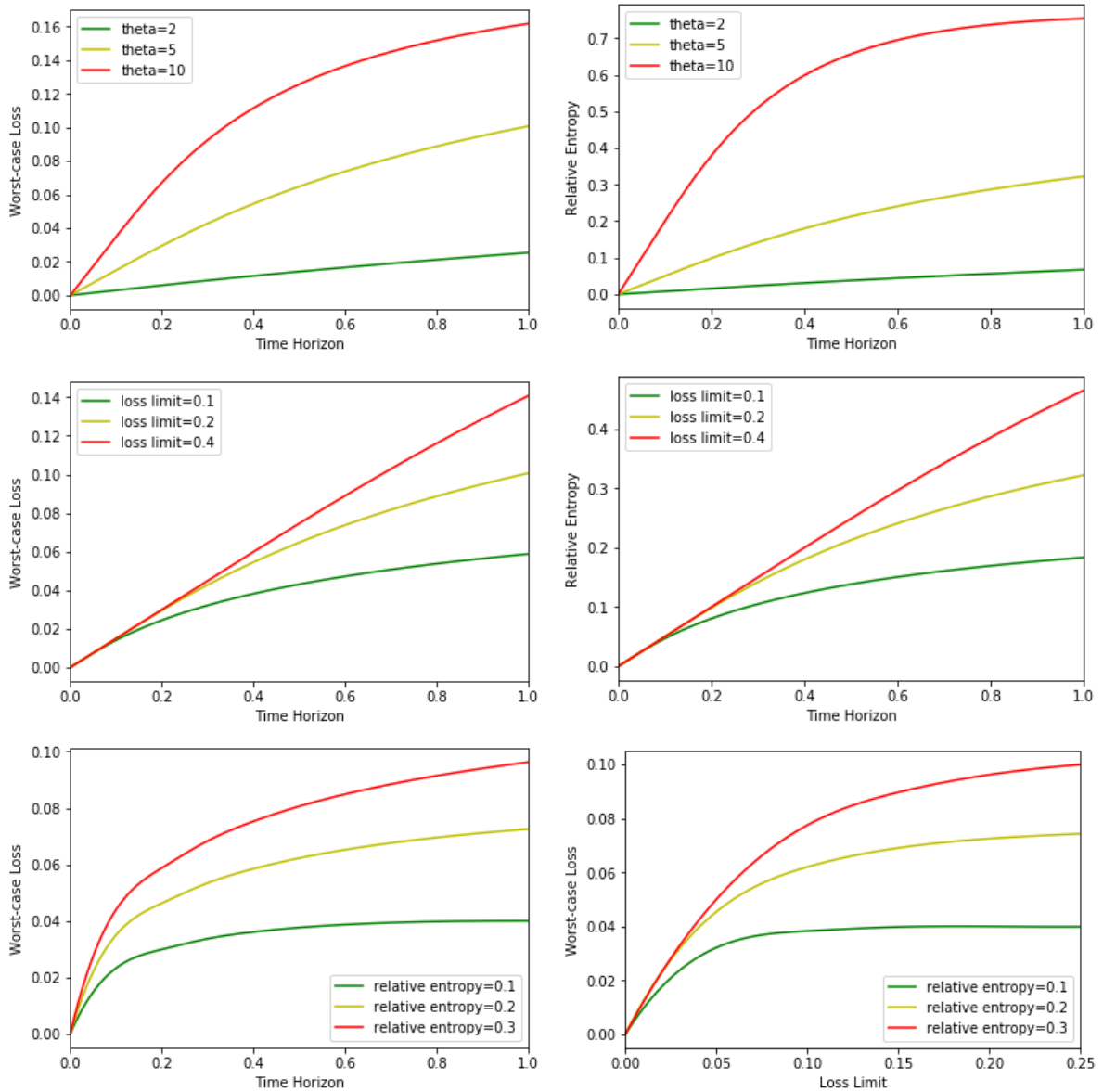


Fig 5.9.2: (Upper Left) Worst-case loss (measured by negative logarithmic return) and (Upper Right) relative entropy at three ϑ levels. (Middle Left) Worst-case loss and (Upper Right) relative entropy under three loss constraints. (Bottom Left) Worst-case loss changes with the time horizon T at three η levels. (Bottom Right) Worst-case loss changes with the loss constraint l at three η levels. The parameter values are $\vartheta = 5$, $\mu_t = 0.05$, $\sigma_t = 0.2$ and $l = 0.2$ unless specified in the figures.

5.10 Concluding Remarks

This paper proposes three approaches to model risk quantification. The mathematical generality provided in Sec. 4 calls for the application in various areas ranging from robust derivative pricing to portfolio management. In particular, it manages to measure the model risk for path-dependent problems in a tractable way.

To formulate a practical problem, we need several ingredients, including terminal time T , a loss function ℓ , a reference model (i.e. a canonical process $(X_t)_{t \in [0, T]}$ under a reference measure \mathbb{P}) and some f -divergence. We rely on the relative entropy to constrain the set of proper alternative models. Other f -divergence may as well be applicable, especially in cases of heavy tails.

The output of the three approaches involves three processes that characterize the penalized risk, the worst-case risk and the relative entropy budget. The relative entropy budget can be estimated on time series of financial market data, allowing disparate sources of model risk to be evaluated in a unified fashion, in a similar manner as developed for static, path-independent case in Sec. 4. Thus, the results presented here represent useful tools to quantify risks due to model ambiguity.

Chapter 6

Model Risk Measurement Using Wasserstein Metric

6.1 Introduction

This chapter presents a new approach to model risk measurement based on the Wasserstein distance between two probability measures. It provides practically feasible results that overcome the restriction of considering only models implying probability measures equivalent to the reference model. The Wasserstein approach is suitable for all types of model risk problems, ranging from the single-asset hedging risk problems to the multi-asset allocation problems. The originality of this approach is also reflected by its capability of accounting for the correlation risk, which results in a sub-linear capital allocation line that is not achievable with other non-parametric approaches.

Most current work on robust risk management either focuses on parameter uncertainty or relies on comparison between models. To go beyond that, Glasserman and Xu (2014) proposed a non-parametric approach. Under this framework, a worst-case model is found among all alternative models in a neighborhood of the reference model. Glasserman and Xu used the Kullback-Leibler divergence (i.e. relative entropy) to measure the distance between an alternative model and the reference model. They also proposed the use of α -divergence to avoid heavy tails that cause integrability issues under Kullback-Leibler divergence.

Both Kullback-Leibler divergence and α -divergence are special examples of the f -divergence (Ali and Silvey 1966, Csiszár 1967, Ahmadi-Javid 2012). A big problem of using f -divergence is that it is well-defined only when the alternative measure is absolutely continuous with respect to the reference measure. This limits the range of the alternative models under consideration. In some cases, we may want to search over all possible probability measures, whether they are absolutely continuous or not. This is especially true when we apply this approach to volatility, which corresponds to the quadratic variation of a process. If the process is driven by a Brownian motion, then searching over absolutely continuous measures rules out any model risk with respect to volatility. In Fig. 6.1.1(a), the distribution of the volatility is a Dirac-delta function

under the reference model. The worst-case scenario that accounts for the volatility risk has a widely spread distribution of volatility. However, f -divergence is not well-defined in this case, and therefore the worst-case scenario simply gets ignored.

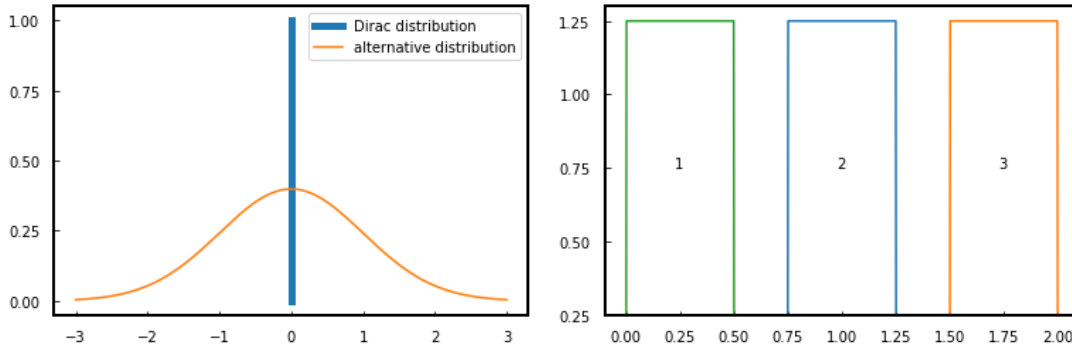


Fig 6.1.1: (a) Dirac measure has a support of a single point. An alternative model with a widespread distribution cannot be related to the reference model using f -divergence. (b) State transition in a metric space. Since f -divergence does not involve the metric, Approaches to model risk measurement based on f -divergence would treat the transitions from State 1 to 2 and 1 to 3 as if they have the cost.

Furthermore, the state space considered by financial practitioners is usually equipped with a natural metric. For instance, the price of a security takes values from the set of positive real numbers, and thus naturally inherits the Euclidean metric. Assuming a diffusion process, the price of the security moves along a continuous path. This means that a large price change is less probable than a small price change, implying a stronger deviation from the reference model. However, the distance of the move, measured by the natural metric, is not explicitly taken into account when using f -divergence, as the models have disjoint support. Fig. 6.1.1(b) shows three models corresponding to three distributions of the security price. Assuming Model 1 is adopted as the reference model, then Model 2 as an alternative model is apparently more probable than Model 3. However, one cannot tell the difference using any type of f -divergence, as the models have disjoint support.

As an attempt to solve these issues, we suggest adopting the Wasserstein metric to measure the distance between probability measures. Relying on the metric of the state space, the Wasserstein metric works for any two measures, even if their supports are disjoint. As a result, the proposed Wasserstein approach accounts for all alternative measures instead of merely the absolutely continuous ones. These features allow us to resolve the two issues of the f -divergence as mentioned above. For financial practitioners, the proposed approach is especially useful when dealing with reference measures with a subspace support (such as a Dirac measure).

This chapter is organized in the following manner. Sec. 6.2 offers a conceptual introduction, including the intuitive motivation and the basics of the Wasserstein metric and its associated transportation theory. Sec. 6.3.1 is the theoretical part that provides

the problem formulation and main results. It also includes practical considerations and comparison between different approaches. Sec. 6.4 gives a few interesting applications in mathematical finance, ranging from the volatility risk in option pricing and hedging to robust portfolio optimisation.

6.2 Basic Concepts

To illustrate the idea of model risk in an intuitive way, we start from a simple discrete state space. An example is credit rating, which is ordinal, e.g. A+, A, A-, BBB+, etc. Suppose we have a reference model that states that in a month the credit rating of an institution could be A+, A- or BBB+. The reference model assigns probabilities of 25%, 50% and 25% to the three states. Since we do not possess complete information, model risk exists either because the actual probabilities of the three states are different or because other ratings are possible. Glasserman and Xu (2014) proposed the so-called “adversary” interpretation, which suggests a fictitious adversary that perturbs the probabilities against us. By perturbing the probabilities, the adversary adds new information, limited by its information entropy budget. For example, if the adversary would like to move a 5% chance from A+ to BBB+, its consumption of relative entropy would be

$$0.2 \ln \left(\frac{0.2}{0.25} \right) + 0.3 \ln \left(\frac{0.3}{0.25} \right) = 0.01 \quad (6.2.1)$$

Now suppose the adversary would like to move the 5% chance to BBB, which is a state of zero probability under the reference measure. The consumption of relative entropy

$$0.2 \ln \left(\frac{0.2}{0.25} \right) + 0.05 \ln \left(\frac{0.05}{0} \right) \quad (6.2.2)$$

becomes infinite. This simply means that such a perturbation is impossible, no matter how much control the adversary has. In the language of probability theory, relative entropy is well-defined only when the new measure is absolutely continuous with respect to the nominal one.

To allow for a more generic quantification of model risk, we may re-define the requested cost of perturbation. Instead of using the relative entropy, we consider the cost of a state transition (termed as the transportation cost). This transportation cost is usually given by some metric on the state space. For simplicity we assume that the distance between two credit ratings is given by the number of ratings in between, e.g. $d(A+, A-) = 2$ and $d(A+, BBB+) = 3$. We calculate the weighted average transportation costs for the two types of perturbations discussed in the last paragraph:

1. shift 5% chance from A+ to BBB+: transportation cost = $5\% \times 3 = 0.15$
2. shift 5% chance from A+ to BBB: transportation cost = $5\% \times 4 = 0.2$

The second-type perturbation only involves a cost slightly larger than the first type, instead of being infinite.

Using the transportation cost described above, one can measure the adversary's cost for all alternative measures, rather than merely the absolutely continuous ones. It may provide state transitions that are highly concentrated. To illustrate this point, think about the transition from state $A+$. The fictitious adversary would push the rating only in one direction. This implies that the transportation performed by the fictitious agent can be represented by a (deterministic) map on the state space $T : \Omega \rightarrow \Omega$. T is called a **transportation map** (Villani 2008). In fact, suppose it is optimal for the fictitious agent (thus the worst case scenario) to transit the state $A+$ to a state, say $BBB+$. There is no motivation for the agent to transport any probability mass from $A+$ to other states. This results from the linearity of the transportation cost, and will be illustrated further in Sec. 6.3.1.

Glasserman and Xu's interpretation of model risk involves an fictitious adversary, but without explicit consideration of its economic nature. They assume that the adversary has a consensus target which is to maximise our expected loss. In reality, such an adversary can only be achieved by a single agent or institution. The actual market structure, however, is usually more competitive. In economic terms, the fictitious adversary may consist of heterogeneous agents who act independently. This asks for approaches that quantify the model risk based on the actual market structure.

Now get back to the credit rating example. In reality there might be multiple agents that are capable of impacting the rating, among which some prefer to upgrade the rating while others prefer to downgrade the rating. This asks for a different formulation of state transitions, for the final state transited from a given initial state becomes a random variable. All we know is a probability measure conditional to the given initial state (or a transition density). Overall, the transportation is described by a joint probability density $\gamma : \Omega \times \Omega \rightarrow \mathbb{R}^+$ instead of a deterministic map. The joint density (or the corresponding measure on $\Omega \times \Omega$) is referred as the **transportation plan** (Villani 2008). This allows us to formulate the optimisation problem w.r.t the transportation plan instead of the transportation map. Such a formulation leads to more general results capable of accounting for different types of market structure.

From a practical perspective, the main advantage of using the Wasserstein metric is to deal with reference measures supported by strict subspaces. Still in the example of credit rating, the reference measure is supported by $\{A+, A-, BBB+\}$, which is a strict subspace of the entire state space (of rating). Approaches based on f -divergence are only capable of incorporating alternative measures with the same support. Using the Wasserstein approach, on the other hand, does allow us to alter the support. In particular, if we formulate the problem using a transportation map T , then the new support is $\{T(A+), T(A-), T(BBB+)\}$, still a strict subspace. Therefore, although different transportation maps provide us with different supports, none of them is capable of spreading to the entire state space. On the other hand, by formulating the problem with a transportation plan, we indeed account for alternative measures that are supported by the entire space. It is reasonable to believe that the distribution is widely spread.

Thus we are interested in an approach to model risk measurement that formulates

the transportation cost based on a transportation plan. In the remaining part, we will apply the Wasserstein metric and its associated transportation theory to the problem of model risk quantification. We will always assume a continuous-state space unless otherwise stated.

6.3 Theory

6.3.1 Wasserstein Formulation of the Model Risk Problem

The core part of model risk measurement is to determine the alternative model under the worst-case scenario. In the language of probability theory, we need to determine the alternative probability measure that maximizes our expected loss. We may formulate the problem in the following way. Given a nominal probability measure P on the state space Ω , we would like to find a worst-case measure Q^* that realizes the following supremum:

$$\begin{aligned} \sup_Q E^Q[V(X)] & \tag{6.3.1} \\ \text{s.t. } D(P||Q) & \leq \eta \end{aligned}$$

The expectation is taken under the alternative measure Q , on a loss function $V : \Omega \rightarrow \mathbb{R}$. Only alternative measures that are close enough to the reference measure are deemed as legitimate. This restriction is formulated by constraining the statistical distance $D(P||Q)$ to be equal to or less than a constant η .

Glasserman and Xu suggest using the relative entropy (or Kullback-Leibler divergence) for $D(P||Q)$. Like any f -divergence, relative entropy has limited feasibility as only equivalent measures are legitimate. Based on the discussion in the last section, we suggest to apply the Wasserstein metric instead. The actual formulation of the model risk problem, on the other hand, has a slightly different form than Eq. 6.3.1. Specifically, instead of optimizing the expectation w.r.t the alternative measure Q (or its density function $q : \Omega \rightarrow \mathbb{R}^+$), we optimize the expectation w.r.t the transportation plan $\gamma : \Omega \times \Omega \rightarrow \mathbb{R}^+$ directly (Villani 2008). The single constraint on q is replaced by two constraints applied to γ , including the marginalisation condition given in Eq. 2.5.44. This formulation is based on the idea of state transition and is illustrated below.

Based on the discussion in the last section, for any pair of states $x, y \in \Omega$ all we need to find is the transition density from x to y , $p_{Y|X}(y|x)$. Given a function of transportation cost from x to y , $c(x, y)$, the expected transportation cost conditional to an initial state x is

$$W(x) = \int_{\Omega} p_{Y|X}(y|x)c(x, y)dy \tag{6.3.2}$$

The initial state x follows a distribution $p_X(x)$ given by the reference model. Take expectation under the reference measure, we get the unconditional transportation cost

$$W = \int_{\Omega} p_X(x)W(x)dx = \int_{\Omega \times \Omega} p_{X,Y}(x, y)c(x, y)dxdy \tag{6.3.3}$$

where the joint distribution $p_{X,Y}(x, y) = p_X(x)p_{Y|X}(y|x)$. To be consistent with the notation used previously, we denote the marginal distributions p_X, p_Y by p, q , and the joint distribution $p_{X,Y}$ by the transportation plan γ . It is noted that the transition converts the initial distribution $p(x)$ to a final distribution $q(y)$, inducing a change of measure on the state space Ω .

One of the key tasks of the model risk measurement is to solve for the worst-case model under certain constraints. These constraints set the criteria for legitimate alternative models. Now denote the loss function by $V(x)$ ($x \in \Omega$), the probability density function of the reference model by $p(x)$, and the probability density function of an alternative model by $q(x)$. We formulate the problem by the supremum of the expected loss over all legitimate models:

$$\sup_q \int_{\Omega} q(y)V(y)dy \tag{6.3.4}$$

According to the discussion in the last section, we regard the change of measure as probabilistic state transitions. The probability density function $q(y)$ of the alternative model is merely the marginalisation of a joint density (or transportation plan) $\gamma(x, y)$, i.e. $q(y) = \int_{\Omega} \gamma(x, y)dx$. This allows us to take the supremum over $\gamma(x, y)$ instead of $q(y)$:

$$\sup_{\gamma} \int_{\Omega \times \Omega} \gamma(x, y)V(y)dx dy \tag{6.3.5}$$

The first constraint of the supremum problem comes from the marginalisation of the joint density w.r.t x , as it is given by the reference model:

$$\int_{\Omega} \gamma(x, y)dy = p(x) \tag{6.3.6}$$

In a similar way to Glasserman and Xu’s work, we restrict all alternative measures by their distances from the reference model. The distance is now measured by the average transportation cost given in Eq. 6.3.3. It reflects the expected cost paid by a fictitious adversary who attempts to transit a state x to an alternative state y according to the transportation plan $\gamma(x, y)$. This results in the following constraint which defines the set of legitimate measures:

$$\int_{\Omega \times \Omega} \gamma(x, y)c(x, y)dx dy \leq \eta \tag{6.3.7}$$

The constant η in Eq. 6.3.7 is termed as the Wasserstein distance budget, just as the relative entropy budget in Glasserman and Xu’s approach. In order to account for a specific density function $q^*(y)$ in the constrained supremum problem given by Eq. 6.3.4-6.3.7, the Wasserstein distance, defined in Eq. 2.5.44, between $q^*(y)$ and the nominal density $p(x)$ cannot exceed η . In fact, if $q^*(y)$ can be obtained by marginalizing

a transportation cost $\gamma^*(x, y)$ that satisfies Eq. 6.3.6-6.3.7, then according to Eq. 2.5.44 its Wasserstein distance with the nominal density function $p(x)$ is

$$\begin{aligned} W(p, q^*) &= \inf_{\gamma} \int_{\Omega \times \Omega} \gamma(x, y) c(x, y) dx dy \\ &\leq \int_{\Omega \times \Omega} \gamma^*(x, y) c(x, y) dx dy \leq \eta \end{aligned} \quad (6.3.8)$$

On the other hand, if $W(p, q^*) < \eta$, then the density function $q^*(y)$ can always be expressed by the marginalisation of a transportation plan $\gamma^*(x, y)$ that satisfies Eq. 6.3.6-6.3.7. Otherwise, in the definition of the Wasserstein distance, Eq. 2.5.44, η sets a lower bound for the term

$$\int_{\Omega \times \Omega} \gamma(x, y) c(x, y) dx dy \quad (6.3.9)$$

Therefore the Wasserstein distance, as the infimum of the term above, is equal to or larger than η . This immediately violates the assumption $W(p, q^*) < \eta$. In summary, η sets the maximum level (budget) of Wasserstein distance for an alternative measure to be legitimate.

Interestingly, even though the problem (Eq. 6.3.5-6.3.7) is formulated using the transportation plan, its solution can be expressed by a transportation map $T^* : \Omega \rightarrow \Omega$,

$$T^*(x) = \arg \max_{y \in \Omega} \left[V(y) - \frac{c(x, y)}{\beta} \right] \quad (6.3.10)$$

where $\beta \in \mathbb{R}_{++}$ is a constant. The underlying reason is the linearity of Eq. 6.3.9 w.r.t the transportation plan γ . Suppose the worst case scenario is to transit a state x to another state $T^*(x)$. Then there is no motivation for the fictitious adversary to transit x to states other than $T^*(x)$, say $T'(x)$, for the adversary could continue improving the target by increasing $\gamma(x, T^*(x))$ while reducing $\gamma(x, T'(x))$ (by the same amount). See Appendix A for a sketch of the derivation of Eq. 6.3.10.

6.3.2 Entropy Constraint on Transportation Plan

Eq. 6.3.10 provides the worst-case transportation map for the problem formulated in Eq. 6.3.5-6.3.7. In practice, it is also advantageous to have a widely distributed transition density. For the purpose of risk management, we need to consider a wide range of alternative measures due to model ambiguity. As a result a wide distribution is usually more representative than a narrow distribution. From the information-theoretic point of view, a widespread distribution contains less information (more entropy) thus more appropriately representing the model ambiguity. Practical situations where the approaches based on f -divergence are not applicable usually have reference measures that are too restrictive in the sense that they are supported merely by subspaces (of the state space). To correctly quantify the model risk one should consider wide distributions supported by the entire state space. However, these distributions do not have

well-defined f -divergence w.r.t the reference measure, providing an inherent issue of these approaches.

One of the primary purposes of using Wasserstein metric instead of f -divergence is to tackle this issue. Specifically, we would like to include all measures regardless of their support. This purpose is achieved by using the transportation plan formulation as illustrated in Sec. 2.5.6. However, without further constraint the worst-case model can still be achieved with a transportation map, as illustrated by Eq. 6.3.10. This causes the worst-case measure to be restrictive if the reference measure is supported merely by a subspace. To achieve a widespread worst-case distribution, one may need to impose further constraints to Eq. 6.3.5-6.3.7.

A Dirac reference measure, denoted by P , provides a special example where Eq. 6.3.10 is not suitable for characterizing the worst-case scenario. Applying the transportation map T^* results in the worst-case measure supported by $\{T(x)\}$ where x is the sole element in $\text{supp}(P)$. The worst-case measure is Dirac as well. In most cases, this worst-case measure inappropriately accounts for model ambiguity. To resolve this issue, we may further impose an entropy constraint that guarantees the worst-case measure to be supported by the entire state space:

$$-\int_{\Omega \times \Omega} \gamma(x, y) \ln \gamma(x, y) dx dy \geq \mu \quad (6.3.11)$$

The LHS is the (differential) entropy (Cover and Thomas 2012) of the joint distribution (transportation plan) $\gamma(x, y)$, and the RHS is a constant $\mu \in \mathbb{R}$ (or a positive constant $\mu \in \mathbb{R}_{++}$ for discrete-state space). This constraint excludes every transportation plan that can be expressed by a transportation map. In fact, every transportation map T gives a transportation plan with a δ -function transition density (see Eq. 2.5.45). For such a transportation plan, the δ -function makes the LHS of Eq. 6.3.11 approach negative infinity (or zero for discrete-state space), and is therefore excluded.

Alternatively, Eq. 6.3.11 can be interpreted with respect to the transition density function $p_{Y|X}(y|x)$. We may rewrite Eq. 6.3.11 by

$$-\int_{\Omega \times \Omega} \gamma(x, y) \ln p_{Y|X}(y|x) dx dy \geq \mu + \int_{\Omega} p(x) \ln p(x) dx \quad (6.3.12)$$

Eq. 6.3.12 imposes a restriction on the transition density function. A tighter restriction (with a larger μ) implies a wider transition density. On the other hand, if we relax the constraint completely by shifting μ towards negative infinity (or zero for discrete-state space), then we permit transition densities to take the form of δ -functions.

We may further introduce terms from information theory, and rewrite Eq. 6.3.12 by

$$\int_{\Omega} p(x) H(Y|X = x) dx \geq \mu - H(X) \quad (6.3.13)$$

where $H(X)$ denotes the entropy of the random variable X (Cover and Thomas 2012). Since its distribution $p(x)$ is given by the reference model, $H(X)$ is deemed as a constant. $H(Y|X = x)$, on the other hand, is the information entropy w.r.t the transition

density $p_{Y|X}(y|x)$. It is interpreted as the entropy of the random variable Y , conditional to X taking a given value x . $H(Y|X = x)$ quantifies the uncertainty of the transportation from a given state x . It is noted that in information theory, the LHS of Eq. 6.3.14 is termed as the conditional (differential) entropy and is denoted by $H(Y|X)$ (Cover and Thomas 2012). This leads to an equivalent information-theoretic version of the constraint Eq. 6.3.11:

$$H(Y|X) \geq \mu - H(X) \quad (6.3.14)$$

6.3.3 Main Result and Discussion

The supremum problem Eq. 6.3.5, subject to the three constraints Eq. 6.3.6, 6.3.7 and 6.3.11, formulates the complete version of the Wasserstein approach to model risk measurement. Now suppose there exists a joint distribution $\gamma^*(x, y)$ that solves the problem. Then the worst-case model is characterised by a probability density function

$$q^*(y) = \int_{x \in \Omega} \gamma^*(x, y) dx, \quad \forall y \in \Omega \quad (6.3.15)$$

To solve the constrained supremum problem, we introduce two multipliers $\alpha \in \mathbb{R}_+$ and $\beta \in \mathbb{R}_+$, and transform the original problem to a dual problem. Solving the inner part of the dual problem leads to our main result (see Appendix B for derivation):

$$q^*(y) = \int_{\Omega} dx \frac{p(x) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right)}{\int_{\Omega} \exp\left(\frac{V(z)}{\alpha} - \frac{c(x,z)}{\alpha\beta}\right) dz} \quad (6.3.16)$$

It is noted that the multipliers α and β are in fact controlling variables that determine the levels of restriction of the entropy constraint Eq. 6.3.11 and the transportation constraint Eq. 6.3.7, respectively.

The limit when α approaches zero corresponds to complete relaxation of the entropy constraint Eq. 6.3.11. In this limit Eq. 6.3.15 degenerates to the probability density function induced by the transportation map given by Eq. 6.3.10. On the other side of the spectrum, Eq. 6.3.15 approaches a uniform distribution when α approaches infinity, as a result of the tight entropy constraint.

In the extreme case of $\beta = 0$, Eq. 6.3.15 leads to a simple result $q^*(x) = p(x)$. This is because the transportation constraint Eq. 6.3.7 reaches its tightest limit ($\eta = 0$). No state transition is allowed, thus preserving the reference model. On the other hand, when β approaches infinity, the worst-case distribution $q^*(y) \sim \exp(V(y)/\alpha)$ is exponentially distributed. In this case, the transportation cost is essentially zero. As a result, the worst-case measure is the one that maximises the expected value of $V(Y)$ with a reasonably large entropy (the maximum expected value is given by a Dirac measure at $\arg \max_y V(y)$ but this results in a very low entropy). Special cases of Eq. 6.3.16 are tabulated in Tab. 6.3.1 for different values of α and β .

Table 6.3.1: Worst-case probability density function at different (α, β) combinations. p is the nominal distribution and u is the uniform distribution. δ denotes the Dirac δ -function and T^* is the transportation map given by Eq. 6.3.10.

	$\alpha = 0$	α	$\alpha \rightarrow \infty$
$\beta = 0$	$p(x)$		
β	$p(T^{*-1}(x))/ J_T $	given by Eq. 6.3.15	$\rightarrow u(x)$
$\beta \rightarrow \infty$	$\delta(x - \arg \max V(x))$	$\propto e^{V(x)/\alpha}$	

6.3.4 Practical Considerations

According to Table. 6.3.1, the worst-case measure approaches a uniform distribution when α approaches infinity (i.e. under the most restrictive entropy constraint). In practice, we may want the worst-case distribution to converge to a given density function q_0 instead of being uniform. This requires modification on the formulation of the problem, by generalising the entropy constraint Eq. 6.3.14 to

$$-D_{KL}(P(Y|X)||Q_0(Y)) \geq \mu - H(X) - H(Y) \quad (6.3.17)$$

$D_{KL}(P(Y|X)||Q_0(Y))$ denotes the conditional relative entropy, given by the expected value of the KL divergence, $D_{KL}(P(Y|X = x)||Q_0(Y))$, of the two probability density functions w.r.t y , $p_{Y|X}(\cdot|x)$ and $q_0(\cdot)$. Written explicitly, the conditional relative entropy takes the form

$$\begin{aligned} D_{KL}(P(Y|X)||Q_0(Y)) &= \int_{\Omega} p(x) \left(\int_{\Omega} p_{Y|X}(y|x) \ln \left(\frac{p_{Y|X}(y|x)}{q_0(y)} \right) dy \right) dx \\ &= \int_{\Omega \times \Omega} \gamma(x, y) \ln \frac{\gamma(x, y)}{q_0(y)} dx dy - \int_{\Omega} p(x) \ln p(x) dx \end{aligned} \quad (6.3.18)$$

Substituting Eq. 6.3.18 into Eq. 6.3.17 allows us to obtain an explicit version of the constraint:

$$- \int_{\Omega \times \Omega} \gamma(x, y) \ln \frac{\gamma(x, y)}{q_0(y)} dx dy - \int_{\Omega} q_0(y) \ln q_0(y) dy \geq \mu \quad (6.3.19)$$

It is clear that the previous entropy constraint Eq. 6.3.11 is merely a special case of Eq. 6.3.19 in which q_0 is a uniform distribution. Under this formulation, the problem that we need to solve consists of Eq. 6.3.5, 6.3.6, 6.3.7 and 6.3.19. The result differs from Eq. 6.3.16 by a weighting function q_0 (see Appendix B for derivation):

$$q^*(y) = \int_{\Omega} dx \frac{p(x) q_0(y) \exp \left(\frac{V(y)}{\alpha} - \frac{c(x, y)}{\alpha \beta} \right)}{\int_{\Omega} q_0(z) \exp \left(\frac{V(z)}{\alpha} - \frac{c(x, z)}{\alpha \beta} \right) dz} \quad (6.3.20)$$

It is noted that Eq. 6.3.20 takes a similar form to Bayes' theorem and q_0 serves as the prior distribution. In fact, if the conditional distribution takes the following form:

$$p_{X|Y}^*(x|y) \propto \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right) \quad (6.3.21)$$

Then Bayes' theorem states that

$$\begin{aligned} p_{Y|X}^*(y|x) &= \frac{p_{X|Y}^*(x|y)q_0(y)}{\mathbf{E}_Y\left(p_{X|Y}^*(x|\cdot)q_0(\cdot)\right)} \\ &= \frac{q_0(y) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right)}{\int_{\Omega} q_0(z) \exp\left(\frac{V(z)}{\alpha} - \frac{c(x,z)}{\alpha\beta}\right) dz} \end{aligned} \quad (6.3.22)$$

which is the posterior distribution of Y given the observation $X = x$. Now if we observe a distribution $p(x)$ over X , then we may infer the distribution of Y to be

$$\begin{aligned} q^*(y) &= \int_{\Omega} p(x)p_{Y|X}^*(y|x)dx \\ &= \int_{\Omega} \frac{p(x)q_0(y) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right)}{\int_{\Omega} q_0(z) \exp\left(\frac{V(z)}{\alpha} - \frac{c(x,z)}{\alpha\beta}\right) dz} dx \end{aligned} \quad (6.3.23)$$

which is exactly the worst-case distribution given in Eq. 6.3.20.

The connection between Bayes' theorem and Eq. 6.3.20 is not just a coincidence. In fact, the worst-case distribution of Y , given in Eq. 6.3.23, can be regarded as the posterior distribution of a latent variable. On the other hand, the reference model of X , given by $p(x)$, is considered as the distribution that is actually observed. Assuming no reference model exists (i.e. no observation on X has been made), then our best guess on the latent variable Y is given solely by its prior distribution $q_0(y)$. Now if the observable variable X does take a particular value x , then we need to update our estimation according to the Bayes' theorem (Eq. 6.3.22). The conditional probability density $p_{X|Y}^*(x|y)$ takes the form of Eq. 6.3.21, reflecting the fact that the observable variable X and the latent variable Y are not far apart. Imagining that we generate a sampling set $\{x_i\}$ following the nominal distribution $p(x)$, then for each x_i we get a posterior distribution $p_{Y|X}^*(y|x_i)$ from Eq. 6.3.22. Overall, the best estimation of the distribution over the latent variable Y results from the aggregation of these posterior distributions. This is achieved by averaging them weighted by their probabilities $p(x_i)$, as given in Eq. 6.3.23. This leads to the Bayesian interpretation of the model risk measurement, which concludes that by "observing" the reference model $p(x)$ over the observable variable X , the worst-case model is given by updating the distribution of the latent variable Y , from the prior distribution $q_0(y)$ to the posterior distribution $q^*(y)$.

If we know nothing about the reference model, setting the prior q_0 to a uniform distribution seems to make the most sense (because a uniform distribution maximizes

Table 6.3.2: Worst-case density function with prior q_0 at different (α, β) combinations. p is the nominal distribution. δ denotes the Dirac δ -function and T^* is the transportation map given by Eq. 6.3.10.

	$\alpha = 0$	α	$\alpha \rightarrow \infty$
$\beta = 0$	$p(x)$		
β	$p(T^{*-1}(x))/ J_T $	given by Eq. 6.3.20	$\rightarrow q_0(x)$
$\beta \rightarrow \infty$	$\delta(x - \arg \max V(x))$	$\propto q_0(x)e^{V(x)/\alpha}$	

the entropy thus containing least information). This leads to the main result given by Eq. 6.3.16. However, it is sometimes much more convenient to choose a prior other than the uniform distribution. A particularly interesting case is to set q_0 the same as the nominal distribution p . In this case, the limit as $\beta \rightarrow \infty$ (complete relaxation of the transportation constraint) is given by

$$q^*(x) = \frac{p(x)e^{\theta V(x)}}{\int_{\Omega} p(x)e^{\theta V(x)} dx} \tag{6.3.24}$$

where we replace the parameter α^{-1} by θ . This limit is exactly the worst-case distribution given by the relative entropy approach (Glasserman and Xu 2014). Despite the simplicity of Eq. 6.3.24, it is not recommended to set $q_0 = p$, because by doing so we lose the capability of altering the support of the reference measure.

In practice, a common problem of the relative entropy approach is that the denominator in Eq. 6.3.16 may not be integrable. To see this point, we examine the worst-case density function under the relative entropy approach:

$$q_{KL}^*(x) \propto p(x)e^{\theta V(x)} \tag{6.3.25}$$

The RHS of Eq. 6.3.25 may not be integrable if $V(x)$ increases too fast (or $p(x)$ decays too slowly, as in the case of heavy tails). As an example, we consider the worst-case variance problem, where $V(x) = x^2$. If the reference model follows an exponential distribution, then Eq. 6.3.25 is not integrable.

Using the proposed Wasserstein approach, however, the flexibility of choosing a proper prior q_0 helps us bypass this issue. In fact, one may choose a prior distribution q_0 , different from the nominal distribution p , to guarantee that it decays sufficiently fast. According to Eq. 6.3.22, all we need to guarantee is that

$$q_0(y) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x, y)}{\alpha\beta}\right) \tag{6.3.26}$$

is integrable w.r.t y . Fortunately, it is always possible to find some q_0 that satisfies this criterion. As a simple choice, we may set $q_0(y) \propto e^{-V(y)/\alpha}$ to ensure integrability. Such a choice makes Eq. 6.3.26 proportional to

$$\exp\left(-\frac{c(x, y)}{\alpha\beta}\right) \tag{6.3.27}$$

We suppose that the state space Ω is a Euclidean space with finite dimension and the transportation cost $c(x, y)$ is given by its Euclidean distance. Then for all $x \in \Omega$, Eq. 6.3.27 is integrable w.r.t y , since the integrand diminishes exponentially as y moves away from x .

In summary, applying the relative entropy constraint Eq. 6.3.19 allows for flexibility of choosing a prior distribution q_0 . This is practically useful as one can avoid integrability issues by selecting a proper prior. This flexibility is not shared by the relative entropy approach in Glasserman and Xu, which is a special case where the prior q_0 equals the nominal distribution p .

6.4 Application

6.4.1 Jump risk under a diffusive reference model

We start from a price process that takes the form of a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{6.4.1}$$

The logarithmic return at time T follows a normal distribution:

$$x := \ln \left(\frac{S_T}{S_0} \right) \sim \mathcal{N} \left(\left(\mu - \frac{\sigma^2}{2} \right) T, \sigma^2 T \right) \tag{6.4.2}$$

In the limit, when the volatility parameter approaches zero, the return becomes deterministic and the distribution density is

$$p(x) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi T} \sigma} e^{-\frac{[x - (\mu - \sigma^2/2)T]^2}{2\sigma^2 T}} = \delta(x - \mu T) \tag{6.4.3}$$

In this case, model risk cannot be quantified using f -divergence. In fact, the reference measure is a Dirac measure, and therefore no equivalent alternative measure exists. Under the KL divergence in particular, the worst-case measure is calculated by

$$\frac{p(x)e^{\theta V(x)}}{\int_{\Omega} p(x)e^{\theta V(x)} dx} = \delta(x - \mu T) \tag{6.4.4}$$

which is the same as the reference measure. This is consistent with the Girsanov theorem for diffusion processes which states that the drift term is altered by some amount proportional to the volatility, i.e. $\tilde{\mu} = \mu - \lambda\sigma$. When the volatility parameter under the reference model decreases to zero, the alternative measure becomes identical to the reference measure.

Approaches based on f -divergence exclude the existence of model risk given a zero volatility. This is, however, not true in practice, as the nominal diffusion process may still “regime-switch” to some discontinuous process. In fact, to quantify risks, one usually takes into account the possibility of discontinuous changes of state variables

(i.e. jumps). Using the Wasserstein approach, quantifying such jump risk becomes possible, even if the reference model is based on a pure diffusion process. Substituting Eq. 6.4.3 into Eq. 6.3.16 gives the worst-case distribution (see Appendix C for details)

$$q_W(x) = \frac{\exp\left(\frac{V(x)}{\alpha} - \frac{c(x, \mu T)}{\alpha\beta}\right)}{\int_{\Omega} \exp\left(\frac{V(y)}{\alpha} - \frac{c(y, \mu T)}{\alpha\beta}\right) dy} \quad (6.4.5)$$

Notice that Eq. 6.4.5 is suitable for any application where the reference model is given by a Dirac measure. Under f -divergence, the limitation to absolutely continuous measures keeps the reference model unchanged. The Wasserstein approach, on the other hand, relaxes this limitation, allowing for a worst-case model that differs from a Dirac measure. This allows us to measure risk in variables assumed to be deterministic in the reference model. A particularly interesting example is the quadratic variation process, which is deemed to be deterministic under the Black-Scholes model. We will discuss this in detail later with regard to model risk in dynamic hedging.

To illustrate Eq. 6.4.5, we consider the expected value of x under the worst-case scenario. This problem is formulated using Eq. 6.3.1 with a linear loss function $V(x) = x$. We further assume a quadratic transportation cost function $c(x, y) = (x - y)^2$. The worst-case distribution given by Eq. 6.4.5 turns out to be

$$q_W(x) = \frac{1}{\sqrt{\pi\alpha\beta}} e^{-\frac{(x - \mu T - \beta/2)^2}{\alpha\beta}} \quad (6.4.6)$$

One can see that the worst-case scenario is associated with a shift of the mean by $-\beta/2$, even if the reference measure is deterministic (i.e. Dirac). The change in mean is also associated with a proportional variance of $\alpha\beta/2$, if α is assigned a positive value. The resulting normal distribution, with a finite variance, is a reflection of model ambiguity. This is in contrast with approaches based on f -divergences, which are incapable of altering the reference model in this case, as its support is a single point.

6.4.2 Volatility Risk and Variance Risk

In this section, we consider the risk of volatility, uncertainty given the nominal Black-Scholes model. When an option approaches maturity, the reference measure (on the price of the underlying asset) approximates a Dirac measure. This is visualised by the normal distribution of the return narrowing at a rate of \sqrt{T} . When the time to maturity T goes to zero, the normal distribution shifts to a Dirac distribution with zero variance.

Under the Kullback-Leibler divergence (or any f -divergence), model risk vanishes when the reference model converges to a Dirac measure. As a result, for a short time to maturity, a sufficient amount of variance uncertainty can only be produced with a large cost (parametrised by θ). To illustrate this point, consider a normal distribution (say Eq. 6.4.3 before taking the limit). For the purpose of measuring the variance risk, we need to adopt a quadratic loss function $V(x) = x^2$. Under the Kullback-Leibler

divergence, the variance of the worst-case distribution is given by (Glasserman and Xu 2014)

$$\sigma_{KL}^2 T = \frac{\sigma^2 T}{1 - 2\theta\sigma^2 T} \tag{6.4.7}$$

As the time to maturity T goes to zero, the worst-case volatility σ_{KL} converges to σ , with a fixed θ . This is not consistent with what we see in the market. In fact, with short time to maturity the fear of jumps can play an important role. Such fear of risks is priced into options and variance swaps, and is called the volatility (or variance) risk premium.

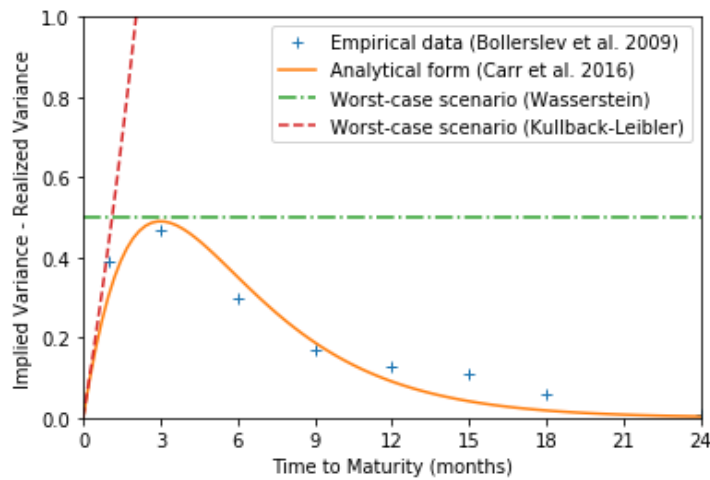


Fig 6.4.1: Worst-case volatility as a function of time under the (a) Wasserstein approach, (b) KL divergence.

The volatility (or variance) risk premium can be considered as compensation paid to option sellers for bearing volatility risk (Bakshi and Kapadia 2003, Low and Zhang 2005). It is practically quantified as the difference between the implied volatility (or variance) and the realised volatility (or variance). As it is priced based on the volatility risk, its quantity is directly linked to the risk associated with the reference measure used to model the underlying asset. Therefore by analyzing the term structure of such premium, one can get some insight into the worst-case volatility risk. Under the assumption of diffusive price dynamics, Carr and Wu (2016) developed a formula for the at-the-money implied variance. Illustrated in Fig. 6.4.1, the formula matches well with the empirical data (Bollerslev, Tauchen and Zhou 2009) for maturities longer than 3 months. For maturities shorter than 3 months, however, the formula seems to underestimate the variance risk premium. Other empirical work also shows that option buyers consistently pay a higher risk premium for shorter maturity options (Low and Zhang 2005).

The underestimation of the volatility risk premium for short maturities is an intrinsic problem with diffusive models. Indeed, the work mentioned above reveals the

importance of quantifying jumps when time to maturity remains short. Other work shows that the risk premium due to jumps is fairly constant across different maturities (Ait-Sahalia, Karaman and Mancini n.d.). This implies a very different time dependency from that due to continuous price moves (Eq. 6.4.7). In fact, any approach based on f -divergence is incapable of producing sufficient model risk for short maturities, suggesting a decaying term structure of the risk premium. On the other hand, the Wasserstein approach does not suffer from this issue. In fact, it produces a worst-case volatility that has little time dependence (Fig. 6.4.1). Therefore, the Wasserstein approach provides a particularly useful tool for managing variance risk and quantifying its risk premium for short times to maturity.

With the Wasserstein approach, the worst-case variance takes the form of (see Appendix C)

$$\sigma_W^2 T = \frac{\sigma^2 T}{(1 - \beta)^2} + \frac{\alpha \beta}{2(1 - \beta)} \quad (6.4.8)$$

This provides a worst-case variance that is independent of time to maturity. It scales the nominal variance by a constant factor $(1 - \beta)^{-2}$. In addition, it introduces a constant additional variance $\alpha \beta / 2(1 - \beta)$. This variance term is modulated by the parameter α . If we set α to zero, then the worst-case volatility σ_W is merely a constant amplification of the nominal volatility σ . This model risk measure, however, may not be sufficient if the nominal volatility is very close to zero. The extra variance term serves to account for additional risks (e.g. jumps) that are not captured by the nominal volatility. In practice, the Lagrange multiplier α and β may be determined by statistical approaches.

6.4.3 Model Risk in Portfolio Variance

The Wasserstein approach can be applied to quantify the risk associated with modelling the variance of a portfolio, assuming the asset returns follow a multivariate normal distribution. Suppose there are n assets under consideration and their returns are reflected by a state vector x : $x \in \mathcal{V}$ where \mathcal{V} is a n -dimensional vector space. For generality, we consider the following target function $V : \mathcal{V} \rightarrow \mathbb{R}_+$

$$V(x) = x^T A x \quad (6.4.9)$$

where A is a positive-definite symmetric matrix. If we replace x by $x' = x - E(x)$ and A by $w w^T$, then the expected value of the target function reflects the portfolio variance:

$$E[V(x)] = E(x^T w w^T x) = w^T \Sigma w \quad (6.4.10)$$

where w is the vector of portfolio weights and Σ is the covariance matrix of the normally distributed asset returns (under the reference model).

To find the worst-case model using the Wasserstein approach, we need to first define a metric in the vector space \mathcal{V} . Suppose the vector space is equipped by a norm

$\|x\|$ then the metric is naturally defined by $c(x, y) = \|x - y\|$. Here we focus on the kind of norm that has an inner-product structure:

$$\|x\| = \sqrt{x^T B x}, \quad \forall x \in \mathcal{V} \quad (6.4.11)$$

where B is a positive-definite symmetric matrix (constant metric tensor). The resulting worst-case distribution is still multivariate normal, with the vector of means and covariance matrix replaced by (see Appendix D for the derivation)

$$\mu_W = (B - \beta A)^{-1} B \mu \quad (6.4.12)$$

$$\Sigma_W = (B - \beta A)^{-1} B \Sigma B (B - \beta A)^{-1} + \frac{\alpha \beta}{2} (B - \beta A)^{-1} \quad (6.4.13)$$

Apart from a constant term that vanishes if $\alpha = 0$, the worst-case distribution is transformed from the nominal distribution via a measure-preserving linear map (see Appendix D). This result is more intuitive than the result obtained using the KL divergence, given by (Glasserman and Xu 2014)

$$\mu_{KL} = (I - 2\theta \Sigma A)^{-1} \mu \quad (6.4.14)$$

$$\Sigma_{KL} = (I - 2\theta \Sigma A)^{-1} \Sigma \quad (6.4.15)$$

Fig. 6.4.2 provides an example illustrating that the worst-case distribution is indeed a measure-preserving transform with the Wasserstein approach. The black arrows illustrate the basis vectors under transformation.

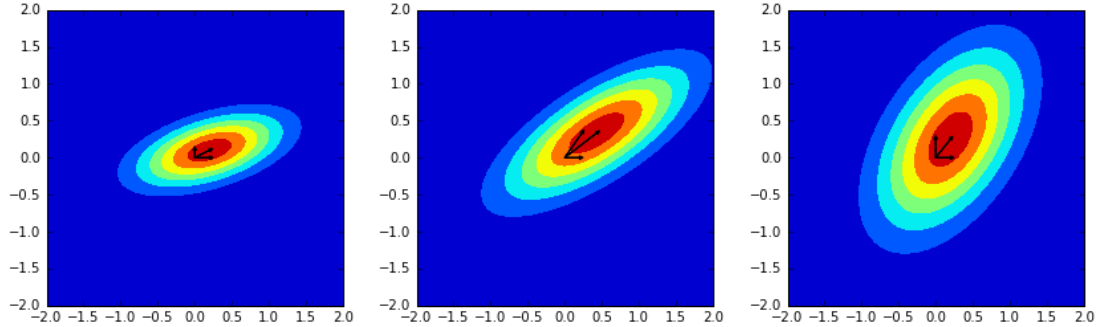


Fig 6.4.2: Multivariate nominal distributions (a) reference model, (b) worst case under the KL divergence, (c) worst case under the Wasserstein approach (as a measure-preserving transform).

The second term in Eq. 6.4.13 reflects residual uncertainty when the reference model has vanishing variances. This term is especially useful when some of the assets are perfectly correlated (either 1 or -1) and the vector space \mathcal{V} is not fully supported by the reference measure. In this case, the Wasserstein approach provides results that differ significantly from the f -divergence approach. In particular, approaches based

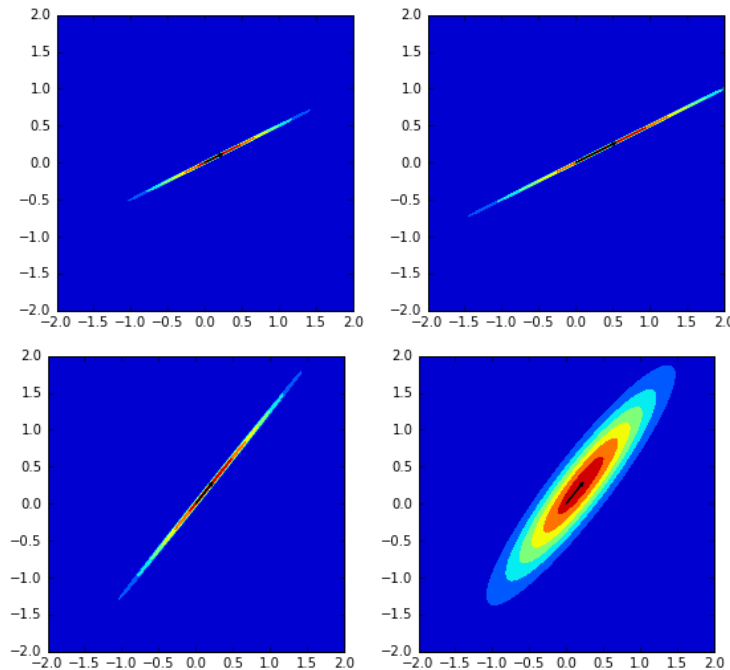


Fig 6.4.3: Multivariate nominal distributions (a) reference model, (b) worst case under the KL divergence, when the support is a low-dimensional subspace. Worst-case multivariate nominal distributions under the Wasserstein approach (c) $\theta = 0$ (d) $\theta = 0.5$.

on KL divergence (or any f -divergence) cannot alter the support; they merely reweight the states within the support. This is illustrated in Fig. 6.4.3, where two assets are perfectly correlated. The reference model shown in (a) provides a measure supported by a one-dimensional vector subspace of \mathcal{V} . The worst-case measure under the KL divergence is supported by the same subspace, as illustrated in (b). This conclusion can actually be derived from the worst-case measure given by Eq. 6.4.15 (see Appendix E for a proof).

On the other hand, the Wasserstein approach is capable of considering measures supported by other vector subspaces. We first ignore the constant variance term by setting α to zero in Eq. 6.4.13. The Wasserstein approach “rotates” the original support by applying linear maps to the reference measure. In the case illustrated by Fig. 6.4.3(c), essentially all measures supported by a one-dimensional vector subspace are within the scope of the approach (see Appendix F for a proof). Among those measures, the Wasserstein approach picks the worst one supported by a vector subspace different from the original one. It essentially searches for the optimal transform over the entire space. In practice, we may want to account for the risk associated with the assumption of perfect correlation. This is accomplished by assigning positive value to α , allowing the distribution to “diffuse” into the entire vector space as illustrated in Fig. 6.4.3(d).

It is worthwhile noting that the Wasserstein approach also has a practical advan-

tage over the approach based on KL divergence. If we examine the worst-case variances resulting from the two approaches, Eq. 6.4.13 and 6.4.15, we can find that their positive definiteness is not guaranteed. This requires practitioners to carefully parametrise either approach to ensure the positive definiteness. However, under KL divergence the positive definiteness is dependent on the original covariance matrix. This makes it harder to parametrise and generalise the approach. In cases where the asset returns have time-varying correlations, one may need to switch parameters (θ) to ensure a positive definite matrix. On the other hand, the Wasserstein approach only requires $B - \beta A$ to be positive-definite, independent of the covariance matrix Σ . The reference probability measure thus no longer affects the feasibility of quantifying the worst-case risk.

6.4.4 Robust Portfolio Optimisation and Correlation Risk

In modern portfolio theory, one considers n risky securities with the excess logarithmic returns following a multivariate normal distribution, i.e. $X \sim \mathcal{N}(\mu, \Sigma)$. The standard mean-variance optimisation problem is formulated as

$$\min_a a^T \Sigma a \quad (6.4.16)$$

$$\text{s.t. } \mu^T a = C \quad (6.4.17)$$

where $a \in \mathbb{R}^n$ is the vector of portfolio weights. It can take any values assuming it is always possible to borrow or lend at the risk-free rate, and to short sell any asset. The problem is solved by introducing a Lagrange multiplier λ , in which case the optimal weights are given by:

$$a^* = \frac{\lambda}{2} \Sigma^{-1} \mu \quad (6.4.18)$$

The optimal portfolio weight a^* depends on λ . However, the Sharpe ratio of the optimal portfolio is independent of λ :

$$\frac{a^{*T} \mu}{\sqrt{a^{*T} \Sigma a^*}} = \sqrt{\mu^T \Sigma^{-1} \mu} \quad (6.4.19)$$

The reference model assumes a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$. The worst-case model is an alternative measure dependent on the security positions a . To formulate the problem of the worst-case measure, we may first express the mean-variance optimisation problem by

$$\min_a E [(x - \mu)^T a a^T (x - \mu) - \lambda x^T a] \quad (6.4.20)$$

where the expectation is taken under the reference measure. Taking into account the model risk, we may formulate a robust version of Eq. 6.4.20 that is consistent with the literature (Glasserman and Xu 2014):

$$\min_a \max_{Q \in \mathcal{M}} E^Q [(X - \mu)^T a a^T (X - \mu) - \lambda X^T a] \quad (6.4.21)$$

where \mathcal{M} is the space of alternative measures constrained by different criteria. For the approach based on the Kullback-Leibler divergence, the constraint is given by a maximum amount of relative entropy w.r.t the reference model (i.e. relative entropy budget). Under the Wasserstein approach, the constraints are given by Eq. 6.3.7 and 6.3.11.

To solve the inner problem of Eq. 6.4.21, we may further simplify the problem to

$$\begin{aligned} & \max_{Q \in \mathcal{M}} E^Q [(X - \mu)^T a a^T (X - \mu) - \lambda X^T a] \\ & = \max_{Q \in \mathcal{M}} E^Q [(X - \mu - k)^T a a^T (X - \mu - k)] - \lambda \mu^T a - \frac{\lambda}{4} \end{aligned} \quad (6.4.22)$$

where k is a vector that satisfies $a^T k = \lambda/2$. It is noted that this is an approximation as the change of measure would also alter the mean from μ to μ' . The variance should be calculated by $E^{Q(m)} [(X - \mu')^T a a^T (X - \mu')]$. However, the difference is proportional to $(\mu' - \mu)^2$ and is thus secondary for a small change of measure (i.e. $\beta \ll 1$). The solution to Eq. 6.4.21 is also multivariate normal under both KL divergence (see Appendix G) and under the Wasserstein metric (see Appendix H). The two approaches result in robust MVO portfolios with different weights (up to the first order w.r.t θ or β):

$$\begin{aligned} a_{KL}^* &= \left(\frac{\lambda}{2} - \frac{\theta \lambda^3}{2} (1 + \mu^T \Sigma^{-1} \mu) \right) \Sigma^{-1} \mu \\ a_W^* &= \left(\frac{\lambda}{2} - \frac{\beta \lambda^3}{4} \mu^T \Sigma^{-1} B^{-1} \Sigma^{-1} \mu \right) \Sigma^{-1} \mu - \frac{\beta \lambda^3}{4} (1 + \mu^T \Sigma^{-1} \mu) \Sigma^{-1} B^{-1} \Sigma^{-1} \mu \end{aligned} \quad (6.4.23)$$

Comparing Eq. 6.4.23 with the standard MVO portfolio given by Eq. 6.4.18, we can see that the robust MVO portfolios provide first-order corrections, resulting in more conservative asset allocation in general.

Despite of being more conservative, a_{KL}^* is in fact parallel to the standard MVO portfolio a^* . As a result, the robust MVO portfolio does not change the relative weights of component assets. In fact, all the weights are reduced by the same proportion ($c < 1$) to account for model risk. This, however, inappropriately accounts for the correlation risk. For example, two highly-correlated assets have extremely high weights in the nominal MVO portfolio. Because of the correlation risk, we would expect the robust MVO portfolio to assign them lower weights relative to other assets. This is the case for a_W^* . In fact, a_W^* not only reduces the overall portfolio weights in order to be more conservative, but also adjusts the relative weights of component assets for a less extreme allocation. One may notice that the term inside the bracket of the expression for a_W^* is a square matrix (see Eq. 6.4.23), which serves to linearly transform the vector of portfolio weights. By adjusting their relative weights, Eq. 6.6.82 correctly accounts for the correlation risk (see Appendix H for details).

The robust optimal portfolio parametrised by λ allows us to plot the robust capital market line (CML). Unlike the standard CML, it is no longer a straight line and the Sharpe ratio is now dependent on λ .

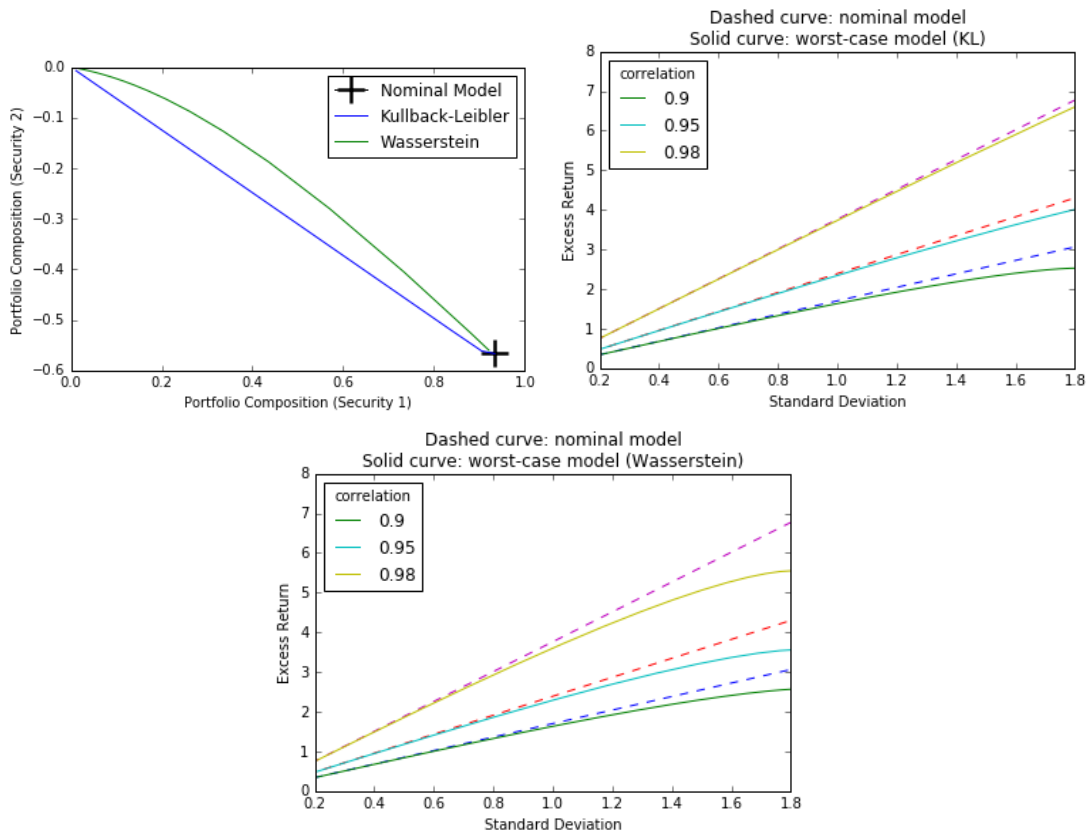


Fig 6.4.4: The normalised optimal composition of a portfolio consisting of two securities, calculated by a^* divided by $\lambda/2$. The normalised optimal composition under the reference model is give by a constant vector $\Sigma^{-1}\mu$, while those under the worst-case models are dependent on λ . In particular, the Kullback-Leibler approach reduces both compositions proportionally, while the Wasserstein approach reduces compositions in a nonlinear way.

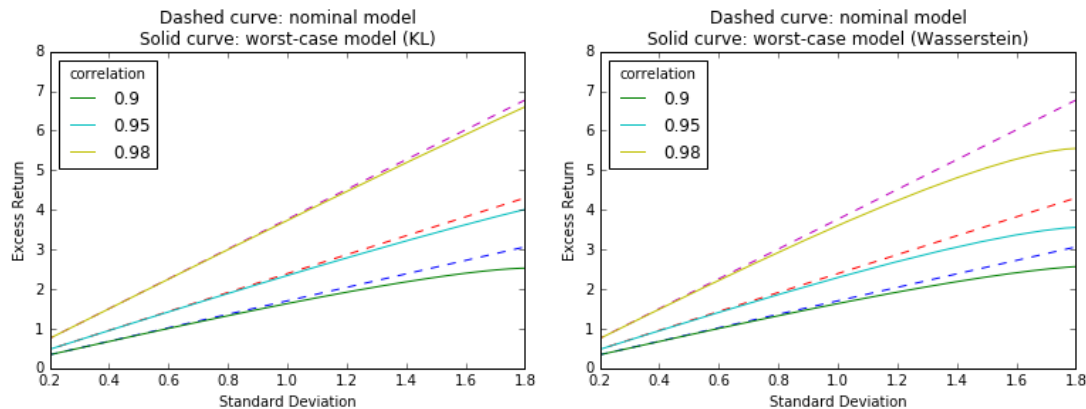


Fig 6.4.5: Robust capital allocation lines (CALs) using (a) the Kullback-Leibler divergence and (b) the Wasserstein approach.

Under the reference model, the optimal composition of a portfolio is given by $\lambda \Sigma^{-1} \mu / 2$. The proportionality of this solution suggests that we should double the weights if the expected excess return doubles. However, this may end up with excessive risk due to increased leverage. Model risk is the major source of risk here, as we are unsure if the expected excess return and the covariance matrix correctly reflect the return distribution in the future (for a given holding period). Since higher leverage amplifies model risk, increasing leverage proportionally is in fact sub-optimal under the worst-case model.

Eq. 6.4.23, on the other hand, provides the optimal solutions under the respective model risk approaches. The robustness of these solutions allow the practitioners to allocate assets in a safer way. It is shown in Fig. 6.4.4 that the normalised optimal compositions reduce with λ . This is because a larger λ indicates higher leverage, and hence the optimal composition is reduced further away from that of the reference model. The normalised optimal compositions approach zero as λ increases. In Fig. 6.4.4, the weights of both securities get reduced proportionally under the KL approach. Using the Wasserstein approach, on the other hand, allows the weights to move in a non-parallel way.

In this example, we have two highly correlated ($\rho = 0.5$) stocks but with very different expected excess returns (Stock 1 0.65 and Stock 2 -0.1). Because of the high correlation we can profit from taking the spread (long Stock 1 and short Stock 2). Under the reference model, taking the spread of a highly correlated pair does not add too much risk. However, the true risk could be underestimated due to the existence of model risk. The spread is more sensitive to model risk than an overall long position, and thus requires reduction when optimising with model risk. This point is well reflected by the non-linearity of the capital market line under the Wasserstein approach, which shows sub-linear increase in excess return as risk (standard deviation) increases. We reduce the position of the spread more than the long position of Stock 1 (or the overall long position). In the KL approach, however, we reduce the spread position and the overall

long position at the same rate.

The effect of robust optimality under the worst-case model is most significant when the reference model is close to having a low-dimensional support. A low-dimensional support means that the covariance matrix does not have the full rank. Stated differently, there exists a risk-free portfolio with non-zero weights in risky assets. In this case, there is an arbitrage opportunity that has close-to-zero risk but high excess returns. The optimal portfolio under the reference model could be unrealistically optimistic, and the arbitrage opportunity might disappear in the face of model risk.

Fig. 6.4.4 illustrates an example of two securities with a high correlation. Under the reference model, the Sharpe ratio (slope of the excess return vs risk line) increases quickly with the correlation coefficient, demonstrated by the dashed lines in Fig. 6.4.5. This results from taking excessive positions in the spread (long the one with higher Sharpe ratio and short the other). It is clear from Fig. 6.4.5 that the approach based on the Kullback-Leibler divergence cannot solve this issue systematically. In fact, when the correlation increases, the capital market line under the worst-case model is even closer to the nominal one. On the other hand, the Wasserstein approach does provide a more plausible adjustment. The robust capital market line given by the Wasserstein approach deviates more from the nominal straight line for an increasing correlation.

This difference is a direct result of their capabilities of altering the support of the reference measure. The KL approach cannot alter the support. So a spurious arbitrage relation under the reference measure may persist under the worst-case measure. On the other hand, the Wasserstein approach breaks the ostensible arbitrage opportunity by transforming the support to a different vector subspace.

6.4.5 Model Risk in Dynamic Hedging

Hedging error is measured by the absolute profit-and-loss (PnL) of a dynamically hedged option until its maturity. Using the Black-Scholes model as the reference model, the hedging risk decreases with hedging frequency. Ideally if hedging is done continuously, then the hedging error is zero almost surely. This is true even under alternative measures, as long as they are equivalent to the reference model. The underlying reason is that the quadratic variation does not change under equivalent measures. In fact, if we consider a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{6.4.24}$$

the quadratic variation $\langle \ln S \rangle_t = \sigma^2 t$ almost surely. Therefore the equation holds under all equivalent measures. Given the Black-Scholes price of an option $C_t = C(t, S_t)$, the PnL of a continuously hedged self-financing portfolio between time 0 and T is

$$\begin{aligned} & \int_0^T dC_t - \int_0^T \frac{\partial C_t}{\partial S_t} dS_t + r \int_0^T \left(S_t \frac{\partial C_t}{\partial S_t} - C_t \right) dt \\ &= \int_0^T \left(\left(\frac{\partial C_t}{\partial t} + r S_t \frac{\partial C_t}{\partial S_t} - r C_t \right) dt + \frac{S_t^2}{2} \frac{\partial^2 C_t}{\partial S_t^2} d\langle \ln S \rangle_t \right) = 0 \end{aligned} \tag{6.4.25}$$

where the last equality results from the Black-Scholes partial differential equation.

Since any f -divergence is only capable of searching over equivalent alternative measures, the worst-case hedging error given by these approaches has to be zero with continuous hedging frequency. One can imagine that as hedging frequency increases, the worst-case hedging risk decreases towards zero (Fig. 6.4.6(b)). This is, however, inconsistent with practitioners' demand for risk management. In fact, if the volatility of the underlying asset differs from the nominal volatility, then Eq. 6.4.25 no longer holds. Such volatility uncertainty is a major source of hedging risk, and thus has to be measured and managed properly. The most straightforward way of doing that is to assume a distribution of volatility, and then run a Monte Carlo simulation to quantify the hedging error (Fig. 6.4.6(a)).

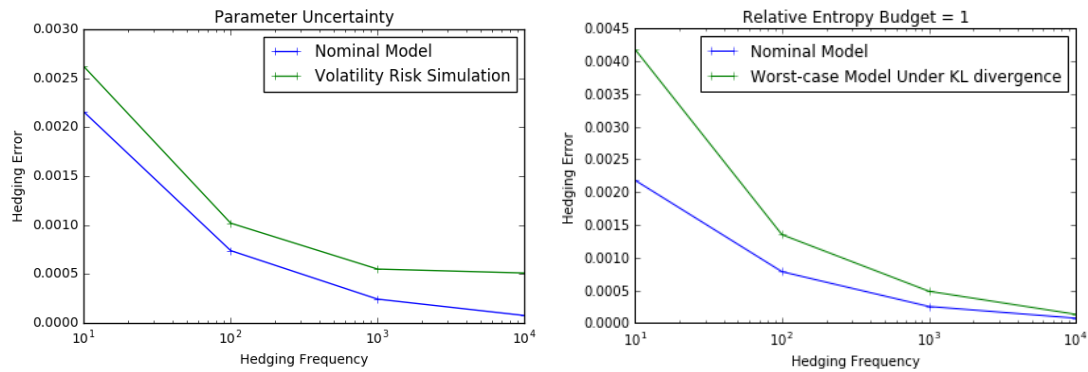


Fig 6.4.6: (a) Worst-case hedging risk under the KL divergence, and (b) hedging risk simulated by randomly sampling volatilities.

Despite of its simplicity, volatility sampling is a parametric approach, for it is only capable of generating alternative Black-Scholes models with different parameter. This approach cannot account for alternatives such as local volatility models or stochastic volatility models. This calls for a non-parametric approach relying on the formulation given in Eq. 6.3.1.

We have already seen that using approaches based on f -divergence one cannot correctly quantify the hedging risk. The Wasserstein approach, on the other hand, does not have this issue, for it is capable of searching over non-equivalent measures. Using Monte Carlo simulation, we obtain the worst-case hedging risk under the Wasserstein approach (see Fig. 6.4.7). Compared to the approach based on Kullback-Leibler divergence (Fig. 6.4.6(b)), the hedging risk given by the Wasserstein approach is more consistent with the simulated results using volatility sampling (Fig. 6.4.6(a)). In the limit of continuous hedging, the Wasserstein approach results in a worst-case risk slightly higher than volatility sampling, for it may involve jumps that cannot be hedged.

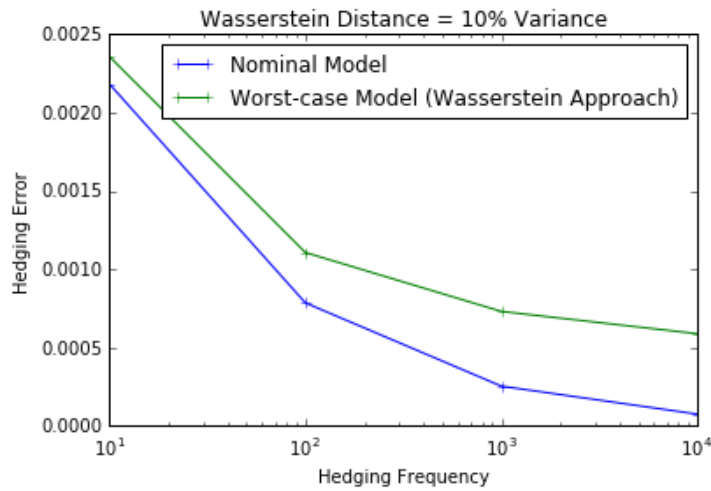


Fig 6.4.7: (a) Worst-case hedging risk under the Wasserstein approach.

In practice, the Wasserstein approach requires some tricks as fully sampling the infinite-dimensional path space is impossible. Therefore only paths close to the sampled paths (under the reference measure) are sampled, as the importance of an alternative path decays exponentially with its distance to these sampled paths. This point is shown in Fig. 6.4.8(a), in which the alternative paths are illustrated by the crosses close to the nominal sampled paths (dots). By increasing the average distance of the alternative paths to the nominal paths, the hedging risk is increased until convergence (Fig. 6.4.8(b)).

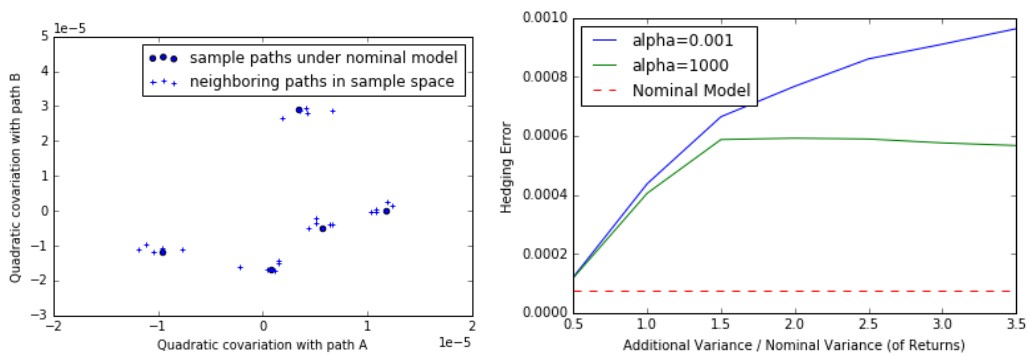


Fig 6.4.8: (a) Sample paths generated for the Wasserstein approach, (b) convergence of the worst-case hedging risk.

Here we list the procedure of the Monte Carlo simulation described in the last paragraph:

- 1 create N sample paths from the reference model
- 2 For each sample paths, create M sample paths by deviating X_t by a normally distributed random variable $\mathcal{N}(0, \sigma^2)$
- 3 collect all MN sample paths and the original N paths, we have $N(M + 1)$ points in the path space. Calculate the hedging error for each of the $N(M + 1)$ paths.
- 4 Apply Eq. 6.3.16 to calculate the worst-case probability of each path where $d(X, Y) = [X - Y]$.
- 5 To find the (worst-case) hedging risk, we average the hedging errors of all $N(M + 1)$ paths, weighted by their worst-case probabilities.
- 6 Repeat steps 2-5 with a larger σ^2 . Continue to increase the deviation until the calculated hedging risk converges (Fig. 6.4.8(b)).

6.5 Conclusion

Non-parametric approaches to model risk measurement are theoretically sound and practically feasible. Adopting the Wasserstein distance allows us to further extend the range of legitimate measures from merely the absolutely continuous ones. The Wasserstein approach is rooted in optimal transport theory, and is well suited for the adversary interpretation of model risk. It may result in a worst-case model that is more robust, in the sense that it is no longer restricted by the support of the reference measure. This is especially useful when the reference measure is supported only by a subspace determined, for instance, by the volatility of a diffusion process or the prices of perfectly correlated assets. This approach has additional practical advantage due to its ability of guaranteeing integrability.

To illustrate the Wasserstein approach, we presented four applications, ranging from single-asset variance risk and hedging risk to the multi-asset allocation problem. All the applications are connected in the sense that their reference measures are (approximately) supported by a subspace. In the example of single-asset variance risk, we looked at the limit of small variances (i.e. when the time to maturity is close to zero or the volatility close to zero). The Wasserstein approach is capable of jumping out of the family of diffusion processes, and accounts for the possibility of jumps. In the application of portfolio variance risk, the Wasserstein approach provides us with a worst-case measure induced by a linear map, thus altering the support. Its advantage of dealing with multi-asset problems is even more apparent when treating the asset allocation problem, in which the Wasserstein approach captures the effect of correlation risk. This approach results in a robust mean-variance optimal portfolio that adjusts the relative weights of the assets according to their correlations. It produces a curved capital market line, with the Sharpe ratio reduced by a larger amount for higher standard deviations or higher asset correlations. The final application is related to the

hedging risk of a vanilla option. f -divergence is incapable of quantifying the model risk associated with a continuously hedged position, because the profit-and-loss is zero almost surely. The Wasserstein approach, on the other hand, leads to a positive hedging error and therefore a more realistic assessment of model risk. In conclusion, the Wasserstein approach provides a useful tool to practitioners who aim to manage risks and optimize positions accounting for model ambiguity.

6.6 Appendix

6.6.1 A. Derivation of Eq. 6.3.10

In this part, we derive the solution Eq. 6.3.10 to the problem expressed by Eq. 6.3.5-6.3.7. For simplicity, we denote the transition density $p_{Y|X}(y|x)$ by $\gamma_x(y) := \gamma(x, y)/p(x)$. This transforms the problem into

$$\begin{aligned} & \sup_{\gamma_x \in \Gamma} \int_{\Omega} p(x) \left[\int_{\Omega} \gamma_x(y) V(y) dy \right] dx & (6.6.1) \\ & \text{s.t.} \int_{\Omega} p(x) \left[\int_{\Omega} \gamma_x(y) c(x, y) dy \right] dx \leq \eta \end{aligned}$$

where Γ is the space of probability density functions. The Karush-Kuhn-Tucker (KKT) condition in convex optimisation ensures the existence of a KKT multiplier λ such that the solution to Eq. 6.6.1 also solves

$$\sup_{\gamma_x \in \Gamma} \int_{\Omega} p(x) \left\{ \int_{\Omega} \gamma_x(y) [V(y) - \lambda c(x, y)] dy \right\} dx \quad (6.6.2)$$

The solution to Eq. 6.6.2 is a δ -function transition density $\gamma_x^*(y) = \delta(y - y^*(x))$, resulting in a transportation plan

$$\gamma^*(x, y) = p(x) \delta(y - y^*(x)) \quad (6.6.3)$$

where

$$y^*(x) = \arg \max_{y \in \Omega} [V(y) - \lambda c(x, y)] \quad (6.6.4)$$

The solution to the model risk problem is expressed either by a transportation plan (Eq. 6.6.3) or a transportation map (Eq. 6.6.4). It is noted that $\lambda = 0$ is a trivial case that we will not consider. To be consistent with the main result Eq. 6.3.16, we replace λ by its inverse $\beta = \lambda^{-1}$:

$$y^*(x) = \arg \max_{y \in \Omega} \left[V(y) - \frac{c(x, y)}{\beta} \right] \quad (6.6.5)$$

6.6.2 B. Derivation of Eq. 6.3.16 and 6.3.20

Eq. 6.3.16 is the solution of the problem formulated by Eq. 6.3.5-6.3.7 plus the additional entropy constraint Eq. 6.3.11. As in Appendix A, we introduce KKT multipliers λ and α . This converts the original constrained supremum problem to the following dual problem (same as in Appendix A we denote the transition density by $\gamma_x(y)$):

$$\begin{aligned} & \inf_{\beta, \theta \in \mathbb{R}^+} \sup_{\gamma} \int_{\Omega \times \Omega} \gamma(x, y) (V(y) - \lambda [c(x, y) - \eta] - \alpha [\ln \gamma(x, y) - \mu]) dx dy \quad (6.6.6) \\ & = \inf_{\beta, \theta \in \mathbb{R}^+} \left(\int_{\Omega} p(x) dx \left[\sup_{\gamma_x} \int_{\Omega} \gamma_x(y) (V(y) - \lambda c(x, y) - \alpha \ln \gamma_x(y)) dy \right] \right. \\ & \quad \left. + \lambda \eta + \alpha \left[\mu - \int_{\Omega} \ln p(x) dx \right] \right) \end{aligned}$$

Same as the relative entropy approach proposed by Glasserman and Xu (2014), we derive a closed-form solution to the inner part of the problem:

$$\sup_{\gamma_x} \int_{\Omega} \gamma_x(y) (V(y) - \lambda c(x, y) - \alpha \ln \gamma_x(y)) dy \quad (6.6.7)$$

It is noted that Eq. 6.6.7 asks for the supremum w.r.t the density function p_x for a given $x \in \Omega$. The solution to this problem is given by (for consistency we replace λ by its inverse γ):

$$\gamma_x^*(y) = \frac{\exp\left(\frac{V(y)}{\alpha} - \frac{c(x, y)}{\alpha\beta}\right)}{\int_{\Omega} \exp\left(\frac{V(z)}{\alpha} - \frac{c(x, z)}{\alpha\beta}\right) dz} \quad (6.6.8)$$

The worst-case probability density function is the marginal distribution of y , induced by the transition density function $\gamma_x^*(y)$:

$$\begin{aligned} p^*(y) &= \int_{\Omega} p(x) \gamma_x^*(y) dx \\ &= \int_{\Omega} dx \frac{p(x) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x, y)}{\alpha\beta}\right)}{\int_{\Omega} \exp\left(\frac{V(z)}{\alpha} - \frac{c(x, z)}{\alpha\beta}\right) dz} \end{aligned} \quad (6.6.9)$$

Eq. 6.3.20 is derived in a similar way. Since we lift the entropy constraint Eq. 6.3.11 into a relative entropy constraint Eq. 6.3.14, the inner problem Eq. 6.6.7 requires slight modification:

$$\sup_{\gamma_x} \int_{\Omega} \gamma_x(y) \left(V(y) - \lambda c(x, y) - \alpha \ln \frac{\gamma_x(y)}{q_0(y)} \right) dy \quad (6.6.10)$$

This problem has the same formulation as the supremum problem given in Glasserman and Xu's work, and therefore shares the same solution

$$\gamma_x^*(y) = \frac{q_0(y) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x, y)}{\alpha\beta}\right)}{\int_{\Omega} q_0(z) \exp\left(\frac{V(z)}{\alpha} - \frac{c(x, z)}{\alpha\beta}\right) dz} \quad (6.6.11)$$

This equation differs from Eq. 6.6.8 merely by a prior distribution q_0 . It takes Eq. 6.6.8 as its special case where q_0 is a uniform distribution. Marginalizing the transition density Eq. 6.6.11 gives the worst-case distribution shown in Eq. 6.3.20.

6.6.3 C. Jump Risk and Variance Risk

Under a diffusive model, the logarithmic return of an asset follows a normal distribution with mean of μT and variance of $\sigma^2 T$, where σ is the volatility and T is the time to maturity, and the drift coefficient of this process is assumed to be $\mu + \sigma^2/2$. The probability density function of the return x is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu T}{2\sigma^2 T}\right)^2} \quad (6.6.12)$$

Applying Eq. 6.3.16, one may obtain the probability density function of the worst-case measure, assuming a linear loss function $V(x) = x$ and a quadratic transportation cost function $c(x, y) = (x - y)^2$,

$$\begin{aligned} q^*(y) &\propto \int_{\Omega} \left[p(x) \exp\left(\frac{y}{\alpha} - \frac{(x-y)^2}{\alpha\beta}\right) / \exp\left(\frac{x}{\alpha}\right) \right] dx \\ &= \int_{\Omega} \exp\left(\frac{y-x}{\alpha} - \frac{(x-y)^2}{\alpha\beta} - \frac{(x-\mu T)^2}{2\sigma^2 T}\right) dx \\ &\propto \exp\left(-\frac{(y-\mu T - \beta/2)^2}{2\sigma^2 T + \alpha\beta}\right) \end{aligned} \quad (6.6.13)$$

Unlike the result given by the KL divergence, Eq. 6.6.13 not only shifts the mean the distribution but also enlarges the variance as a result of additional uncertainty. On $\sigma \rightarrow 0$, the worst-case measure is no longer a Dirac measure, showing consideration of jump risks:

$$\lim_{\sigma \rightarrow 0} q^*(y) \propto \exp\left(-\frac{(y-\mu T - \beta/2)^2}{\alpha\beta}\right) \quad (6.6.14)$$

This gives Eq. 6.4.6. Alternatively, one may first derive Eq. 6.4.5 followed by substituting $V(x) = x$ to get Eq. 6.4.6. Eq. 6.4.5 is derived by substituting $p(x) = \delta(x - \mu T)$ into Eq. 6.3.16:

$$q^*(y) = \int_{\Omega} \delta(x - \mu T) \frac{\exp\left(\frac{V(y)}{\alpha} - \frac{(x-y)^2}{\alpha\beta}\right)}{\int_{\Omega} \exp\left(\frac{V(z)}{\alpha} - \frac{(x-z)^2}{\alpha\beta}\right) dz} dx \quad (6.6.15)$$

$$= \frac{\exp\left(\frac{V(y)}{\alpha} - \frac{(y-\mu T)^2}{\alpha\beta}\right)}{\int_{\Omega} \exp\left(\frac{V(z)}{\alpha} - \frac{(x-z)^2}{\alpha\beta}\right) dz} \quad (6.6.16)$$

Now we adopt a quadratic type of loss function, $V(x) = (x - \mu T)^2$, following a procedure similar to Eq. 6.6.13 we get

$$q^*(y) \propto \exp\left(-\frac{(y - \mu T)^2}{\frac{2\sigma^2 T}{(1-\beta)^2} + \frac{\alpha\beta}{(1-\beta)}}\right) \quad (6.6.17)$$

the variance of the worst-case measure is

$$\sigma_W^2 T = \frac{\sigma^2 T}{(1-\beta)^2} + \frac{\alpha\beta}{2(1-\beta)} \quad (6.6.18)$$

as provided in Eq. 6.4.8. We may verify that the measure Q^* given by Eq. 6.6.17 does provide the largest variance among all the legitimate alternative measures. In fact, the variance of x under Q^* is

$$E^{Q^*} [(x - E^{Q^*}(x))^2] = E^{Q^*} [(x - \mu T)^2] \quad (6.6.19)$$

According to the definition of the worst-case model, for all $Q \in \mathcal{M}$ (the space of legitimate alternative measures) we have

$$E^{Q^*} [(x - \mu T)^2] \geq E^Q [(x - \mu T)^2] \quad (6.6.20)$$

$$= E^Q [(x - E^Q(x))^2] + (E^Q(x) - \mu T)^2 \quad (6.6.21)$$

$$\geq E^Q [(x - E^Q(x))^2] \quad (6.6.22)$$

This confirms that Eq. 6.6.18 is indeed the worst-case (maximum) variance.

6.6.4 D. Worst-case Portfolio Variance

To find the portfolio variance under the worst-case scenario, we need to formulate the problem using Eq. 6.3.1 with a loss (target) function given by Eq. 6.4.9. The worst-case measure may be evaluated by substituting the loss function into Eq. 6.3.16. In this section we will show the calculation step by step. First, we need to specify the transport cost function $c(x, y)$ as the inner product introduced in Eq. 6.4.11:

$$c(x, y) = \|y - x\|^2 = (y - x)^T B (y - x) \quad (6.6.23)$$

Then we evaluate the following part in Eq. 6.3.16:

$$\begin{aligned} & \exp\left(\frac{V(y)}{\alpha} - \frac{c(x, y)}{\alpha\beta}\right) \\ &= \exp\left(\frac{y^T A y}{\alpha} - \frac{(y - x)^T B (y - x)}{\alpha\beta}\right) \\ &= \exp\left(\frac{1}{\alpha\beta} x^T B ((B - \beta A)^{-1} - I) B x\right) \end{aligned}$$

$$-\frac{1}{\alpha\beta} \left(y - (B - \beta A)^{-1} Bx \right)^T (B - \beta A) \left(y - (B - \beta A)^{-1} Bx \right) \quad (6.6.24)$$

Remember that both A and B are symmetric, positive-definite matrices. Fixing x , Eq. 6.6.24 is proportional to the probability density function of a multivariate normal variable Y , with its mean and covariance matrix

$$\mathbb{E}(Y) = (B - \beta A)^{-1} Bx \quad (6.6.25)$$

$$\Sigma(Y) = \frac{\alpha\beta}{2} (B - \beta A)^{-1} \quad (6.6.26)$$

This means that after normalization w.r.t y , Eq. 6.6.24 gives exactly the probability density function of Y . We may write this down explicitly by noting that y lives in the n -dimensional vector space, i.e. $\Omega = \mathcal{V}$:

$$\begin{aligned} & \frac{\exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right)}{\int_{\mathcal{V}} \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right) dy} \quad (6.6.27) \\ &= (2\pi)^{-\frac{n}{2}} \sqrt{\frac{\alpha\beta}{2} |B - \beta A|} \exp\left(-\left(y - (B - \beta A)^{-1} Bx\right)^T \frac{(B - \beta A)}{\alpha\beta} \left(y - (B - \beta A)^{-1} Bx\right)\right) \end{aligned}$$

Now we need to evaluate the product of Eq. 6.6.27 and the nominal distribution $p(x)$. The nominal distribution is multivariate normal with mean μ and covariance matrix Σ :

$$p(x) = \frac{(2\pi)^{-\frac{n}{2}}}{|\Sigma|} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right) \quad (6.6.28)$$

The product contains many terms of x and y . One may re-arrange the terms to isolate quadratic and linear terms of x :

$$\begin{aligned} & \frac{p(x) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right)}{\int_{\mathcal{V}} \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right) dy} \\ & \propto \exp\left(-\frac{1}{\alpha\beta} \left(y - (B - \beta A)^{-1} Bx \right)^T (B - \beta A) \left(y - (B - \beta A)^{-1} Bx \right) \right. \\ & \quad \left. - \frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right) \\ & \propto \exp\left(-\frac{1}{\alpha\beta} \left[(x - K\mu - Ly)^T M (x - K\mu - Ly) - (K\mu + Ly)^T M (K\mu + Ly) \right] \right. \\ & \quad \left. - \frac{1}{\alpha\beta} y^T (B - \beta A) y \right) \quad (6.6.29) \end{aligned}$$

where

$$\begin{aligned} M &:= B(B - \beta A)^{-1} B + \frac{\alpha\beta}{2} \Sigma^{-1} \\ K &:= \frac{\alpha\beta}{2} M^{-1} \Sigma^{-1} \\ L &:= M^{-1} B \end{aligned} \quad (6.6.30)$$

Fixing y , Eq. 6.6.29 is proportional to the probability density function of a multivariate normal variable X where

$$\mathbb{E}(X) = K\mu + Ly \quad (6.6.31)$$

$$\Sigma(X) = \frac{\alpha\beta}{2}M^{-1} \quad (6.6.32)$$

The following integral

$$\int_{\mathcal{V}} \exp\left(-\frac{1}{\alpha\beta}(x - K\mu - Ly)^T M(x - K\mu - Ly)\right) dy = \frac{2}{\alpha\beta}(2\pi)^{-\frac{n}{2}}|M|^{-1} \quad (6.6.33)$$

is constant irrespective of y . Integrating Eq. 6.6.29 over x gives the worst-case probability density function $q^*(y)$:

$$\begin{aligned} q^*(y) &= \int_{\mathcal{V}} dx \frac{p(x) \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right)}{\int_{\Omega} \exp\left(\frac{V(y)}{\alpha} - \frac{c(x,y)}{\alpha\beta}\right) dy} \\ &\propto \int_{\mathcal{V}} \exp\left(-\frac{1}{\alpha\beta}(x - K\mu - Ly)^T M(x - K\mu - Ly)\right) dx \\ &\quad \times \exp\left[\frac{1}{\alpha\beta}(K\mu + Ly)^T M(K\mu + Ly) - \frac{1}{\alpha\beta}y^T(B - \beta A)y\right] \\ &\propto \exp\left[\frac{1}{\alpha\beta}(K\mu + Ly)^T M(K\mu + Ly) - \frac{1}{\alpha\beta}y^T(B - \beta A)y\right] \\ &= \exp\left[\frac{1}{\alpha\beta}\left(\left(\frac{\alpha\beta}{2}\Sigma^{-1}\mu + By\right)^T M^{-1}\left(\frac{\alpha\beta}{2}\Sigma^{-1}\mu + By\right) - y^T(B - \beta A)y\right)\right] \\ &\propto \exp\left[-\frac{1}{2}(y - B(B - \beta A)^{-1}\mu)^T \Sigma^{*-1}(y - B(B - \beta A)^{-1}\mu)\right] \end{aligned} \quad (6.6.34)$$

where

$$\begin{aligned} \Sigma^{*-1} &= \frac{2}{\alpha\beta}(B^T M^{-1}B - (B - \beta A)) \\ &= \frac{2}{\alpha\beta}\left(\left((B - \beta A)^{-1} + \frac{\alpha\beta}{2}B^{-1}\Sigma^{-1}B^{-1}\right)^{-1} - (B - \beta A)\right) \\ &= \frac{2}{\alpha\beta}\left(\left((B - \beta A)^{-1} + \frac{\alpha\beta}{2}B^{-1}\Sigma^{-1}B^{-1}\right)^{-1}\left(I - \left(I + \frac{\alpha\beta}{2}B^{-1}\Sigma^{-1}B^{-1}(B - \beta A)\right)\right)\right) \\ &= \left((B - \beta A)^{-1} + \frac{\alpha\beta}{2}B^{-1}\Sigma^{-1}B^{-1}\right)^{-1}\left((B - \beta A)^{-1}B\Sigma B\right)^{-1} \\ &= \left((B - \beta A)^{-1}B\Sigma B(B - \beta A)^{-1} + \frac{\alpha\beta}{2}(B - \beta A)^{-1}\right)^{-1} \end{aligned}$$

Eq. 6.6.34 shows that the worst-case distribution is still multivariate normal. The vector of means and the covariance matrix are given respectively by

$$\mu_W = (B - \beta A)^{-1} B \mu \quad (6.6.35)$$

$$\Sigma_W = (B - \beta A)^{-1} B \Sigma B (B - \beta A)^{-1} + \frac{\alpha \beta}{2} (B - \beta A)^{-1} \quad (6.6.36)$$

An interesting observation on Eq. 6.6.35 is that the worst-case measure can be generated by a measure-preserving linear map. In fact, for any vector v of asset returns, the linear map g gives

$$\begin{aligned} g(v) &= (B - \beta A)^{-1} B v \\ &= (I - \beta B^{-1} A)^{-1} v \end{aligned} \quad (6.6.37)$$

We write down the probability density function for the reference measure by

$$f(v) \propto \exp\left(-\frac{1}{2} (v - \mu)^T \Sigma^{-1} (v - \mu)\right) \quad (6.6.38)$$

The measure given by the measure-preserving map g has a probability density function that is proportional to $f(g^{-1}(v))$,

$$\begin{aligned} & f(g^{-1}(v)) \\ & \propto \exp\left(-\frac{1}{2} ((I - \beta B^{-1} A)v - \mu)^T \Sigma^{-1} ((I - \beta B^{-1} A)v - \mu)\right) \\ & = \exp\left(-\frac{1}{2} (v - (I - \beta B^{-1} A)^{-1} \mu)^T (I - \beta B^{-1} A) \Sigma^{-1} (I - \beta B^{-1} A) ((I - \beta B^{-1} A)v - \mu)\right) \\ & = \exp\left(-\frac{1}{2} (v - \tilde{\mu})^T \tilde{\Sigma}^{-1} (v - \tilde{\mu})\right) \end{aligned} \quad (6.6.39)$$

where

$$\tilde{\mu} := (I - \beta B^{-1} A)^{-1} \mu \quad (6.6.40)$$

$$\tilde{\Sigma} := (I - \beta B^{-1} A)^{-1} \Sigma (I - \beta B^{-1} A)^{-1} \quad (6.6.41)$$

that are precisely the mean and covariance matrix given in Eq. 6.6.35 (with $\alpha = 0$). As a result, we generate the worst-case measure by applying the measure-preserving map g .

6.6.5 E. The support of a multivariate normal distribution

In this section, we discuss the support of the reference measure P assuming the asset returns follow a multivariate normal distribution. In addition, we want see how it is altered by different approaches to model risk measurement. Clearly, approaches based on f -divergence cannot alter the support as they only account for measures that are

equivalent to the nominal one. But this conclusion does not tell us explicitly what the support is. In the following work we aim to find the linear subspace that supports the measure.

Formally speaking, returns of the n assets form a n -dimensional vector that lives in a n -dimensional topological vector space \mathcal{V} . If the asset returns follow a multivariate normal distribution with a non-singular covariance matrix, then the support is the entire space \mathcal{V} . However, if the covariance matrix is singular, the support can only be part of \mathcal{V} . We will find this support and show that it is a m -dimensional linear subspace, where m is the rank of the covariance matrix.

The reference model of asset returns defines a probability space $(\mathcal{V}, \mathcal{F}, P)$, where \mathcal{F} is the Borel σ -algebra on \mathcal{V} . Since \mathcal{V} is a vector space, we may consider its dual space \mathcal{V}^* , i.e. the space of linear maps $a : \mathcal{V} \rightarrow \mathbb{R}$. Any element of the dual space is regarded as a vector of portfolio weights. To see this, suppose the asset returns are $v = (v_1, v_2, \dots, v_n) \in \mathcal{V}$, and the portfolio weights are $a = (a_1, a_2, \dots, a_n) \in \mathcal{V}^*$. The pairing of a and v results in a real number, which is exactly the portfolio return:

$$a(v) = \sum_{j=1}^n a_j v_j \quad (6.6.42)$$

If we treat the asset returns v_i as random variables, we may calculate of portfolio variance on a given vector of weights $a \in \mathcal{V}$ by $\text{Var}(a(v)) = a^T \Sigma a$, where Σ is the covariance matrix of the asset returns. For convenience, we use the same symbol v for both the vector of random variables (random vector) and its realization (i.e. a specific element in \mathcal{V}). Now take the positive semi-definite matrix Σ as a linear map $\Sigma : \mathcal{V}^* \rightarrow \mathcal{V}$:

$$\Sigma(a) = \Sigma a \in \mathcal{V}, \quad \forall a \in \mathcal{V}^* \quad (6.6.43)$$

The portfolio variance is formed by applying the linear map $a : \mathcal{V} \rightarrow \mathbb{R}$ to $\Sigma(a) \in \mathcal{V}$: $\text{Var}(a(v)) = a(\Sigma(a))$. If the square matrix Σ is singular, then its kernel $\ker \Sigma$ is not trivial (i.e. contains elements other than the zero vector). \mathcal{V}^* can therefore be decomposed into two subspaces:

$$\mathcal{V}^* = \ker \Sigma \oplus \ker \Sigma^\perp \quad (6.6.44)$$

Suppose $\ker \Sigma^\perp$ has dimension n . $\ker \Sigma$ has dimension $m - n$ for the dimensions of subspaces sum up to the dimension of \mathcal{V}^* . We may switch to a new orthonormal basis $\{e_1^*, e_2^*, \dots, e_m^*, k_1^*, k_2^*, \dots, k_{m-n}^*\}$ in consistency with the decomposition Eq. 6.6.44, in the sense that $e_1^*, e_2^*, \dots, e_m^*$ span $\ker \Sigma^\perp$ and $k_1^*, k_2^*, \dots, k_{m-n}^*$ span $\ker \Sigma$. Now get back the original space of asset returns \mathcal{V} , we may select a new basis $\{e_1, e_2, \dots, e_m, k_1, k_2, \dots, k_{m-n}\}$, dual to $\{e_1^*, e_2^*, \dots, e_m^*, k_1^*, k_2^*, \dots, k_{m-n}^*\}$, i.e.

$$\begin{aligned} e_i^*(e_j) &= \delta_{i-j} \\ k_i^*(k_j) &= \delta_{i-j} \\ e_i^*(k_j) &= 0 \end{aligned}$$

$$k_i^*(e_j) = 0$$

Any $v \in \mathcal{V}$ can be expressed by

$$v = \sum_{i=1}^m u_i e_i + \sum_{i=1}^{m-n} w_i k_i \quad (6.6.45)$$

Suppose \mathcal{U} denotes the linear subspace spanned by e_1, e_2, \dots, e_m . \mathcal{U} is in fact the dual space of $\ker \Sigma^\perp$. We will show that the support of the reference measure P is indeed the linear subspace \mathcal{U} shifted by the vector of average asset returns μ :

Theorem Given a finite-dimensional topological vector space \mathcal{V} and its Borel α -algebra \mathcal{F} , the support of a measure P on $(\mathcal{V}, \mathcal{F})$ is $\{v \in \mathcal{V} : v - \mu \in \mathcal{U}\}$ if P provides a multivariate distribution $\mathcal{N}(\mu, \Sigma)$.

Proof For every $v \in \ker \Sigma$, consider the variance of $a(v)$ (v is a random vector here):

$$\text{Var}(a(v)) = a^T \Sigma a = 0 \quad (6.6.46)$$

The zero variance implies that a carries the measure P on \mathcal{V} to a Dirac measure P_a on \mathbb{R}

$$P_a(A) = P(a^{-1}(A)), \quad \forall A \in \{A \subseteq \mathbb{R} : a^{-1}(A) \in \mathcal{F}\} \quad (6.6.47)$$

Suppose $\text{supp}(P_a) = \{s_a\}$ where $s_a \in \mathbb{R}$. We can show that $\text{supp}(P)$ should only include elements in \mathcal{V} that is projected to s_a . More formally, with the projection map $\mathcal{P} : \mathcal{V} \rightarrow \ker \Sigma$, we have

$$\{v \in \mathcal{V} : \exists a \in \ker \Sigma, a(v) \neq s_a\} \cap \text{supp}(P) = \emptyset \quad (6.6.48)$$

In fact, for a given $v \in \mathcal{V}$, suppose there exists $a \in \ker \Sigma$ such that $a(v) \neq s_a$. $a(v)$ is not in the $\text{supp}(\tilde{P})$, suggesting the existence of an open neighborhood $N_{a(v)} \subseteq \mathbb{R}$ such that $P_a(N_{a(v)}) = 0$. Since the linear map a is continuous, $a^{-1}(N_{a(v)})$ is an open neighborhood of v and

$$P(a^{-1}(N_{a(v)})) = P_a(N_{a(v)}) = 0 \quad (6.6.49)$$

As a result, $v \notin \text{supp}(P)$ which proves Eq. 6.6.48.

Now we consider the set $S := \{v \in \mathcal{V} : a(v) = s_a, \forall a \in \ker \Sigma\}$. For a given $v_s \in S$, every $v \in S$ satisfies

$$a(v - v_s) = a(v) - a(v_s) = 0, \quad \forall a \in \ker \Sigma \quad (6.6.50)$$

suggesting that $v - v_s \in \mathcal{U}$. Therefore $S = \{v \in \mathcal{V} : v - v_s \in \mathcal{U}\}$. Regard \mathcal{U} as a topological linear subspace of \mathcal{V} equipped with the relative topology. $\tilde{\mathcal{F}}$ is the Borel σ -algebra on \mathcal{U} . We may define a new probability space $(\mathcal{U}, \tilde{\mathcal{F}}, \tilde{P})$ by

$$\tilde{P}(A \cap \mathcal{U}) = P(A), \quad \forall A \in \mathcal{F} \quad (6.6.51)$$

One can verify that this probability space is well defined. Now we would like to show that $\text{supp}(\tilde{P}) = \mathcal{U}$. In fact, assuming this is true, then for arbitrary $v \in S$ every open neighborhood $N(v)$ has positive measure:

$$P(N(v)) = \tilde{P}(N(v) \cap \mathcal{U}) > 0 \quad (6.6.52)$$

This immediately leads to the result $\text{supp}(P) = S$. In particular, from the property of the multivariate normal distribution, $\text{supp}(P)$ includes the vector μ of average asset returns. This means that $\mu \in S$, and thus the support of P can be written as $\text{supp}(P) = S = \{v \in \mathcal{V} : v - \mu \in \mathcal{U}\}$.

Now we only need to show that $\text{supp}(\tilde{P}) = \mathcal{U}$. Consider the projection map $\mathcal{P} : \mathcal{V} \rightarrow \mathcal{U}$ that sends $v = (u_1, u_2, \dots, u_m, w_1, w_2, \dots, w_{m-n})$ to $u = (u_1, u_2, \dots, u_m)$. The projection results in the marginal distribution w.r.t u_1, u_2, \dots, u_m . This marginal distribution characterises a measure P' on the subspace \mathcal{U} :

$$P'(A) = P(\mathcal{P}^{-1}(A)), \quad \forall A \in \{A \subseteq \mathcal{U} : \mathcal{P}^{-1}(A) \in \mathcal{F}\} \quad (6.6.53)$$

For any $A \in \mathcal{F}$,

$$\begin{aligned} P'(A \cap \mathcal{U}) &= P(\mathcal{P}^{-1}(A \cap \mathcal{U})) \\ &= \tilde{P}(\mathcal{P}^{-1}(A \cap \mathcal{U}) \cap \mathcal{U}) \\ &= \tilde{P}(A \cap \mathcal{U}) \end{aligned} \quad (6.6.54)$$

Therefore, the two measures \tilde{P} and P' coincide, and we only need to prove $\text{supp}(P') = \mathcal{U}$. The marginal distribution from projection \mathcal{P} is apparently multivariate normal (every linear combination of the elements in u is also a linear combination of the elements in $v \in \mathcal{P}^{-1}(u)$ thus normally distributed).

The covariance matrix $\tilde{\Sigma}$ of the truncated random vector u is invertible. In fact, because $\Sigma(a)$ is not a zero vector for every non-zero $a \in \ker \Sigma^\perp$, the linear map between two m -dimensional vector spaces $\Sigma|_{\ker \Sigma^\perp} : \ker \Sigma^\perp \rightarrow \Sigma(\ker \Sigma^\perp)$ is invertible. Represented by a $m \times m$ matrix, $\Sigma|_{\ker \Sigma^\perp}$ has only non-zero eigenvalues. Since it is also positive semi-definite (for $\text{Var}(a(v)) = a(\Sigma(a)) \geq 0, \forall a \in \ker \Sigma^\perp \subseteq \mathcal{V}^*$), it must be positive definite. We conclude that for every non-zero $a \in \ker \Sigma^\perp$, $\text{Var}(a(v)) = a(\Sigma|_{\ker \Sigma^\perp}(a)) > 0$. If we expand $a(v)$ component-wise according to Eq. 6.6.45,

$$\begin{aligned} a(v) &= \sum_{i=1}^m u_i a(e_i) + \sum_{i=1}^{m-n} w_i a(k_i) \\ &= a \left(\sum_{i=1}^m u_i e_i \right) = a(u) \end{aligned} \quad (6.6.55)$$

Therefore $a^T \tilde{\Sigma} a = \text{Var}(a(u)) = \text{Var}(a(v)) > 0$ for every nonzero $a \in \ker \Sigma^\perp$. As a result, $\tilde{\Sigma} a$ is positive-definite and thus invertible. Under the measure P' , the random vector u follows a multivariate normal distribution with a non-singular covariance matrix. It is supported by the entire subspace \mathcal{U} , i.e. $\text{supp}(P') = \mathcal{U}$. \square

For a multivariate distribution $\mathcal{N}(\mu, \Sigma)$, the support $\text{supp}(P) = \{v \in \mathcal{V} : v - \mu \in \mathcal{U}\}$ only depends on the vector μ and the kernel of Σ . It is clear that under the Kullback-Leibler divergence the worst-case measure shares the same support. In fact, the worst-case distribution is $\mathcal{N}(\mu_{KL}, \Sigma_{KL})$ where μ_{KL} and Σ_{KL} are given in Eq. 6.4.14. Assuming θ is sufficiently small so that $I - 2\theta\Sigma A$ is invertible, $\Sigma_{KL}a = 0$ if and only if $\Sigma_{KL} = 0$ for every $a \in \mathcal{V}^*$. Therefore, Σ_{KL} and Σ share the same kernel and therefore the same subspace $\mathcal{U} \subseteq \mathcal{V}$. In addition, $\mu_{KL} - \mu \in \mathcal{U}$ because for every $a \in \ker\Sigma$ we have

$$\begin{aligned} a(\mu_{KL} - \mu) &= a(2\theta\Sigma A(I - 2\theta\Sigma A)^{-1}\mu) \\ &= 2\theta(\Sigma a)^T A(I - 2\theta\Sigma A)^{-1}\mu \\ &= 0 \end{aligned} \tag{6.6.56}$$

As a result, the support of the worst-case measure is $\{v \in \mathcal{V} : v - \mu_{KL} \in \mathcal{U}\} = \{v \in \mathcal{V} : v - \mu \in \mathcal{U}\}$, same as the support of the reference measure.

On the other hand, the worst-case measured resulted from the Wasserstein approach can have different support. According to Eq. 6.4.12, the worst-case covariance matrix Σ_W has a different kernel in general. In addition, $\mu_W - \mu = \beta A(B - \beta A)^{-1}\mu$ is not dependent on Σ , thus not linked to the subspace \mathcal{U} . Setting $\alpha = 0$ in Eq. 6.4.12 provides a particularly interesting case, where the worst-case measure is given by a measure-preserving linear map $g : \mathcal{V} \rightarrow \mathcal{V}$ given by Eq. 6.6.37. As a result, the support of the worst-case measure can be obtained using the same map, i.e.

$$\begin{aligned} &\{v \in \mathcal{V} : g^{-1}(v) - \mu \in \mathcal{U}\} \\ &= \{v \in \mathcal{V} : (I - \beta B^{-1}A)v - \mu \in \mathcal{U}\} \\ &= \{v \in \mathcal{V} : v - (I - \beta B^{-1}A)^{-1}\mu \in \{(I - \beta B^{-1}A)^{-1}u : u \in \mathcal{U}\}\} \\ &= \{v \in \mathcal{V} : v - \mu_W \in \mathcal{U}_W\} \end{aligned} \tag{6.6.57}$$

$\mathcal{U}_W := \{(I - \beta B^{-1}A)^{-1}u : u \in \mathcal{U}\} \subseteq \mathcal{V}$ is the linear subspace (perpendicular to $\ker\Sigma_W$) that corresponds to the worst-case scenario under the Wasserstein approach.

6.6.6 F. Verification of the Wasserstein approach

Sec. 6.6.5 shows that under the Wasserstein approach the worst-case measure does alter the support. Now the question is whether the approach searches over all alternative measures. Unlike f -divergence that is only capable of measuring distance between equivalent measures, the Wasserstein metric provides a finite distance between non-equivalent measures as well. Therefore the Wasserstein approach should be able to find out the worst-case measure from all equivalent and non-equivalent measures. In this section, we will verify it for the example of portfolio variance. In particular, we will find out a worst-case linear map $g^* : \mathcal{V} \rightarrow \mathcal{V}$ by searching over the entire space of linear maps. We will verify that Eq. 6.4.14 (with $\alpha = 0$) can be given by the worst-case linear map.

Theorem Given a probability space $(\mathcal{V}, \mathcal{F}, P)$ where \mathcal{V} is a finite-dimensional vector space and P provides a multivariate distribution $\mathcal{N}(\mu, \Sigma)$, there exists a worst-case linear map $g^* : \mathcal{V} \rightarrow \mathcal{V}$ in the sense of Eq. 6.3.10, i.e.

$$g^*(x) = \arg \max_{y \in \mathcal{V}} \left[y^T A y - \frac{(x - y)^T B (x - y)}{\beta} \right] \quad (6.6.58)$$

for every non-zero $x \in \mathcal{V}$, as long as $B - \beta A$ is positive definite.

Proof Given a non-zero $x \in \mathcal{V}$, every non-zero $y \in \mathcal{V}$ can be expressed by $y = g(x)$ where g is some linear map (not unique) $g : \mathcal{V} \rightarrow \mathcal{V}$. The problem Eq. 6.6.58 is therefore equivalent to

$$g^*(x) = \arg \max_{g \in \mathfrak{L}(\mathcal{V}, \mathcal{V})} \left[g(x)^T A g(x) - \frac{(x - g(x))^T B (x - g(x))}{\beta} \right] (x) \quad (6.6.59)$$

where $\mathfrak{L}(\mathcal{V}, \mathcal{V})$ is the space of all linear maps from \mathcal{V} to \mathcal{V} . Choosing an orthonormal basis for \mathcal{V} allows us to represent g by a square matrix, and the linear map $g(x)$ by matrix multiplication gx . The expression inside the square bracket in Eq. 6.6.59 is then transformed into

$$\begin{aligned} & (gx)^T A gx - \frac{(x - gx)^T B (x - gx)}{\beta} \\ &= -\frac{1}{\beta} x^T (g^T - B(B - \beta A)^{-1}) (B - \beta A) (g - (B - \beta A)^{-1} B) x \\ & \quad - \frac{1}{\beta} x^T (B - B(B - \beta A)^{-1} B) x \end{aligned} \quad (6.6.60)$$

Since $B - \beta A$ is positive-definite, the first term in Eq. 6.6.60 is either zero or negative. It reaches zero (and hence Eq. 6.6.60 reaches its maximum value) if and only if

$$(g - (B - \beta A)^{-1} B) x = 0 \quad (6.6.61)$$

or equivalently

$$g(x) = (B - \beta A)^{-1} B x \quad (6.6.62)$$

This allows to rewrite Eq. 6.6.59 by

$$g^*(x) = (B - \beta A)^{-1} B x \quad (6.6.63)$$

The linear map g^* given by the square matrix $(B - \beta A)^{-1} B$ satisfies Eq. 6.6.63 and thus solves Eq. 6.6.59 for every non-zero $x \in \mathcal{V}$. \square

It is noted that in the problem of portfolio variance risk, both square matrices A and B are symmetric and positive definite. Therefore if the positive multiplier β is sufficiently small, $B - \beta A$ is also positive definite satisfying the condition assumed in the theorem above. Now the worst-case linear map g^* transforms the vector of asset returns from μ to $(B - \beta A)^{-1} B \mu$, and the covariance matrix from Σ to $(B - \beta A)^{-1} B \Sigma B (B - \beta A)^{-1}$, same as the expressions given in Eq. 6.4.14 (with $\alpha = 0$). This verifies that the Wasserstein approach indeed searches over the entire space $\mathfrak{L}(\mathcal{V}, \mathcal{V})$ of linear maps. It results in a measure that corresponds to the worst-case linear map g^* .

6.6.7 G. Robust MVO Portfolio (Kullback-Leibler divergence)

According to Eq. 6.4.22, we consider the problem

$$\max_{Q \in \mathcal{M}} E^Q \left((X - \mu - k)^T a a^T (X - \mu - k) \right) \quad (6.6.64)$$

Since $X - \mu - k \sim \mathcal{N}(-k, \Sigma)$, under the Kullback-Leibler divergence the covariance matrix and the mean of the worst-case measure are given according to Eq. 6.4.14 (remember $a^T k = \lambda/2$):

$$\begin{aligned} \Sigma_{KL} &= (I - 2\theta \Sigma A)^{-1} \Sigma \\ \mu_{KL} &= (\mu + k) - (I - 2\theta \Sigma A)^{-1} k \\ &= \mu - \lambda \theta (I - 2\theta \Sigma A)^{-1} \Sigma a \end{aligned} \quad (6.6.65)$$

Using the Wassertein approach, however, the worst-case measure has different covariance matrix and mean (Eq. 6.4.12):

$$\begin{aligned} \Sigma_W &= (I - \beta B^{-1} A)^{-1} \Sigma (I - \beta A B^{-1})^{-1} \\ \mu_W &= (\mu + k) - (I - \beta B^{-1} A)^{-1} k \\ &= \mu - \frac{\lambda}{2} \beta (I - \beta B^{-1} A)^{-1} B^{-1} a \end{aligned} \quad (6.6.66)$$

We may then formulate the optimal asset allocation a^* under the worst-case measure. According to Eq. 6.4.20, the problem is formulated in the following form under the Kullback-Leibler divergence.

$$\begin{aligned} \min_a \quad & a^T \Sigma_{KL} a - \lambda a^T \mu_{KL} \\ &= a^T (I - 2\theta \Sigma A)^{-1} \Sigma a - \lambda a^T (\mu - \lambda \theta (I - 2\theta \Sigma A)^{-1} \Sigma a) \\ &= a^T \Sigma a - \lambda a^T \mu + \theta a^T \Sigma a (\lambda^2 + 2a^T \Sigma a) + O(\theta^2) \end{aligned} \quad (6.6.67)$$

Note that in the last equality we apply the Taylor expansion $(I - 2\theta \Sigma A)^{-1} = I + 2\theta \Sigma A + 4\theta^2 \Sigma A \Sigma A + \dots = I + 2\theta \Sigma A + O(\theta^2)$. To find out a closed-form solution, we need to ignore the higher order terms $O(\theta^2)$. Then the stationary condition of the minimisation problem is given by a non-linear equation:

$$2\Sigma a - \lambda \mu + 2\theta (\lambda^2 + 4a^T \Sigma a) \Sigma a = 0 \quad (6.6.68)$$

Notice that

$$a^* := \frac{\lambda}{2} \Sigma^{-1} \mu \quad (6.6.69)$$

is the MVO portfolio weight under the reference measure. For the robust MVO portfolio, we may consider its first-order deviation from a^* . To do that, we substitute $a = a^* + \theta b$ into Eq. 6.6.68 allowing us to cancel the term $\lambda \mu$.

$$2\theta \Sigma b + \theta \lambda (\lambda^2 + \lambda^2 \mu^T \Sigma^{-1} \mu) \mu + O(\theta^2) = 0 \quad (6.6.70)$$

By matching the first-order term w.r.t θ , we find the expression for b :

$$b = -\frac{\lambda^3}{2} (1 + \mu^T \Sigma^{-1} \mu) \Sigma^{-1} \mu \quad (6.6.71)$$

Therefore the optimal MVO portfolio under the worst-case scenario is

$$\begin{aligned} a_{KL}^* &= \left(\frac{\lambda}{2} - \frac{\theta \lambda^3}{2} (1 + \mu^T \Sigma^{-1} \mu) \right) \Sigma^{-1} \mu \\ &= c a^* \end{aligned} \quad (6.6.72)$$

where the coefficient c is defined by

$$c := 1 - \theta \lambda^2 (1 + \mu^T \Sigma^{-1} \mu) \quad (6.6.73)$$

The robust MVO portfolio, as a vector a_{KL}^* , is parallel to the normal MVO portfolio a^* . As a result, the robust MVO portfolio does not change the relative weights of component assets. In fact, all the weights are reduced by the same proportion ($c < 1$) to account for model risk. This is, however, inappropriately account for the correlation risk. For example, two highly-correlated assets have extremely high weights in the nominal MVO portfolio. Because of the correlation risk, we would expect the robust MVO portfolio to assign them lower weights relative to other assets.

The Sharpe ratio of the robust MVO portfolio obviously equals the Sharpe ratio under the reference measure, denoted by S ($S = \sqrt{\mu^T \Sigma^{-1} \mu}$). Sometimes we may be interested in the Sharpe ratio under the worst-case measure. This requires us to examine the mean and variance of the robust MVO portfolio given by Eq. 6.6.72. Assuming that we are under the worst-case scenario given by Eq. 6.6.65, the portfolio mean and variance can be obtained by substituting $a_{KL}^* = c \lambda \Sigma^{-1} \mu / 2$:

$$\mu_{KL}^T a_{KL}^* = \left(\mu - \lambda \theta (I - 2\theta \Sigma a_{KL}^* a_{KL}^{*T})^{-1} \Sigma a_{KL}^* \right)^T a_{KL}^* \quad (6.6.74)$$

$$= \frac{\lambda c}{2} \mu^T \Sigma^{-1} \mu - \theta \frac{\lambda^3 c^2}{4} \mu^T \Sigma^{-1} \mu + O(\theta^2) \quad (6.6.75)$$

$$a_{KL}^{*T} \Sigma_{KL} a_{KL}^* = a_{KL}^{*T} (I - 2\theta \Sigma a_{KL}^* a_{KL}^{*T})^{-1} \Sigma a_{KL}^* \quad (6.6.76)$$

$$= \frac{\lambda^2 c^2}{4} \mu^T \Sigma^{-1} \mu + \theta \frac{\lambda^4 c^4}{8} (\mu^T \Sigma^{-1} \mu)^2 + O(\theta^2) \quad (6.6.77)$$

By using the portfolio mean and variance given in Eq. 6.6.74, we may calculate the Sharpe ratio of the robust MVO portfolio (under the worst-case scenario):

$$\begin{aligned} S_{KL} &= \frac{\mu_{KL}^T a_{KL}^*}{\sqrt{a_{KL}^{*T} \Sigma_{KL} a_{KL}^*}} \\ &= \frac{2 - \theta \lambda^2 c + O(\theta^2)}{2 + \frac{\theta}{2} \lambda^2 c^2 \mu^T \Sigma^{-1} \mu + O(\theta^2)} \sqrt{\mu^T \Sigma^{-1} \mu} \\ &= \left(1 - \frac{\theta}{4} \lambda^2 c (c S^2 + 2) + O(\theta^2) \right) S \end{aligned} \quad (6.6.78)$$

We can see that the robust Sharpe ratio (defined as the Sharpe ratio of the robust MVO portfolio under the worst-case model) is a function of the nominal Sharpe ratio S . The MVO portfolio corresponds to $c = 1$, suffering from more reduction in Sharpe ratio than the robust MVO portfolio ($c < 1$) under the worst-case measure. This simple relation, however, no longer holds for the Wasserstein approach.

6.6.8 H. Robust MVO Portfolio (Wasserstein approach)

In this section, we will switch to the Wasserstein approach to model risk measurement. We will derive the robust MVO portfolio with the Wasserstein approach. Using Eq. 6.6.66, we may formulate the robust portfolio optimisation problem in the following form:

$$\begin{aligned}
 \min_a a^T \Sigma_W a - \lambda a^T \mu_W \\
 &= a^T (I - \beta B^{-1} A)^{-1} \Sigma (I - \beta A B^{-1})^{-1} a - \lambda a^T \left(\mu - \frac{\lambda}{2} \beta (I - \beta B^{-1} A)^{-1} B^{-1} a \right) \\
 &= a^T \Sigma a - \lambda a^T \mu + \beta \left(2a^T B^{-1} a a^T \Sigma a + \frac{\lambda^2}{2} a^T B^{-1} a \right) + O(\beta^2)
 \end{aligned} \tag{6.6.79}$$

Ignoring the higher order terms, the minimisation problem is solved using

$$2\Sigma a - \lambda\mu + \beta(4a^T B^{-1} a \Sigma a + (4a^T \Sigma a + \lambda^2) B^{-1} a) = 0 \tag{6.6.80}$$

Substituting $a = a^* + \beta b$ into Eq. 6.6.80, we find the expression for the perturbation b by matching the first-order terms of β :

$$b = -\frac{\lambda^3}{4} \left(\mu^T \Sigma^{-1} B^{-1} \Sigma^{-1} \mu + (1 + \mu^T \Sigma^{-1} \mu) \Sigma^{-1} B^{-1} \right) \Sigma^{-1} \mu \tag{6.6.81}$$

Therefore, the Wasserstein approach results in a robust MVO portfolio with weights

$$\begin{aligned}
 a_W^* &= \left(\frac{\lambda}{2} - \frac{\beta\lambda^3}{4} \mu^T \Sigma^{-1} B^{-1} \Sigma^{-1} \mu - \frac{\beta\lambda^3}{4} (1 + \mu^T \Sigma^{-1} \mu) \Sigma^{-1} B^{-1} \right) \Sigma^{-1} \mu \\
 &= ca^* - Da^*
 \end{aligned} \tag{6.6.82}$$

where c is a coefficient while C is a square matrix defined by

$$\begin{aligned}
 c &:= 1 - \frac{\beta\lambda^2}{2} \mu^T \Sigma^{-1} B^{-1} \Sigma^{-1} \mu \\
 D &:= \frac{\beta\lambda^2}{2} (1 + \mu^T \Sigma^{-1} \mu) \Sigma^{-1} B^{-1} := d \Sigma^{-1} B^{-1}
 \end{aligned} \tag{6.6.83}$$

c serves just as the coefficient under the Kullback-Leibler divergence, reducing the portfolio weights by the same fraction. D is a matrix that serves to linearly transform the normal MVO portfolio weights.

Eq. 6.6.82 correctly accounts for the correlation risk. When two assets are highly correlated, Σ is close to be singular. This results in extremely large weights under the normal MVO portfolio. Eq. 6.6.82, on the other hand, not only scales the weights down simultaneously by the coefficient c , but also reduces the relative weights of the highly-correlated assets by the linear map D . To see how the linear map D changes the relative weights, we may re-arrange Eq. 6.6.82 to the following form:

$$a_W^* = \frac{\lambda}{2} (\Sigma(cI - D)^{-1})^{-1} \mu \quad (6.6.84)$$

Therefore, the robust MVO portfolio has the same weights as a normal MVO portfolio with an effective covariance matrix

$$\Sigma^* = \Sigma(cI - d\Sigma^{-1}B^{-1})^{-1} \quad (6.6.85)$$

One can show by induction that $\Sigma v = xv$ (x and v are respectively the eigenvalue and the eigenvector) leads to $\Sigma^n v = xv$ for every integer n . This is to say, x is an eigenvalue of Σ only if x^n is an eigenvalue of Σ^n corresponding to the same eigenvector. As a result, for every eigenvalue $x > 0$ of the positive-definite covariance matrix, there exists a corresponding eigenvalue of the effective covariance matrix (here we only consider the special case where B is the identity matrix I):

$$\begin{aligned} \Sigma^* v &= \Sigma(cI - d\Sigma^{-1})^{-1} v \\ &= \frac{1}{c} \sum_{i=0}^{\infty} \left(\frac{d}{c}\right)^i \Sigma^{1-i} v \\ &= \frac{1}{c} \sum_{i=0}^{\infty} \left(\frac{d}{c}\right)^i x^{1-i} v \\ &= \frac{x}{c - d/x} v \end{aligned} \quad (6.6.86)$$

The corresponding eigenvalue

$$\begin{aligned} x^* &:= \frac{x}{c - d/x} \\ &= x + \frac{\beta\lambda^2}{2} (1 + \mu^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} B^{-1} \Sigma^{-1} \mu) + O(\beta^2) \end{aligned} \quad (6.6.87)$$

Any eigenvalue x close to zero is adjusted according to Eq. 6.6.87, resulting in a corresponding eigenvalue x^* that is at least as large as $\beta\lambda^2/2$. This results in an effective matrix Σ^* that is less "singular" than Σ , and therefore a robust MVO portfolio that accounts for the correlation risk.

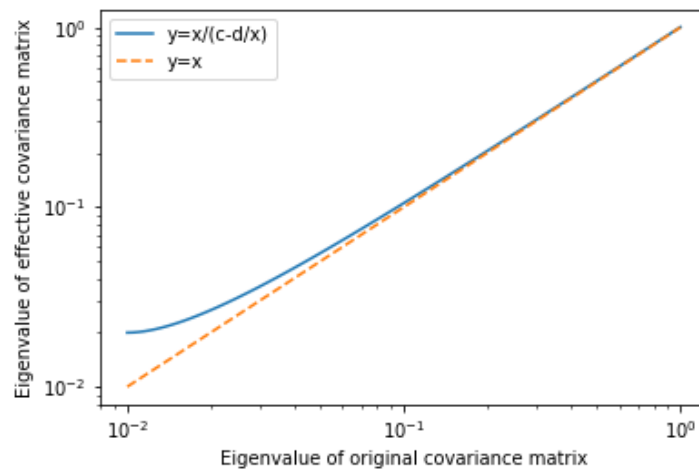


Fig 6.6.1: Eigenvalue x^* of the effective covariance matrix Σ^* increases by a greater amount when the original eigenvalue x gets closer to zero.

Bibliography

- Acerbi, C. and Tasche, D.: 2002, Expected shortfall: a natural coherent alternative to value at risk, *Economic notes* **31**(2), 379–388.
- Ahmadi-Javid, A.: 2012, Entropic value-at-risk: A new coherent risk measure, *Journal of Optimization Theory and Applications* **155**(3), 1105–1123.
- Ait-Sahalia, Y., Karaman, M. and Mancini, L.: n.d., The term structure of variance swaps and risk premia, *Journal of Econometrics* p. forthcoming.
- Ali, S. M. and Silvey, S. D.: 1966, A general class of coefficients of divergence of one distribution from another, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 131–142.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D.: 1999, Coherent measures of risk, *Mathematical finance* **9**(3), 203–228.
- Bakshi, G. and Kapadia, N.: 2003, Delta-hedged gains and the negative market volatility risk premium, *The Review of Financial Studies* **16**(2), 527–566.
- Bally, V., Caramellino, L. and Cont, R.: 2016, Functional Kolmogorov equations, *Stochastic Integration by Parts and Functional Itô Calculus*, Springer, pp. 183–207.
- Bannör, K. F. and Scherer, M.: 2013, Capturing parameter risk with convex risk measures, *European Actuarial Journal* **3**, 97–132.
- Bartl, D., Drapeau, S. and Tangpi, L.: 2018, Computational aspects of robust optimized certainty equivalents and option pricing, *Technical Report 1706.10186*, arXiv preprint.
- Basseville, M.: 2013, Divergence measures for statistical data processing—An annotated bibliography, **93**(4), 621–633.
- Black, F. and Scholes, M.: 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* **81**(3), 637–654.
- Blanchet, J., Chen, L. and Zhou, X. Y.: 2018, Distributionally robust mean–variance portfolio selection with Wasserstein distances, *Technical Report 1802.04885*, arXiv preprint.

- Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency: 2011, Supervisory guidance on model risk management, *Technical Report OCC 2011-12 Attachment*, Federal Reserve.
- Bollerslev, T., Tauchen, G. and Zhou, H.: 2009, Expected stock returns and variance risk premia, *The Review of Financial Studies* **22**(11), 4463–4492.
- Boucher, C. M., Danielsson, J., Kouontchou, P. S. and Maillet, B. B.: 2014, Risk models-at-risk, *Journal of Banking & Finance* **44**, 72–92.
- Box, G. E. P. and Draper, N. R.: 1987, *Empirical Model-Building and Response Surfaces*, Wiley.
- Boyd, S. and Vandenberghe, L.: 2004, *Convex optimization*, Cambridge University Press.
- Branger, N. and Schlag, C.: 2004, Model risk: A conceptual framework for risk measurement and hedging.
- Breeden, D. T. and Litzenberger, R. H.: 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* pp. 621–651.
- Breuer, T. and Csiszár, I.: 2016, Measuring distribution model risk, *Mathematical Finance* **26**(2), 395–411.
- Buchen, P. W. and Kelly, M.: 1996, The maximum entropy distribution of an asset inferred from option prices, *Journal of Financial and Quantitative Analysis* **31**(01), 143–159.
- Carr, P. and Wu, L.: 2016, Analyzing volatility risk and risk premium in option contracts: A new theory, *Journal of Financial Economics* **120**(1), 1–20.
- Cheng, B., Nikitopoulos, C. S. and Schlögl, E.: 2017, Pricing of long-dated commodity derivatives: Do stochastic interest rates matter?, *Journal of Banking & Finance* **95**, 148–166.
- Cont, R.: 2006, Model uncertainty and its impact on the pricing of derivative instruments, *Mathematical Finance* **16**(3), 519–547.
- Cont, R.: 2016, Functional Itô calculus and functional Kolmogorov equations, *Stochastic Integration by Parts and Functional Itô Calculus*, Birkhäuser, Basel, pp. 115–207.
- Cornish, E. A. and Fisher, R. A.: 1938, Moments and cumulants in the specification of distributions, *Revue de l'Institut international de Statistique* pp. 307–320.
- Cover, T. M. and Thomas, J. A.: 2012, *Elements of information theory*, John Wiley & Sons.

- Cox, J. C., Ross, S. A., Rubinstein, M. et al.: 1979, Option pricing: A simplified approach, *Journal of Financial Economics* **7**(3), 229–263.
- Csiszár, I.: 1967, Information-type measures of difference of probability distributions and indirect observations, *Studia Scientiarum Mathematicarum Hungarica* **2**, 299–318.
- Detering, N. and Packham, N.: 2016, Model risk of contingent claims, *Quantitative Finance* **16**(9), 1357–1374.
- Dowd, K.: 2007, *Measuring market risk*, John Wiley & Sons.
- Duffie, D.: 2010, *Dynamic asset pricing theory*, Princeton University Press.
- Feng, Y. and Schlögl, E.: 2018, Model risk measurement under Wasserstein distance, *Technical report*, SSRN Working Paper.
- Fleming, W. H. and Souganidis, P. E.: 1989, On the existence of value functions of two-player, zero-sum stochastic differential games, *Indiana University Mathematics Journal* **38**(2), 293–314.
- Föllmer, H. and Schied, A.: 2002, Convex measures of risk and trading constraints, *Finance and stochastics* **6**(4), 429–447.
- Gatheral, J.: 2006, *The Volatility Surface: A Practitioner's Guide*, John Wiley & Sons.
- Glasserman, P. and Xu, X.: 2014, Robust risk measurement and model risk, *Quantitative Finance* **14**(1), 29–58.
- Gupta, A., Reisinger, C. and Whitley, A.: 2010, Model uncertainty and its impact on derivative pricing, *Rethinking risk measurement and reporting* **122**.
- Hagan, P. S., Kumar, D., Lesniewski, A. S. and Woodward, D. E.: 2002, Managing smile risk, *The Best of Wilmott* p. 249.
- Hansen, L. P. and Sargent, T. J.: 2006, *Robustness*, Princeton University Press.
- Hansen, L. P., Sargent, T. J., Turmuhambetova, G. and Williams, N.: 2006, Robust control and model misspecification, *Journal of Economic Theory* **128**(1), 45–90.
- Hansen, L. and Sargent, T. J.: 2001, Robust control and model uncertainty, *American Economic Review* **91**(2), 60–66.
- Heston, S. L.: 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* **6**, 327–343.

- Hull, J. and Suo, W.: 2002, A methodology for assessing model risk and its application to the implied volatility function model, *Journal of Financial and Quantitative Analysis* **37**(02), 297–318.
- Karatzas, I. and Shreve, S. E.: 1991, *Brownian Motion and Stochastic Calculus*, Springer.
- Kerkhof, J., Melenberg, B. and Schumacher, H.: 2010, Model risk and capital reserves, *Journal of Banking & Finance* **34**, 267–279.
- Kerkhof, J., Schumacher, J. and Melenberg, B.: 2002, Model risk and regulatory capital.
- Lindström, E.: 2010, Implications of parameter uncertainty on option prices, *Advances in Decision Sciences* **2010**.
- Löffler, G.: 2003, The effects of estimation error on measures of portfolio credit risk, *Journal of Banking & Finance* **27**, 1427–1453.
- Low, B. S. and Zhang, S.: 2005, The volatility risk premium embedded in currency options, *Journal of Financial and Quantitative Analysis* **40**(4), 803–832.
- McNeil, A. J., Frey, R. and Embrechts, P.: 2015, *Quantitative risk management: Concepts, techniques and tools*, Princeton university press.
- Merton, R. C.: 1973, Theory of rational option pricing, *The Bell Journal of Economics and Management Science* **4**(1), 141–183.
- Meyer, P. A.: 1966, *Probability and potentials*, Vol. 1318, Blaisdell Pub. Co.
- Morini, M.: 2011, *Understanding and Managing Model Risk: A practical guide for quants, traders and validators*, John Wiley & Sons.
- Platen, E. and Heath, D.: 2006, *A benchmark approach to quantitative finance*, Springer Science & Business Media.
- Revuz, D. and Yor, M.: 2013, *Continuous martingales and Brownian motion*, Vol. 3.8, Springer Science & Business Media.
- Rockafellar, R. T. and Uryasev, S.: 2002, Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* **26**(7), 1443–1471.
- Schlogl, E.: 2013, *Quantitative Finance: An Object-Oriented Approach in C++*, CRC Press.
- Schlögl, E.: 2015, Toward quantifying model risk, *Presentation Slides* .
- Slater, M.: 2014, Lagrange multipliers revisited, *Traces and Emergence of Nonlinear Programming*, Springer, pp. 293–306.
- Villani, C.: 2008, *Optimal transport - Old and new*, Vol. 338.