Sets of Indistinguishable Models for Robust Optimisation

Anne Balter*
Maastricht University†, Netspar

Antoon Pelsser‡
Maastricht University, Netspar, Kleynen Consultants

First version: December 15, 2013
This version: September 2, 2015

Abstract

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

Keywords: Model Uncertainty, Robust Stochastic Control

*Phone: +31433884962. Email: `a.balter@maastrichtuniversity.nl`
†Tongersestraat 53, 6211LM, Maastricht, The Netherlands
‡Phone: +31433883899. Email: `a.pelsser@maastrichtuniversity.nl`

1 Introduction

Models can be wrong and recognising their limitations is important in financial and economic decision making. In asset pricing, model uncertainty has implications for the valuation of derivatives and long-dated contracts. In such situations the value of an asset is not unique but falls in a range, requiring extensions of standard pricing methods. We develop a method that provides a credible set of models surrounding the baseline model to use in robust decision making. Making financial decisions robust calls for applying worst case scenarios. Therefore even if an optimisation problem is considered under the allowance of model uncertainty, still the question remains how to determine the set of alternative models. The choice of the uncertainty set is what we will focus on, contrary to most literature that solves a robust control problem for a given set of alternatives. In the second review of Chapter 1 we discussed other methods in literature that cope with uncertainty. The shared intuition is that an agent is concerned about model misspecification, consequently the robust control decisions are based on the worst case scenario. This is obtained by evaluating the control function for every plausible model. Intuitively, these models come from a set surrounding the baseline model. Our objective is to determine explicitly the set that incorporates all plausible alternative models using statistical testing theory. We use the Neyman-Pearson Lemma and impose a Type I and II error to construct the set of indistinguishable models.

1.1 Literature Review

Hansen and Sargent (2008) (hereafter H&S) motivate their approach of uncertainty to robust optimisation by choosing models surrounding the baseline model with bounded entropy (i.e. with bounded Kullback-Leibler divergence). However, when implementing their method they make a subtle switch: they replace the endogenous Lagrange multiplier of the entropy bound by a fixed entropy penalty. As a consequence of this switch H&S obtain a time-consistent operator, which is a very desirable property for dynamic optimisation problems. But the explicit connection with the size of the set is lost, and it is unclear how to choose the fixed penalty parameter. Breuer and Csiszár (2013) base stress tests on a similar setting of a robust optimisation problem with a Kullback-Leibler bound. Similarly, also Glasserman and Xu (2014) use a Kullback-Leibler penalty.

H&S start with an optimisation problem, then they pick a Lagrange multiplier and calculate the worst case path which depends on the multiplier and the specific optimisation problem. Next they calculate the Type I and II errors for the specific multiplier and the associated worst case path. If the probability of the average of the two incorrect rejections is too high, then the worst case choice from mother nature is too extreme, with other words too far from the approximating model. Therefore it is deemed unlikely that these two models cannot be distinguished from each other. Hence this multiplier is rejected and will not belong to the set of alternatives. By this procedure one plausible worst-case based multiplier is selected rather than a set. The main difference with our approach is the order of the procedure. H&S start with an explicit optimisation problem whereas we focus on the creation of the set of plausible alternative models that can be applied to and is independent of the particular choice of the optimisation problem.

Ben-Tal et al. (2013) focus on ϕ-divergences in robust optimisation. General optimisation problems, examples ranging from finance to operations research, are solved robustly over an uncertainty region $\mathscr{U}$. The uncertainty region is identified by the confidence set using a specific ϕ-divergence functions. The described distance measures are Kullback-Leibler, Burg Entropy, J-divergence, χ^2-distance, Hellinger distance, Variation distance and Cressie-Read. The method proposed by Ben-Tal et al. is a procedure how a set could be created with multiple possibilities, such as the choice of the divergence function and the bounds on these, rather than they specify the characterisation of the set of indistinguishable models. In section 4 we will quantify the bounds on these different measures.

Hansen et al. (2011) introduce the concept of Model Confidence Set (MCS). This method, or actually algorithm, is a sequential method that starts with a collection of possible models and ends with a subset of these that contains the best models with a given level of confidence. Best is in terms of a chosen test statistic, this optionality highlights the first difficulty in this method. Moreover, in comparison with the indistinguishable method discussed in this paper, the main difference is that the MCS method start with a collection of competing models $\mathscr{M}^0$. That is why the set of alternatives has to be selected upfront, whereas no such decisions have to be made ex ante during the indistinguishable method. Therefore all models are considered in the indistinguishable method rather than a discrete number of models. Contrary to the MCS method that considers a list of models which will be removed from the set of plausible ones by judgement of the data, we would like to consider a class of models as large as possible. The standard parameter approach where a confidence interval is constructed considers the smallest class of models as it is restricted to constant parameters only. The MCS method is already fundamentally larger as it can contain also deterministic and stochastic models, while we do not even need to push the agent to introduce a list of concrete models.

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

1.2 Intuition

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

Suppose we have an optimisation problem which extends over the time interval $[0, T]$. With other words, suppose there would be T years of extra information available. Our baseline model is specified on a filtered probability space
$\left(\Omega, \{\mathscr{F}_t\}_{0\leq t\leq T}, \mathbb{P}\right)$, where $\mathbb{P}$ denotes the probability measure that corresponds to our baseline model. The idea is that we define the credible set of alternative models, as those models that cannot be distinguished statistically from the baseline model if one would take the

observations accumulated over $[0, T]$ into consideration. In other words, we want to exclude ex ante (at time 0) those models which could possibly be rejected by a statistical test procedure at time T with a reasonable level of confidence. We want to use the most powerful tests possible, which are likelihood ratio tests, as stated by the Neyman-Pearson Lemma.

2 Statistically Indistinguishable Models

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

$$dX(t) = \mu(t,\omega)\,dt + \sigma(t,\omega)\,dW(t) \tag{2.1}$$

For the specification of possible alternative models, we consider Brownian Motion with a (stochastic) drift process $dW(t) + \lambda(t,\omega)\,dt$. Such an alternative model specification of the Brownian Motion can be captured as a change in probability measure from $\mathbb{P}$ to a new probability measure $\mathbb{Q}$. With slight abuse of notation we will denote both the alternative model and the alternative probability measure by $\mathbb{Q}$.

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

$$dR(t) = \lambda(t,\omega)R(t)\,dW^{\mathbb{P}}(t) \tag{2.2}$$

The superscript $\mathbb{P}$ denotes the probability measure we are considering. The solution to the stochastic differential equation (2.2) can be represented as

$$R(T) = \exp\left\{-\tfrac{1}{2}\int_0^T \lambda(t,\omega)^2\,dt + \int_0^T \lambda(t,\omega)\,dW^{\mathbb{P}}(t,\omega)\right\} \tag{2.3}$$

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

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

$$\mathbb{P}[R(T) \ge \gamma] = \alpha \tag{2.4}$$

We set the critical value γ such that probability of incorrectly rejecting model $\mathbb{P}$ when model $\mathbb{P}$ is the true model is equal to α. This is known as the Type I error. The probability α is the significance level of the test, and is typically set at 0.05. [1]

[1] Inclusion/exclusion of the equality sign in the expression $R(T) \ge \gamma$ is potentially relevant when there are point-masses in the probability distribution of $R(T)$. For ease of exposition, we assume this is not the case. When we do have point-masses we can still handle this mathematically, but then we must consider randomised tests.

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

$$\mathbb{Q}[R(T) < \gamma] = \beta \tag{2.5}$$

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

$$\mathbb{Q}[R(T) \geq \gamma] = 1 - \beta \tag{2.6}$$

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

The Type I and Type II probabilities can be computed ex ante at time 0 for a given alternative model $\mathbb{Q}$. The model selection procedure we propose is based on the Type II error of the likelihood ratio test. The intuition is a follows. For small values of $\lambda(t,\omega)$ we will have a model $\mathbb{Q}$ that is "close" to the baseline model $\mathbb{P}$. This closeness can be identified by the fact that the likelihood ratio $R(T)$ will be a random variable with a probability distribution tightly concentrated around the value $R(T) = 1$. Even though we can define a critical value γ for any model $\mathbb{Q}$, for models that are "close" there will be almost no difference between the $\mathbb{P}$-probability and the $\mathbb{Q}$-probability of the event $R(T) \geq \gamma$. Hence, the power of the statistical test will be very low. In the limiting case when $\mathbb{P} = \mathbb{Q}$ the power of the (randomised) likelihood ratio test will be as low as α.

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

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

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

2.1 Deterministic Drift Term

When Mother Nature is only allowed to use alternative models with a deterministic drift term $\lambda(t)$, we can compute the probability distribution of the Radon-Nikodym derivative defined in equation (2.2) explicitly. For this case we obtain

$$R(T) = \exp\left\{-\tfrac{1}{2}\int_0^T \lambda(t)^2\,dt + \int_0^T \lambda(t)\,dW^{\mathbb{P}}(t)\right\} \tag{2.7}$$

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

$$r(T) = \int_0^T \lambda(t)\,dW^{\mathbb{P}}(t) \tag{2.8}$$

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

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

$$\mathbb{Q}[r(T) > \gamma] = \Phi\left(\Phi^{-1}(\alpha) + \left(\int_0^T \lambda(t)^2\,dt\right)^{\frac{1}{2}}\right) \tag{2.9}$$

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

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

The deterministic example we have formulated can be generalised easily to the *multi-dimensional case*. For a vector-valued Brownian Motion all alternative models are specified by the deterministic vector-valued process $\lambda(t)$. The test statistic $r(t)$ is then given by

$$r(T) = \int_0^T \lambda(t) \cdot dW^{\mathbb{P}}(t) \tag{2.10}$$

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

2.2 Stochastic Drift Term

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

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

$$R(T) := e^{-\frac{1}{2}a^2 T}\cosh(aW(T)) = \tfrac{1}{2}\left(e^{-\frac{1}{2}a^2 T + aW(T)} + e^{-\frac{1}{2}a^2 T - aW(T)}\right) \tag{2.11}$$

which is strictly positive and has expectation $\mathbb{E}[R(T)] = 1$, hence this is a valid Radon-Nikodym derivative. This $R(T)$ corresponds to a $\mathbb{Q}$-model where the probability distribution of $W(T)$ at time T is given by a mixture distribution of two normal distributions with mean $+aT$ and $-aT$, the same variance T, and mixing probabilities $\frac{1}{2}$. Note that this mixture distribution is not a normal distribution, and has mean 0 and variance equal to $T + (aT)^2$ (see Appendix A), which is larger than the variance T under the $\mathbb{P}$-model.

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

2.3 General Case

The deterministic case and the stochastic hyperbolic cosine served as an illustration of our approach. However, we are interested in the generalisation of this method. With other words, we would like to allow for a wide class of alternative models. As such we have shown in the previous subsection that stochastic deviations can lead to fundamentally different models. Therefore we would like to generate the set of stochastic alternative models surrounding the baseline model that are indistinguishable based on an insufficient power and sufficient size.

For general $\lambda(t, \omega)$ the power calculation is based on (2.3) from which the distribution is unknown.

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

Because of the difficulty of the power calculation for general $\lambda(t, \omega)$, we impose time-consistency on the set of indistinguishable models (with insufficient power). We search for the distribution of the Radon-Nikodym derivative that generates the maximum power possible in Section 3. All the models that imply a power lower than a maximum of $1 - \beta$ are defined to be indistinguishable.

3 Time Consistency

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

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

The set of indistinguishable models we have defined thus far is *not* time-consistent: the set is defined as those models that have sufficiently low power at time T using the information over the whole path $[0, T]$. We have established in Section 2 that the set of indistinguishable models defines a coherent risk measure. But, this risk measure is "static" at time 0 and not time-consistent.

We can however look at a smaller class of risk measures: the class of time-consistent risk (dynamic) measures. This class has been extensively studied in recent years, and we know how to characterise this class of risk measures. Delbaen (2006) proves that time-consistent (coherent and convex) risk measures are generated by m-stable sets of probability measures. A similar structure (albeit with less mathematical rigour) was already proposed by Epstein and Schneider (2003). An alternative characterisation of time-consistent risk measures has been provided by Rosazza Gianin (2006). She proves that every time-consistent risk measure is equivalent to a g-expectation $\mathcal{E}^g[]$. These non-linear g-expectations can be computed as the solution of a backward stochastic differential equation (BSDE) with a driver $g(t,Y,Z)$. A further characterisation has been provided by Barrieu and El Karoui (2007): they prove that time-consistent coherent risk measures are generated by drivers $g(t,Z)$ that satisfy a Lipschitz growth constraint in Z, and time-consistent convex risk measures are generated by drivers that satisfy a quadratic growth constraint in Z.

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

3.1 Maximum Power Calculation

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

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

$$\begin{aligned} \max_{\gamma,|\lambda(t,\omega)|\le k} \; & \mathbb{E}\left[R(T)\mathbb{1}(R(T)\ge\gamma)\right] \\ \text{s.t. } & \mathbb{E}\left[\mathbb{1}(R(T)\ge\gamma)\right]=\alpha \\ & dR=\lambda(t,\omega)R\,dW, R_0=1 \end{aligned} \tag{MP}$$

where γ is defined by $\mathbb{E}\left[\mathbb{1}(R(T)\ge\gamma)\right]=\alpha$.

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

$$\begin{aligned} F_\alpha(R,\gamma) &= \alpha\gamma+\mathbb{E}\left[(R-\gamma)^+\right] \\ &= \alpha\gamma+\mathbb{E}\left[R\mathbb{1}(R(T)\ge\gamma)\right]-\gamma\mathbb{E}\left[\mathbb{1}(R(T)\ge\gamma)\right] \\ &= \mathbb{E}\left[R\mathbb{1}(R(T)\ge\gamma)\right]+\left(\alpha-\mathbb{E}\left[\mathbb{1}(R(T)\ge\gamma)\right]\right)\gamma \end{aligned} \tag{3.1}$$

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

$$\max_{|\lambda(t,\omega)|\leq k} \min_\gamma F_\alpha(R,\gamma) \quad \text{s.t. } dR = \lambda(t,\omega) R dW \tag{MaMi}$$

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

$$\begin{aligned} \frac{\partial F_\alpha(R,\gamma)}{\partial \gamma} &= \alpha - \mathbb{E}\left[\mathbb{1}(R(T) \geq c)\right] \\ &= 0 \end{aligned} \tag{3.2}$$

This implies the Type I error constraint to hold. Since $F_\alpha(R,\gamma)$ is convex in γ, the extreme value is a minimum. In particular, then, we have

$$\begin{aligned} \min_\gamma F_\alpha(R,\gamma) &= \min_\gamma \mathbb{E}\left[R\mathbb{1}(R(T) \geq \gamma)\right] + (\alpha - \alpha)\gamma \\ &= \mathbb{E}\left[R(T)\mathbb{1}(R(T) \geq \gamma)\right] \end{aligned} \tag{3.3}$$

We maximise $F_\alpha(R,\gamma)$ because the objective corresponds with

$$\begin{aligned} \mathbb{E}\left[R(T)\mathbb{1}(R(T) \geq \gamma)\right] &= \min_\gamma \mathbb{E}\left[R\mathbb{1}(R(T) \geq \gamma)\right] + \left(\alpha - \mathbb{E}\left[\mathbb{1}(R(T) \geq \gamma)\right]\right)\gamma \\ &= \min_\gamma F_\alpha(R,\gamma) \end{aligned} \tag{3.4}$$

Therefore solving the initial non-convex optimisation problem is identical with solving the max min. The power from the (MaMi) is always lower than or equal to the reversed order optimisation by the max min inequality.

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

We can summarise the relations between the different optimisation formulations by

$$(\text{MP}) \overbrace{=}^{\text{CVaR}} (\text{MaMi}) \underbrace{\overbrace{\leq}^{\text{Max Min}}}_{\text{Inequality}} (\text{MiMa}) \tag{3.5}$$

First we solve the (MiMa), in specific we start with the inner maximisation problem

$$\max_{|\lambda(t,\omega)|\leq k} F_\alpha(R,\gamma) \quad \text{s.t. } dR = \lambda(t,\omega) R dW \tag{3.6}$$

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

To maximise $V(t,r,\gamma)$, for any value of α and γ the optimised V function w.r.t. λ for $t \le T$ solves the HJB-equation

$$V_t + \max_{|\lambda(t,r)|\le k} \tfrac{1}{2}\lambda(t,r)^2 r^2 V_{rr} = 0. \tag{3.7}$$

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

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

$$V(t,r,\gamma) \quad = \quad \alpha\gamma + rN(d_1) - \gamma N(d_2) \tag{3.8}$$

where $d_1 = \frac{1}{k\sqrt{T-t}}\left(\ln\left(\frac{r}{\gamma}\right) + \frac{1}{2}k^2(T-t)\right)$ and $d_2 = d_1 - k\sqrt{T-t}$. This function solves the HJB equation for the boundary condition. Moreover,

$$\left|\lambda^*(t,r)\right| = \begin{cases} k & \text{for } V_{rr}(t,r,\gamma) \ge 0 \\ 0 & \text{for } V_{rr}(t,r,\gamma) < 0 \end{cases} \tag{3.9}$$

The value function is convex as

$$V_{rr}(t,r,\gamma) \quad = \quad \frac{n(d_1)}{rk\sqrt{T-t}} \ge 0 \ \forall\ r, t < T \tag{3.10}$$

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

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

The optimal value function is $V(0,1,\gamma) = \mathbb{E}\left[F_\alpha(R(T),\gamma)|\mathscr{F}_0\right] = \alpha\gamma +$ $\mathbb{E}\left[(R(T)-\gamma)^+|t=0,R_0=1\right]$, hence this is the optimal power and can be interpreted as a log-normal expectation equal to a constant plus the expectated value of a Black-Scholes call option.

The remaining part is the outer minimisation. In specific,

$$\min_\gamma V(0,1,\gamma) \tag{3.11}$$

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

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

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

$$\mathbb{E}\left[R^*(T)\mathbb{1}\big(R^*(T)>\gamma^*\big)\right] = \mathbb{Q}\left[R^*(T)>\gamma^*\right] = \Phi\Big(\Phi^{-1}(\alpha)+k\sqrt{T}\Big). \tag{3.12}$$

If we want to construct a class of time-consistent coherent risk measures with stochastic $\lambda(t,\omega)$ such that every element in the class is statistically indistinguishable, then a sufficient condition is to choose a k such that the "worst case" power (3.12) does not exceed $(1-\beta)$. For example, if we take $\alpha = 0.05$ and $\beta = 0.20$, we find $k = 2.48/\sqrt{T}$.

All previous derivations also hold for vector Brownian Motions and vector $\lambda(t,\omega)$, thus the L2-norm $|\lambda(t,\omega)| \leq \frac{2.48}{\sqrt{T}}$.

The stochastic example of section 2.2 illustrates that the bound on the norm of the stochastic deviations is a sufficient condition to determine the set of indistinguishable models. For hyperbolic cosine as choice for $R(T)$ we can explicitly compute the conditional expectation $R(t) = \mathbb{E}^{\mathbb{P}}[R(T)|\mathscr{F}_t] = e^{-\frac{1}{2}a^2t}\cosh(aW(t))$. If we apply Itô's Lemma to $R(t)$ we obtain the stochastic differential equation

$$dR(t) = ae^{-\frac{1}{2}a^2t}\sinh(aW(t))dW(t) = a\tanh\big(aW(t)\big)R(t)dW(t). \tag{3.13}$$

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

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

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

$$\mathbb{Q}\left[|W(T)/\sqrt{T}|>\gamma'\right] = \Phi\big(\Phi^{-1}(\alpha/2)+a\sqrt{T}\big)+\Phi\big(\Phi^{-1}(\alpha/2)-a\sqrt{T}\big) \tag{3.14}$$

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

4 Bounds on Distance Measures

Ben-Tal et al. (2013) discuss several possible distance measures to generate robust results in optimisation problems. ϕ-Divergence (or f-divergence) functions measure the distance between two probability distributions weighted by the specific function. The choice which measure should be picked is an unanswered issue in their and many other papers, plus the question when the distance is too far is unanswered. In this chapter we explicitly focus on the size of the set of alternatives. We can directly link this to a critical value for each measure. Note that Ben-Tal et al. (2013) consider the discrete versions, (4.1), whereas we consider the continuous distance measures, (4.2) and (4.3), as a function of the Radon-Nikodym derivatives. The numerical value for the set of time-consistent models that cannot be distinguished between with a Type I error of 5% and a power less than 80% is displayed in the last column of Table 1. The derivations can be found in the Appendix C.

The general discrete ϕ-divergence is defined as

$$D_\phi(p,q) = \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right) \tag{4.1}$$

We use the continuous version

$$D_\phi(p,q) = \mathbb{E}^{\mathbb{Q}}\left[\phi\left(\frac{1}{R(T)}\right)\right] \tag{4.2}$$
$$= \mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] \tag{4.3}$$

where the functions for $\phi(\cdot)$ are given for each measure and by definition $\phi(\cdot)$ is convex.

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

Table 1: Distance measures

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

Both the χ-divergence of order $\theta > 1$ and the Cressie-Read can only be solved numerically. The Cressie-Read for θ and k can analytically be expressed by $\frac{1}{\theta(1-\theta)}\left(2-\theta-e^{-\frac{1}{2}k^2\theta(1-\theta)T}\right)$. For $k\sqrt{T} = 2.48$ we get the bounds on both measures displayed in Table 2.

Table 2: Numerical bounds

Divergence θ	1.5	2.0	2.5	3.0
χ-divergence of order θ	10.86	563.64	9.3232×10^4	2.3657×10^7
Cressie-Read	12.72	234.45	2.7180×10^4	1.7183×10^7

5 Conclusion

To conclude, by imposing probabilities on the Type I and II error of the Radon-Nikodym derivative we are able to quantify uncertainty in an intuitive way. Hence if an agent acknowledges that his model might be misspecified, he would like to evaluate the optimal decision rule against all plausible alternative models to incorporate robustness. Applications that build upon the uncertainty need the specific set of indistinguishable models. Examples can be found on a wide range to price and hedge in incomplete markets, for instance long-dated insurance contracts or illiquid assets.

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

A Variance Hyperbolic Cosine

$$\begin{aligned} H_0 : W(t) \quad &\text{versus} \quad H_A : W(t) + \int_0^t \lambda(s,\omega)ds \\ H_0 : W(T) \quad &\text{versus} \quad H_A : W(T) + \int_0^T \lambda(s,\omega)ds \end{aligned} \tag{A.1}$$

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

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

And under $\mathbb{Q}$

$$\begin{aligned} M_{\mathbb{Q}}(t) &= \mathbb{E}^{\mathbb{P}}\left[e^{tW(T)}R(T)\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[e^{tW(T)}\tfrac{1}{2}\left(e^{-\frac{1}{2}a^2T+aW(T)} + e^{-\frac{1}{2}a^2T-aW(T)}\right)\right] \\ &= \tfrac{1}{2}\left(e^{-\frac{1}{2}a^2T+(a+t)^2\frac{1}{2}T} + e^{-\frac{1}{2}a^2T+(-a+t)^2\frac{1}{2}T}\right) \\ \mathbb{E}[W(T)] = \frac{\partial M}{\partial t}(0) &= \tfrac{1}{2}\left((a+t)Te^{-\frac{1}{2}a^2T+(a+t)^2\frac{1}{2}T}\right) + \\ &\quad \tfrac{1}{2}\left((-a+t)Te^{-\frac{1}{2}a^2T+(-a+t)^2\frac{1}{2}T}\right) \\ &= 0 \\ \mathbb{E}\left[W(T)^2\right] = \frac{\partial^2 M}{\partial t^2}(0) &= \tfrac{1}{2}\left((a(a+t)T^2+T)e^{-\frac{1}{2}a^2T+(a+t)^2\frac{1}{2}T}\right) + \\ &\quad \tfrac{1}{2}\left((-a(-a+t)T^2+T)e^{-\frac{1}{2}a^2T+(-a+t)^2\frac{1}{2}T}\right) \\ &= (aT)^2 + T \end{aligned} \tag{A.3}$$

B Power Hyperbolic Cosine

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

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

Since $\lambda(t,\omega) = a\tanh(aW(t))$ and $|a\sqrt{T}| < 2.80 \Rightarrow |a| < 2.80/\sqrt{T}$ this can be one-to-one compared with $|\lambda\sqrt{T}| < 2.48 \Rightarrow |\lambda| < 2.80/\sqrt{T} \Rightarrow |a\tanh(aW(t))| < 2.80/\sqrt{T}$. Because $-a < a\tanh(aW(t)) < a$.

C Distance Measures

C.1 Kullback-Leibler

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

$$D\left(\mathbb{P}\|\mathbb{Q}\right) = \mathbb{E}^{\mathbb{P}}[-\ln R(T)] = \mathbb{E}^{\mathbb{Q}}\left[-\frac{\ln R(T)}{R(T)}\right] \tag{C.1}$$

We can arrive at a similar conclusion for a maximum entropy (= max K-L) calculation. We want to investigate the maximum entropy that can be achieved within the class of Radon-Nikodym derivatives with $|\lambda(t,\omega)| \le k$. We can formulate this as a stochastic optimisation problem of the form

$$\max_{|\lambda(t,\omega)|\le k} \mathbb{E}[-\ln R(T)] \tag{C.2}$$

$$\text{s.t. } dR = \lambda(t,\omega)R\,dW$$

$$R > 0 \tag{C.3}$$

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

$$V_t + \max_{|\lambda(t,r)|\le k} \tfrac{1}{2}\lambda(t,r)^2 r^2 V_{rr} = 0 \tag{C.4}$$

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

$$V(t,r) \quad = \quad \tfrac{1}{2}k^2 T \tag{C.5}$$

Similarly as in the maximum power calculation in section 3.1, the value function $V(t,r)$ is uniformly integrable and thus the optimal control is $|\lambda(t,\omega)| = k$. We arrive at the conclusion that the log-normal Radon-Nikodym derivative with $|\lambda(t,\omega)| \equiv k$ for all $0 \le t \le T$ achieves the maximal entropy of $V(0,1) = \frac{1}{2}k^2 T$. Hence, we can alternatively characterise the class of time-consistent indistinguishable models with $|\lambda(t,\omega)| \le 2.48/\sqrt{T}$, which implies a maximum attainable entropy of $\frac{1}{2}(2.48)^2 = 3.08$.

C.2 Burg Entropy

Burg Entropy or also called Minimum Discrimination Information is defined as

$$D\left(\mathbb{Q}\|\mathbb{P}\right) = \mathbb{E}^{\mathbb{Q}}[\ln R(T)] = \mathbb{E}^{\mathbb{P}}[R(T)\ln R(T)] \tag{C.6}$$

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

$$\max_{|\lambda(t,\omega)|\le k} \mathbb{E}[R(T)\ln R(T)] \tag{C.7}$$
$$\text{s.t. } dR = \lambda(t,\omega) R dW$$

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

$$V_t + \max_{|\lambda(t,R)|\le k} \tfrac{1}{2}\lambda(t,r)^2 r^2 V_{rr} = 0 \tag{C.8}$$

As the terminal condition $V(T,R) = R\ln R$ is a strictly convex function in R, we arrive at the conclusion that the log-normal Radon-Nikodym derivative with $|\lambda(t,\omega)| \equiv k$ for all $0 \le t \le T$ achieves the maximal Burg entropy of $V(0,1) = \frac{1}{2}k^2 T$. Hence, we can alternatively characterise the class of time-consistent indistinguishable models with $|\lambda(t,\omega)| \le 2.48/\sqrt{T}$, which implies a maximum attainable Burg entropy of $\frac{1}{2}(2.48)^2 = 3.08$.

C.3 J-Divergence

Jeffreys (1946) J-divergence is $D\left(\mathbb{P}\|\mathbb{Q}\right)+D\left(\mathbb{Q}\|\mathbb{P}\right)$ which equals $\mathbb{E}^{\mathbb{P}}[-\ln R(T)]+\mathbb{E}^{\mathbb{P}}[R(T)\ln R(T)]$. We can formulate this as a stochastic optimisation problem of the form

$$\max_{|\lambda(t,\omega)|\le k} \mathbb{E}[-\ln R(T)] + \mathbb{E}[R(T)\ln R(T)] \tag{C.9}$$
$$\text{s.t. } dR = \lambda(t,\omega) R dW$$
$$R > 0 \tag{C.10}$$

Again the associated value function is convex in R and hence the maximum J-divergence is obtained for $k^2 T = (2.48)^2 = 6.15$.

C.4 χ^2-Distance

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

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= \mathbb{E}^{\mathbb{P}}\left[R(T)^2 - 2R(T) + 1\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[R(T)^2\right] - 2\mathbb{E}^{\mathbb{P}}\left[R(T)\right] + 1 \\ &= \mathbb{E}^{\mathbb{P}}\left[R(T)^2\right] - 2 + 1 \end{aligned} \tag{C.11}$$

For $|\lambda(t,\omega)| = k$ this boils down to

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)|\mathscr{F}_0\right] &= e^{k^2T} - 2 + 1 \\ &= e^{2.48^2} - 1 = 467.90 \end{aligned} \tag{C.12}$$

C.5 Modified χ^2-Distance

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= \mathbb{E}^{\mathbb{P}}\left[\frac{1}{R(T)} - 2 + R(T)\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[\frac{1}{R(T)}\right] - 2 + 1 \end{aligned} \tag{C.13}$$

For the log-normal Radon-Nikodym derivative the modified χ^2-divergence bounded by

$$\mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] = e^{k^2T} - 1 = 467.90 \tag{C.14}$$

represents the set of indistinguishable models for a probability of 5% for a Type I error occurring and all models yield a power of at most 80%. The derivation is obtained by applying Ito's Lemma

$$\begin{aligned} d\left(\frac{1}{R(t)}\right) &= \frac{1}{R(0)}\left(k^2dt - kW^{\mathbb{P}}(t)\right) \\ \frac{1}{R(t)} &= \frac{1}{R(t)}e^{\frac{1}{2}k^2t - kW^{\mathbb{P}}(t)} \end{aligned} \tag{C.15}$$

Since $R(0) = 1$ we get

$$\mathbb{E}^{\mathbb{P}}\left[\frac{1}{R(T)}\right] = e^{k^2T} \tag{C.16}$$

C.6 Hellinger Distance

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= \mathbb{E}^{\mathbb{P}}\left[1 - 2\sqrt{R(T)} + R(T)\right] \\ &= -2\mathbb{E}^{\mathbb{P}}\left[\sqrt{R(T)}\right] + 2 \end{aligned} \tag{C.17}$$

The optimal control solves for

$$\mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] = -2e^{-\frac{1}{8}k^2T} + 2 = 1.07 \tag{C.18}$$

C.7 Variation Distance

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= \mathbb{E}^{\mathbb{P}}\left[\left|\frac{1}{R(T)}-1\right|R(T)\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\max\left(\frac{1}{R(T)}-1,0\right)R(T)\right]+ \\
&\quad \mathbb{E}^{\mathbb{P}}\left[\max\left(1-\frac{1}{R(T)},0\right)R(T)\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\max\left(\frac{1}{R(T)}-1,0\right)\right]+\mathbb{E}^{\mathbb{Q}}\left[\max\left(1-\frac{1}{R(T)},0\right)\right]
\end{aligned} \tag{C.19}$$

The stochastic differential equation of $\frac{1}{R(T)}$ under $\mathbb{Q}$ is obtained by applying Ito's Lemma

$$\begin{aligned}
d\frac{1}{R(t)} &= -\frac{1}{R(t)}kdW^{\mathbb{Q}}(t) \\
\frac{1}{R(T)} &= e^{-\frac{1}{2}k^2T-kW^{\mathbb{Q}}(T)}
\end{aligned} \tag{C.20}$$

The maximum variation distance can be decomposed in a call and put option where the underlying stock process is $d\frac{1}{R(t)}$, the strike price is 1 and the risk-free rate is 0. Then we get for (C.19)

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= C\left(\frac{1}{R(0)},0\right)+P\left(\frac{1}{R(0)},0\right) \\
&= 4N(-d1)-2 \\
&= 1.57
\end{aligned} \tag{C.21}$$

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

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= \mathbb{E}^{\mathbb{P}}\left[\left|\frac{1}{R(T)}-1\right|^{\theta}R(T)\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\left|\frac{1}{R(T)}-1\right|^{\theta}\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\max\left(\frac{1}{R(T)}-1,0\right)^{\theta}\right]+ \\
&\quad \mathbb{E}^{\mathbb{Q}}\left[\max\left(1-\frac{1}{R(T)},0\right)^{\theta}\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\max\left(\left(\frac{1}{R(T)}-1\right)^{\theta},0\right)\right]+ \\
&\quad \mathbb{E}^{\mathbb{Q}}\left[\max\left(\left(1-\frac{1}{R(T)}\right)^{\theta},0\right)\right]
\end{aligned} \tag{C.22}$$

This can only be solved numerically, e.g. for different values of $\theta = \{1.5, 2, 2.5, 3\}$ and $k^2T = 2.48^2$. Then we get

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\left|\frac{1}{R(T)}-1\right|^2 R(T)\right] &= \mathbb{E}^{\mathbb{Q}}\left[\max\left(\left(\frac{1}{R(T)}-1\right)^2, 0\right)\right] + \\
&\quad \mathbb{E}^{\mathbb{Q}}\left[\max\left(\left(1-\frac{1}{R(T)}\right)^2, 0\right)\right] \\
\frac{1}{R(T)} &= e^{-3.08-2.48/\sqrt{T}W^{\mathbb{Q}}(T)}
\end{aligned}$$

Draw 100,000,000 values z_i from a standard normal distribution. For a fixed parameter θ calculate $\left|e^{-\frac{1}{2}2.48^2-2.48z_i}-1\right|^{\theta}$ and take the average for all draws.

$$\begin{aligned}
\frac{1}{R(T)} &= e^{-\frac{1}{2}k^2T-kW^{\mathbb{Q}}(T)} \\
\mathbb{E}^{\mathbb{Q}}\left[\left|\frac{1}{R(T)}-1\right|^{\theta}\right] &= \int_{-\infty}^{\infty}\left|e^{-\frac{1}{2}k^2T-k\sqrt{T}z}-1\right|^{\theta} n(z)dz \\
&= \int_{-\infty}^{\infty}\left|e^{-\frac{1}{2}2.48^2-2.48z}-1\right|^{\theta}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz
\end{aligned}$$

C.9 Cressie-Read

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= \mathbb{E}^{\mathbb{P}}\left[\frac{1-\theta+\theta\frac{1}{R(T)}-\left(\frac{1}{R(T)}\right)^{\theta}}{\theta(1-\theta)}R(T)\right] \quad (C.23) \\
&= \frac{1}{\theta(1-\theta)}\mathbb{E}^{\mathbb{P}}\left[R(T)-\theta R(T)+\theta-R(T)^{1-\theta}\right] \\
&= \frac{1}{\theta(1-\theta)}\left(1-\theta+1-\mathbb{E}^{\mathbb{P}}\left[R(T)^{1-\theta}\right]\right) \\
&= \frac{1}{\theta(1-\theta)}\left(1-\theta+1-\mathbb{E}^{\mathbb{P}}\left[\left(e^{-\frac{1}{2}k^2T+kW^{\mathbb{P}}(T)}\right)^{1-\theta}\right]\right)
\end{aligned}$$

If $R(T)$ is log-normal distributed with volatility k,

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[R(T)\phi\left(\frac{1}{R(T)}\right)\right] &= \frac{1}{\theta(1-\theta)}\left(2-\theta-e^{-\frac{1}{2}k^2\theta(1-\theta)T}\right) \quad (C.24) \\
&= \frac{1}{\theta(1-\theta)}\left(2-\theta-e^{-3.08\theta(1-\theta)}\right) \\
&= \frac{1}{\theta(1-\theta)}\left(2-\theta-0.05e^{\theta(1-\theta)}\right)
\end{aligned}$$

References

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228. 5

Avellaneda, M., Levy, A., and Parás, A. (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance*, 2(2):73–88. 10

Barrieu, P. M. and El Karoui, N. (2007). Pricing, hedging and optimally designing derivatives via minimization of risk measures. 8

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357. 2, 12

Breuer, T. and Csiszár, I. (2013). Systematic stress tests with entropic plausibility constraints. *Journal of Banking & Finance*, 37(5):1552–1559. 2

Chen, Z. and Epstein, L. (2002). Ambiguity, risk, and asset returns in continuous time. *Econometrica*, pages 1403–1443.

Delbaen, F. (2006). The structure of m–stable sets and in particular of the set of risk neutral measures. In *In Memoriam Paul-André Meyer*, pages 215–258. Springer. 8

Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31. 8

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58. 2

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton university press. 2

Hansen, L. P. and Sargent, T. J. (2015). Sets of models and prices of uncertainty. *University of Chicago manuscript*.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497. 3

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 186, pages 453–461. The Royal Society. 16

Neyman, J. and Pearson, E. S. (1933). The testing of statistical hypotheses in relation to probabilities a priori. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 29, pages 492–510. Cambridge Univ Press. 3

Øksendal, B. (2003). *Stochastic differential equations*. Springer. 10, 12

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42. 5, 9

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471. 5, 9

Rosazza Gianin, E. (2006). Risk measures via g-expectations. *Insurance: Mathematics and Economics*, 39(1):19–34. 8

Vanden, J. M. (2006). Exact superreplication strategies for a class of derivative assets. *Applied Mathematical Finance*, 13(01):61–87. 10